BIODIESEL QUALITY MONITORING USING VIBRATIONAL SPECTROSCOPY

by

MARCELO CORONADO HIGUERO

M.Sc., Technological University of Panama, 2006

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Biological and Agricultural Engineering
College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2012

Abstract

Biodiesel production and utilization has been increasing rapidly worldwide in recent years. A main challenge in the commercialization and public acceptance of biodiesel is its quality control. This work reports the use of infrared spectroscopy to monitor biodiesel quality through the development of models to predict (1) the blending level of biodiesel in biodiesel-diesel mixtures, (2) the fatty acid profile of biodiesel fuels derived from various lipids, and (3) the concentration of most common impurities present in biodiesel including water, glycerol, methanol and triglycerides.

Regressions based on near-infrared (NIR) spectroscopy were developed for relatively inexpensive and rapid on-line measurement of the concentration and specific gravity of biodiesel-diesel blends. Methyl esters of five different oils—soybean oil, canola oil, palm oil, waste cooking oil, and coconut oil—and two different brands of commercial-grade No. 2 on-highway diesel and one brand of off-road No. 2 diesel were used in the calibration and validation processes. The predicted concentration and specific gravity of the biodiesel-diesel blends were compared with the actual values. The maximum and average root-mean-square errors of prediction (RMSEP) of biodiesel concentration were 5.2% and 2.9%, respectively, from the biodiesel type-specific regression. For the general regression, the RMSEP were 3.2% and 0.002 for biodiesel concentration and specific gravity predictions, respectively.

Five different models were developed to determine the concentration of methyl palmitate (C16:0), methyl stearate (C18:0), methyl oleate (C18:1), methyl linoleate (C18:2), and methyl linolenate (18:3) present in biodiesel. Using the NIR range a set of models based on four different types of biodiesel was developed. The maximum RMSEP was 0.553% when the models were validated with biodiesel samples that were used in the calibration, however, prediction accuracy of the model under external samples was poor, therefore, a new set of models was proposed. For this case, six different types of biodiesel were used. The models developed for C18:1, C18:2 and C18:3 presented good accuracy on prediction. However, for C16:0 and C18:0, additional work was necessary to reach reasonable accuracy in prediction. Three sub models for specific ranges of concentration (low, medium, and high) were developed. The RMSEP was

reduced from 2.98% to 1.51% for the C16:0 and from 2.33% to 0.56% for C18:0, when the sub-models were validated under internal and external samples. Similar procedures were followed to develop regression models based on mid infrared (MIR) spectra. The RMSEP for C16:0, C18:0, C18:1, C18:2, and C18:3 were 0.83%, 0.37%, 1.45%, 1.59%, and 0.84%, respectively. Predictions using MIR spectroscopy models were better than those obtained with NIR spectroscopy models for the C16:0 and C18:0 models.

The most common impurities present in biodiesel from production processes, including methanol, free glycerol, triglycerides, and water, were determined by infrared methods using NIR and MIR spectra and partial least square regression (PLSR) methods. The models were developed in two different approaches, one was when a single impurity was present and the other was when all impurities were present. In the single impurity models, the maximum RMSEP obtained in the NIR and MIR models were 647 mg kg^{-1} and 206 mg kg^{-1}, respectively. The models for methanol, glycerol, and water performed better using the NIR data. For the triglycerides model, MIR worked better. Only NIR data were used to develop the models for samples with all impurities. Data pre-treatment (Savitzky-Golay second derivative) was necessary to achieve reasonable accuracy in the predictions in this type of models. The maximum RMSEP was 932 mg kg^{-1} presented in the model for triglycerides. The best performance was obtained in the model developed to predict methanol concentration in biodiesel with RMSEP of 177 mg $kg^{-1}$when all listed impurities were presented.

The feasibility of using NIR and MIR spectroscopy to monitor biodiesel quality was demonstrated in this work. The developed method was accurate, rapid, convenient, yet inexpensive to determine some important characteristics of biodiesel, such as biodiesel blending level in biodiesel-diesel mixtures, the fatty acid profile of biodiesel, and impurities present in the fuel.

BIODIESEL QUALITY MONITORING USING VIBRATIONAL SPECTROSCOPY

by

MARCELO CORONADO HIGUERO

M.Sc., Technological University of Panama, 2006

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Biological and Agricultural Engineering
College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2012

Approved by:

Major Professor
Wenqiao Yuan

Abstract

Biodiesel production and utilization has been increasing rapidly worldwide in recent years. A main challenge in the commercialization and public acceptance of biodiesel is its quality control. This work reports the use of infrared spectroscopy to monitor biodiesel quality through the development of models to predict (1) the blending level of biodiesel in biodiesel-diesel mixtures, (2) the fatty acid profile of biodiesel fuels derived from various lipids, and (3) the concentration of most common impurities present in biodiesel including water, glycerol, methanol and triglycerides.

Regressions based on near-infrared (NIR) spectroscopy were developed for relatively inexpensive and rapid on-line measurement of the concentration and specific gravity of biodiesel-diesel blends. Methyl esters of five different oils—soybean oil, canola oil, palm oil, waste cooking oil, and coconut oil—and two different brands of commercial-grade No. 2 on-highway diesel and one brand of off-road No. 2 diesel were used in the calibration and validation processes. The predicted concentration and specific gravity of the biodiesel-diesel blends were compared with the actual values. The maximum and average root-mean-square errors of prediction (RMSEP) of biodiesel concentration were 5.2% and 2.9%, respectively, from the biodiesel type-specific regression. For the general regression, the RMSEP were 3.2% and 0.002 for biodiesel concentration and specific gravity predictions, respectively.

Five different models were developed to determine the concentration of methyl palmitate (C16:0), methyl stearate (C18:0), methyl oleate (C18:1), methyl linoleate (C18:2), and methyl linolenate (18:3) present in biodiesel. Using the NIR range a set of models based on four different types of biodiesel was developed. The maximum RMSEP was 0.553% when the models were validated with biodiesel samples that were used in the calibration, however, prediction accuracy of the model under external samples was poor, therefore, a new set of models was proposed. For this case, six different types of biodiesel were used. The models developed for C18:1, C18:2 and C18:3 presented good accuracy on prediction. However, for C16:0 and C18:0, additional work was necessary to reach reasonable accuracy in prediction. Three sub models for specific ranges of concentration (low, medium, and high) were developed. The RMSEP was reduced from 2.98% to 1.51% for the C16:0 and from 2.33% to 0.56% for C18:0, when the sub-

models were validated under internal and external samples. Similar procedures were followed to develop regression models based on mid infrared (MIR) spectra. The RMSEP for C16:0, C18:0, C18:1, C18:2, and C18:3 were 0.83%, 0.37%, 1.45%, 1.59%, and 0.84%, respectively. Predictions using MIR spectroscopy models were better than those obtained with NIR spectroscopy models for the C16:0 and C18:0 models.

The most common impurities present in biodiesel from production processes, including methanol, free glycerol, triglycerides, and water, were determined by infrared methods using NIR and MIR spectra and partial least square regression (PLSR) methods. The models were developed in two different approaches, one was when a single impurity was present and the other was when all impurities were present. In the single impurity models, the maximum RMSEP obtained in the NIR and MIR models were 647 mg kg^{-1} and 206 mg kg^{-1}, respectively. The models for methanol, glycerol, and water performed better using the NIR data. For the triglycerides model, MIR worked better. Only NIR data were used to develop the models for samples with all impurities. Data pre-treatment (Savitzky-Golay second derivative) was necessary to achieve reasonable accuracy in the predictions in this type of models. The maximum RMSEP was 932 mg kg^{-1} presented in the model for triglycerides. The best performance was obtained in the model developed to predict methanol concentration in biodiesel with RMSEP of 177 mg $kg^{-1}$when all listed impurities were presented.

The feasibility of using NIR and MIR spectroscopy to monitor biodiesel quality was demonstrated in this work. The developed method was accurate, rapid, convenient, yet inexpensive to determine some important characteristics of biodiesel, such as biodiesel blending level in biodiesel-diesel mixtures, the fatty acid profile of biodiesel, and impurities present in the fuel.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to show my gratitude to Dr. Wenqiao Yuan for his encouragement, guidance and support during the development of this work.

I also thank the members of my graduate committee Professor Donghai Wang, Dr. Floyd E. Dowell, Dr. Richard Nelson, and to the outside chairperson, Professor David Steward, for their guidance and suggestions.

I want to express gratitude to the faculty and staff in the Department of Biological and Agricultural Engineering, especially Barb Moore, Kerri Ebert, Randy Erickson, and Darrell Oard for their cooperation in all the activities related to this research. I also want to express my gratitude to Dr. Joseph Harner, Department Head, and Dr. Naiqian Zhang, Graduate Program Director.

I would like to acknowledge the Technological University of Panama and the IFARHU-SENACYT scholarship program for their financial support.

Finally, I offer my regards and blessings to all of those who supported me in any respect during our stay in Manhattan, Kansas.

Dedication

To my best friend and wife
Deyka

To my kids
Andres, Reggeany, Alan, and Alec

To my parents
Marcelo and Telvina

To my brother and his family
Gregorio

To my sisters and their families
Rita, Nelky, Osmilda, and Rosemary

To my university
Technological University of Panama

To my country
Panamá

Chapter 1 - Introduction

1.1 Problem Statement

Biodiesel is mainly produced from vegetable oils such as waste vegetable oil (WVO), soybean, and canola/rapeseed. New sources of oils or fats in algae, Jatropha, palm, greases, etc., have been used to implement the second generation of biofuel. Biofuels from nonfood crops circumvent the competition between food production and biofuel expansion. The large variety of raw materials for biodiesel production makes biodiesel quality control an immense challenge. However, the fatty acid profile of feed stocks/fuel has been mentioned in literature as the most important factor to determine biodiesel properties. Consequently, determination of the fatty acid profile in biodiesel is an important task in the processes of monitoring and controlling quality of the fuel. In addition, various concentrations of biodiesel are commonly used as blends with petroleum-based diesel fuels, which makes biodiesel monitoring and engine performance optimization even more difficult. Additionally, biodiesel is mainly produced through a transesterification process, using a low molecular weight alcohol and a catalyst. Due to this production method, the final products often contain impurities such as unreacted triglycerides, free glycerol, catalyst, and residual alcohol. These contaminants undermine engine performance and can cause severe engine problems. Most analytical methods currently used in biodiesel are based on chromatographic analysis. Even though this method is suitable and accurate, it is a time-consuming process that requires well-trained personnel and expensive reagents. Therefore, a simple, fast, and accurate method to determine blending level in biodiesel-diesel blend, fatty acid profile, and the most common impurities present in biodiesel is extremely desired.

1.2 Background

Biodiesel is an alternative biofuel produced by chemical reaction, transesterification, in which vegetable oils or animal fat react with a short-chain alcohol in the presence of a catalyst Van Gerpen (2005). Related to the production method biodiesel usually contains residual alcohol, catalyst, and free glycerol therefore a final washing process is necessary to minimize or eliminate these components, considered impurities, in the final product. In order to ensure proper performance of engines and avoid engine problems, biodiesel has to meet all requirements included in ASTM D6751-11B the nationally accepted fuel standard in the United States, or EN

14214 applied in Europe. Table 1-1 summarizes the requirements for biodiesel (B100) according to the ASTM standard.

Table 1-1 Detailed Requirements for Biodiesel (B100) (All Sulfur Levels)

Property	Test Method	Grade S15 Limits	Grade S500 Limits	Units
Calcium and Magnesium, combined	EN 14538	5 max	5 max	ppm (µg/g)
Flash point (closed cup)	D93	93 min	93 min	ºC
Alcohol control, One of the following must be met: 1- Methanol content	EN14110	0.2 max	0.2 max	mass %
2-Flash point	D93	130 min	130 min	ºC
Water and sediment	D2709	0.050 max	0.050 max	% volume
Kinematic viscosity, 40 ºC	D445	1.9 – 6.0	1.9 – 6.0	mm^2/s
Sulfated ash	D874	0.020 max	0.020 max	% mass
Sulfur	D5453	0.0015 max (15)	0.05 max (500)	% mass (ppm)
Copper strip corrosion	D130	No. 3 max	No 3 max	
Cetane number	D613	47 min	47 min	
Cloud point	D2500	Report	Report	ºC
Carbon residue	D4530	0.050 max	0.050 max	% mass
Acid number	D664	0.50 max	0.50 max	mg KOH/g
Cold soak filterability	D7501	360 max	360 max	Seconds
Free glycerin	D6584	0.020 max	0.020 max	% mass
Total glycerin	D6584	0.240 max	0.240 max	% mass
Phosphorus content	D4951	0.001 max	0.001 max	% mass
Distillation temperature, Atmospheric equivalent temperature 90% recovered	D1160	360 max	360 max	ºC
Sodium and potassium, combined	EN 14538	5 max	5 max	ppm (µg/g)
Oxidation stability	EN 15751	3 minimum	3 minimum	Hours

From: Standard ASTM 6751-11B

Biodiesel is commonly used as a blend of biodiesel and diesel fuel. Biodiesel blend indicated as BXX refers to blend of XX % (v) of biodiesel and 100-XX % (v) of diesel fuel. It is important to note that the current methods to test the quality of biodiesel are referring to biodiesel blended until 20% (v) ASTM D7467-10. Little work has been done to monitor important properties of biodiesel after it is blended with diesel fuel in quantities greater than 20%(v). Contaminants in biodiesel such as methanol, residual glycerol, triglycerides (unreacted oil), and water even in very small amounts can damage diesel engines. Triglycerides in biodiesel can generate emissions of noxious pollutants from the combustion and injector coking. Methanol content in biodiesel promotes lacquer deposit formation on the injectors. Furthermore, methanol content in biodiesel can adversely affect some important fuel properties, for instance, heating value and flash point, decreasing performance of the fuel. Glycerol content in biodiesel has been reported as a cause of injection system corrosion and deposit formation in the injector, and also emission problems, which increase aldehydes emission. Finally, water presence in biodiesel can cause corrosion problems in the engine and can also react with triglycerides producing an undesirable compound.

The obtained information from the study of biodiesel blending levels can be used to develop a simple model or device to detect the blending levels of biodiesel in diesel, helping to adjust the combustion timing to reduce the NOx emissions in engines from biodiesel-diesel combustion. Additionally, some properties of biodiesel-diesel blends can be indirectly measured when the blending level is known. With models to determine the fatty acid profile of biodiesel, important information will be available; specifically, information related to fatty acids with double bonds to determine if the biodiesel meets the standard requirements (European), information related to the oil sources, which can be used to improve the biodiesel production, and important information about properties of biodiesel such as viscosity, oxidative stability, and iodine value that can be indirectly determined knowing the fatty acid profile. The models to quantify the impurities found in biodiesel can replace the gas chromatograph analysis throughout all the biodiesel stages, from production until retail sales. Finally these models constitute the necessary tools to determine the quality of biodiesel in accurate and fast ways, with the added advantage of direct applicability because of its online measurement characteristics.

1.2.1 Vibrational spectroscopy

The term vibrational spectroscopy is applied to any technique used to obtain vibrational data from samples which have been specifically developed to study the vibrations of molecules based on the interaction between electromagnetic radiations and matter. The theory of infrared spectroscopy is based on the fact that all atoms in molecules are in continuous vibration. When the frequency of a specific vibration is equal to the frequency of the infrared radiation of the molecule, the molecule absorbs the radiation. It is important to note that the functional groups in molecule samples only absorb infrared radiation at selected frequencies. This radiation corresponds to the different vibration modes of the bonds in the molecules. There are two general types of vibration, bending and stretching as shown in Figure 1-1. Bending is defined as the change in the bond angle; it can be rocking or deformation depending on whether the movement is in the same or opposite directions. Stretching can be symmetrical when it is on the plane or asymmetrical when it is out of the plane.

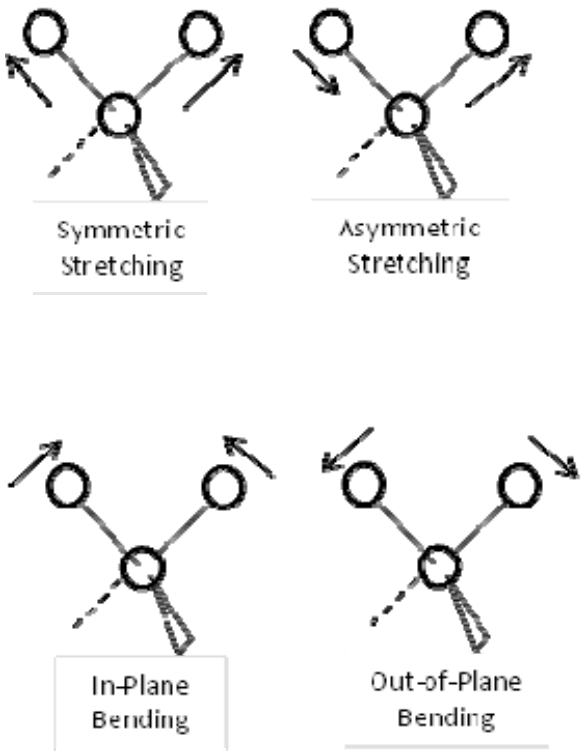

Figure 1-1 Vibration modes of atoms when they are IR radiated.

Infrared spectroscopy is divided into three regions, the far infrared (400 to 10 cm^{-1}), mid infrared (MIR, 4,000 to 400 cm^{-1}), and near infrared (NIR, 12,820 to 4,000 cm^{-1}). Mid and near infrared spectroscopy have been commonly used to analyze a large variety of compounds; in the mid infrared, the fundamental molecular vibration occurs and its combination bands and overtones are presented in the near infrared range. To go from fundamental to the first overtone, the intensity of an absorption band is reduced by a factor from 10 to 100; consequently, the sensitivity of near infrared spectroscopy is lower than the sensitivity of mid infrared spectroscopy.

The vibrational degree of freedom represents the number of fundamental vibrational frequencies of the molecula or normal modes with atoms moving in phase with the same frequency to reach its position of maximun displacement and passing through its equilibrium position at the same time. In the molecula certain vibrational modes are localized as a bond vibration. The recognition of charecteristic local bond or local group frequencies which is associated with absortion spectra represents the success of infrarred spectroscopy as an analitical tool. When a molecula containing two equivalent bond with a common frequency oscillates, one of them will be resonantly excited by vibrations and energy will flow between them at another frequency governed by the strength of interbond coupling . The two frequencies are independent of energy, however this normal mode depends on the harmonic approximation. In a more anharmonic model the individual bond frequencies will vary with energy, typically decreasing, for stretching vibrations, as the energy increases. Overtone bands in an infrared spectrum are analogous and are multiples of the fundamental absorption frequency. Due to the fact that the energy is proportional to the frequency absorbed and this is proportional to the wavenumber the first overtone requires twice the energy of the fundamental.

Two general types of infrared spectroscopy have been used to obtain the spectra; dispersive IR spectroscopy and the Fourier transform infrared (FTIR) spectroscopy. In dispersive spectroscopy, the source of energy travels through samples and the reference path, then it goes to the chopper to adjust the energy level that will reach the detector; finally, the source of energy is sent to the diffraction grating (monochromator), which splits the wavelengths of spectral range and sends each wavelength individually to the detector, one at a time. FTIR spectrophotometer use halogen bulb, the energy for the source is directed into an interferometer where the energy is

transformed by the computer into its actual electromagnetic frequency. This interferometer uses a beam splitter to divide the beam radiation from the source into two parts, one part is sent to the stationary mirror and the other part is sent to a moving mirror. When the beams are reflected to the beam splitter, it generates an interference pattern called interferogram, this interferogram travels from the beam splitter to the sample, where some energy is absorbed and the rest transmitted to the detector. In this case the detector reads the information of every wavelength simultaneously. After that, the signal is sent to the computer where an algorithm denominated Fourier transform is used to transform the interferogram into a single beam spectrum.

1.2.2 Infrared spectra of biodiesel

Biodiesel is a mix of fatty acid methyl ester, with different lengths and degrees of saturation of the chains. Although spectra of biodiesels appear to be similar, they differ in the intensity of their band as well as in the exact frequency at which the maximum absorbance is produced in each type of biodiesel, caused by different nature and composition of oils used to produce the biodiesel. These differences are used to discriminate among biodiesel.
Table 1-2 describes the frequencies of most prominent bands and shoulders which are characteristic of biodiesels on the MIR range.

Table 1-2 The frequencies of band and shoulders of biodiesel in the MIR range.

Frequency (cm^{-1})	Functional Group	Mode of vibration	Intensity
3009	=C-H	Stretching	Medium
2924	-CH	Stretching (Asymmetric)	Very strong
2854	-CH	Stretching (symmetric)	Very strong
1740	-C=O	Stretching	Very strong
1654	-C=C	Stretching	Very weak
1460	-CH	Bending (scissoring)	Medium
1240	$-CH_2$	Stretching (bending)	Medium
1165	$-CH_2$	Stretching (bending)	Strong
720	$-(CH_2)_n$	Bending (rocking)	Medium

Yaakob (2010) and Sablinskas (2003).

Typical infrared spectra of biodiesel are presented in Figure 1-2. Characteristic peaks of –CH stretching asymmetric/symmetric are observed around 2900 cm^{-1} in MIR range.

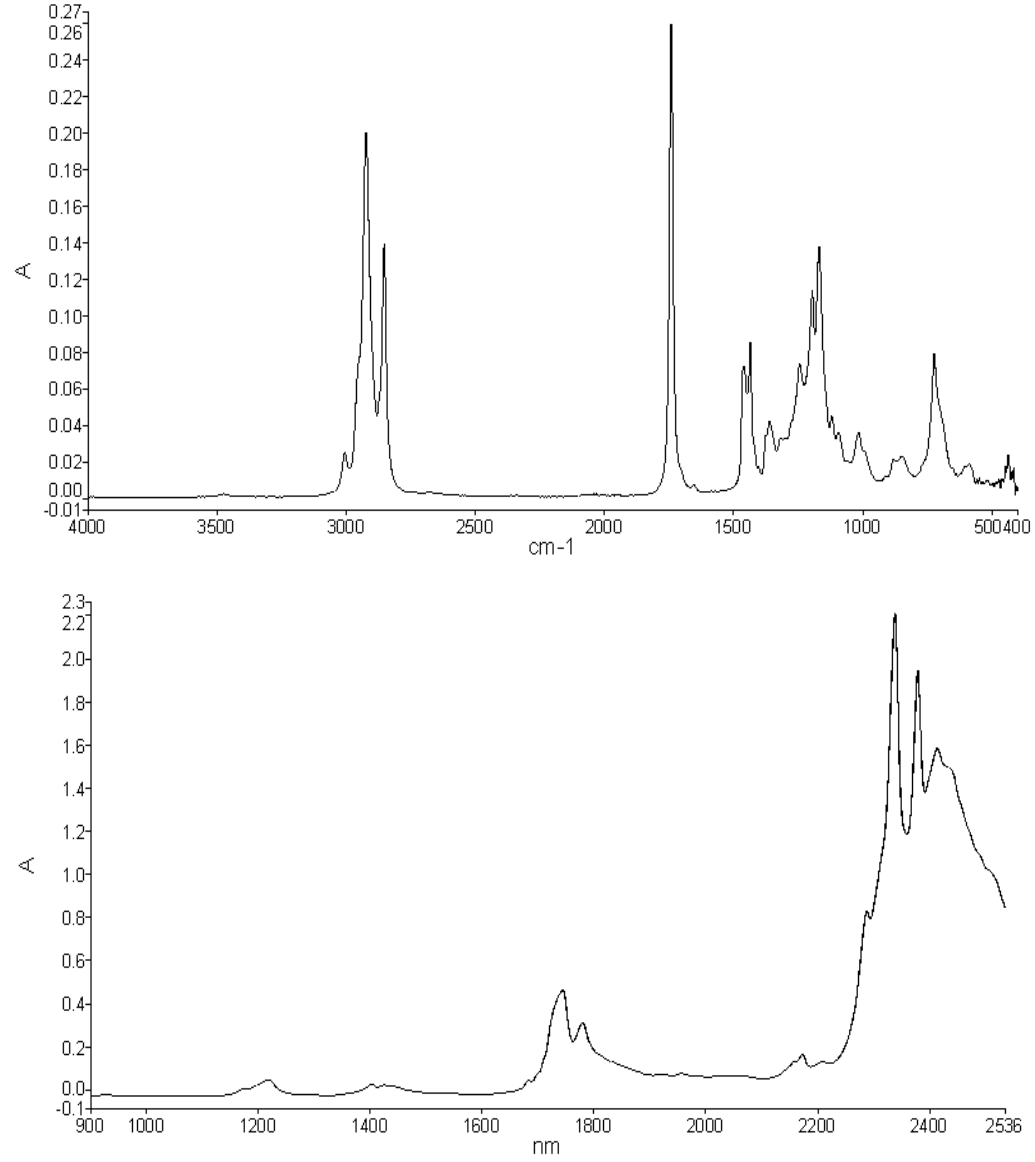

Figure 1-2 Typical spectra in the infrared range: MIR (top) and NIR (bottom)

In the NIR range the bands are a composite of many bands containing information on more than one type of vibration, the most prominent band and shoulder present in the biodiesel spectra are presented in Table 1-3. The first overtones of –CH stretching are observed in the NIR range around 1700 nm.

Table 1-3 The frequencies of band and shoulders of biodiesel in the NIR range.

Frequency (nm)	Functional Group	Mode of vibration	Intensity
1160	-C=O	Stretching fourth overtone	Very weak
1190	$-CH_3$	Stretching second overtone	Weak
1215	$-CH_2$	Stretching second overtone	Weak
1395	-CH	Combination	Weak
1415	-CH	Combination	Weak
1705	$-CH_3$	Stretching first overtone	Medium
1725	$-CH_2$	Stretching first overtone	Very strong
1765	-CH	Stretching first overtone	Strong
2140	-CH/C=O	Stretching, combination or Sym. Def.	Medium
2170	-CH	Stretching, combination or Asym. Def.	Medium

Shenk (2008) and Sablinskas (2003).

1.3 Research Objectives

The overall goal of this work was to apply infrared spectroscopy and statistical methods to monitor biodiesel quality. The biodiesel level in biodiesel-diesel blends, specific gravity, fatty acid profile, and impurities commonly present in biodiesel such as methanol, free glycerol, triglycerides, and water were determined. Specific objectives were as follows:

1) To develop a regression model to determine the concentration of biodiesel in biodiesel-diesel blends using near infrared spectroscopy spectra. The specific gravities of biodiesel-diesel blends were also predicted using the same spectra and multiple linear regression (MLR) method.

2) To develop a prediction model for fatty acid profiles of biodiesel using near and mid infrared spectroscopy. Concentrations of five main fatty acids that are present in most biodiesel,

including palmitic, stearic, oleic, linoleic, and linolenic acids, were determined using the raw spectra in near and mid infrared range.

3) To develop a prediction model to quantify impurities commonly present in biodiesel such as methanol, free glycerol, triglycerides, and water, using mid and near infrared spectroscopy data. Models to predict above listed impurities were developed using raw spectra in near and mid infrared range and partial least square regression method.

4) To evaluate the performance of derivative technique as pre-treatment of the data for biodiesel analysis using infrared spectroscopy. Based on near-infrared spectra data and derivative technique as pre-treatment, the fatty acid profile and impurities present in biodiesel such as methanol, triglycerides, water, and glycerol were determined.

1.4 Organization of Dissertation

This dissertation has seven chapters. The first chapter presents the problem, objectives, and summarizes the importance of the research. Chapter 2 reviews the literature related to biodiesel quality monitoring. In Chapter 3 a method to predict the concentration and specific gravity of biodiesel-diesel blend is proposed and developed. In Chapter 4 methods to determining the fatty acid profile of biodiesel using Fourier-transformed near and mid infrared spectroscopy were developed. Chapter 5 presents models to quantify trace biodiesel impurities based in Fourier-transformed near and mid infrared spectroscopy. In Chapter 6 a model to determine fatty acid composition in biodiesel focusing on derivative technique as pre-treatment was developed. Chapter 7 provides conclusions and future work discussion.

1.5 References

ASTM Standard D6751-11b. Standard Specification for Biodiesel Fuel Blend Stock (B100) for Middle Distillate Fuels, ASTM International, West Conshohocken, PA, 2011, DOI: 10.1520/D6751-11B, www.astm.org.

ASTM Standard D7467-10. Standard Specification for Biodiesel Fuel Oil, Biodiesel Blend (B6 to B20), ASTM International, West Conshohocken, PA, 2010, DOI: 10.1520/D7467-10, www.astm.org.

Banga, S., & Varshney, P. K. (2010). Effect of impurities on performance of biodiesel: A review. Journal of Scientific & Industrial Research, 69(8), 575-579.

European Standard EN 14214:2008: E. Automotive fuels – fatty acid methyl esters (FAME) for diesel engines – requirements and test methods. European Committee for Standardization 2008.

Lapuerta, M. (2009). Effect of ethanol on blending stability and diesel engine emissions. Energy Fuels, 23(9), 4343.

Mittelbach M, Remschmidt C. Biodiesel. The comprehensive handbook. Graz, Austria: Martin Mittelbach; 2004.

Peterson, C. L. (1986). Vegetable oil as a diesel fuel: Status and research priorities. Transactions of the ASAE, 29(5), 1413.

Sablinskas, V., Steiner, G., Hof, M. (2003) Chapter 6-Application. In G. Gauglitz, and T.Vo-Dinh (Eds.), Handbook of Spectroscopy (pp.89-168). Germany: WILEY-VCH.

Sagar, Naik, Meher, (2006). Technical aspects of biodiesel production by transesterification - a review. Renewable Sustainable Energy Reviews, 10(3), 248.

Shenk, J. S., Workman, J. J., Jr., Westerhaus, M. O. (2008) Application of NIR Spectroscopy to Agricultural Products. In D. Burns and E. Ciurczak (Eds.), Handbook of Near-Infrared Analysis (pp.347-386). Boca Raton, FL: Taylor &Francis Group.

Van Gerpen, J. Business Management for Biodiesel Producer, National Renewable Energy Laboratory, SR-510-36242, July 2004.

Yaakob, B. C. M., Syahariza, Z. A., Rohman, A. (2010) Chapter 1. Fourier Transform (FTIR) Spectroscopy: Development, Techniques, and Application in the Analyses of Fats and Oil. In O. J. Rees (Eds.), Fourier Transform Infrared Spectroscopy: Development, Techniques, and Application (pp. 1-26).New York, NY: Nova Science.

Chapter 2 - Related Current and Previous Work

2.1 Biodiesel-diesel Blend Level Detection

Many researchers have reported studies related to detecting concentrations of biodiesel in diesel blends. Knothe (2001) determined the blend level of mixtures of biodiesel with conventional diesel fuel using fiber-optic near infrared spectroscopy. He reported that the peaks at 6005 cm^{-1} and 4600 - 4800 cm^{-1} in the near infrared range could be used to identify the concentration of biodiesel in diesel blends. These studies were conducted using only soybean methyl esters, and may not be applied for biodiesel detection from other sources.

Pimentel et al. (2006) developed models using partial least square regression using near and mid infrared spectra. The models were able to predict biodiesel concentration in biodiesel-diesel blend, with the presence of 0 to 5% raw oil in the sample, based on good correlation coefficient. The region of the near infrared spectra used in this study was between 2200 – 2280 nm. The RMSEP was 0.18 % (v/v) and the relative average error was 6.7 %. For the mid infrared case the range used was between 1700 -1800 cm^{-1}. The root means square error of prediction (RMSEP) was 0.25 % (v/v) and the relative average error was 10.2 %. Pretreatment of the data was necessary to get good results. First derivative and smoothing by a Savitzky-Golay filter were used. Even with these promising results, calibration with a wider range of biodiesel in biodiesel-diesel blend is necessary. Because the models were developed for blending levels from 0 to 2% of biodiesel, it is not applicable in wider ranges of blend. The standard method to determine biodiesel content in biodiesel-diesel blend is the ASTM D 7371-07. This standard is based on FTIR-ATR-PLS method, which has a few limitations. The method was developed for blending levels between 1 and 20%. The effectiveness of the method has not been demonstrated above this range. The method can be applied for fatty acid methyl ester (FAME), but not for fatty acid ethyl ester (FAEE). This method was developed using only soybean methyl ester, and may not be applied for biodiesel detection from other sources. For this reason, further studies are necessary for the use on wider ranges of blends and larger variety of biodiesel.

Zawadzki and Shrestha (2009) developed a sensing model for biodiesel feedstock and blending level using visible light spectra and neural network. The obtained model showed low standard error values (1.85%) at 95% confidence interval in the range from 470 to 490 nm with known feedstock. In their model for an unknown biodiesel source, the range between 380 to 530

nm was used with the neural network approach. This model recognized the biodiesel feedstock; however, it only gave a rough estimation of biodiesel blend.

Considering the results and methods used in previous works related to biodiesel levels in biodiesel-diesel blend, additional study is necessary to obtain a method with demonstrated applicability on different sources of biodiesel as well as diesel fuel types.

2.2 Determining Fatty Acid Profile of Biodiesel Samples

Several studies were found to determine fatty acid profile in a few products using near and mid infrared spectroscopy including the subcutaneous fat of Iberian breed swine by Gonzles et al. (2003), and cow milk by Soyeurt et al. (2006) But only the study developed by Batista et al. (2008) made references to analyzing fatty acid composition in biodiesel.

Gonzales et al. (2003) published a successful determination of fatty acid in fat of swine using NIR spectroscopy. Two different types of samples were used to develop the study, one applying the fiber-optic probe directly on intact subcutaneous fat and the other using the lipid extracted from subcutaneous fat. When the samples from extracted lipid (model I) were used the best standard error of calibration (SEC) was presented in the model for C14:0. It was 0.09 % and the worst was found in the model for C18:1, which was 0.74 %. When the fiber-optic probe was applied directly over the subcutaneous fat (model II) the best and worst SEC were 0.09% and 0.97 % for the C14:0 and C18:1 models, respectively. The best and worst standard error of prediction (SEP) for the models of type I was 0.13 % and 0.97 %, presented in the models for C14:0 and C18:1 respectively. For the cases of the models of type II the best and worst SEP were 0.11 % and 1.20 % presented again in the models for C14:0 and C18:1 , respectively. The authors conclude, in both types of models, results are comparable to the reference method used. This study suggests that NIR could be applied to determine the fatty acid profile in biodiesel.

The study of Soyeurt et al. (2006) showed relatively successful estimation of fatty acid content in cow milk using MIR spectroscopy. The regions from 1736 to 1805 cm^{-1} and from 2823 to 3016 cm^{-1} were used in this study. Models for C4:0, C6:0, C8:0, C10:0, C10:1 cis-9, C12:0, C14:0, C14:1, C15:0, C16:0, C16;1, C18:0, C18:1*cis-9*, C18:2*cis-9, cis-12*, C18:3*cis-9, cis-12, cis-15*, and C18:2*cis-9, trans-11* were developed. The performance of the method was evaluated only by standard error of cross validation (SECV). The SECV were ranged from 0.01 to 0.18 g/dL of milk. The authors conclude that MIR can be used to predict the concentration of

fatty acid in cow milk, even though low performance was observed in the models with low concentration of fatty acid. Considering the results and the fact that cow milk is a very complex chemical structure compared to biodiesel, MIR method could be a promising method to determine the fatty acid profile in biodiesel.

Batista et al. (2008) developed multivariate calibration to determine ester content (total amount) in biodiesel samples using the near-infrared range between 9000 – 4500 cm^{-1}, and between 6102 – 5880 cm^{-1} in two different models. The correlation coefficients (R^2) for calibration were 0.913 and 0.924 for the first and second model. The RMSEP were 0.9 % and 1.0 % for each model, respectively. The spectra were pretreated applying a first order Savitsky-Golay derivative and the models were developed using PLS methods. Additionally, the authors presented models to predict linolenic acid (C18:3) methyl esters content (%) using the near-infrared range between 9000 – 4500 cm^{-1}. In these cases the correlation coefficients (R^2) was 0.995, using first order Savitsky-Golay derivative as pre-treatment. Models to predict the content of myristic acid (C14:0), palmitic acid (C16:0), stearic acid (C18:0), oleic (C18:1), and linoleic acid (C18;2) were presented . The RMSEP were 0.18%, 0.02%, 0.79%, 0.22%, 1.79%, and 2.5 % respectively. The validation process was developed using the same types of biodiesel used in the calibration, the use of external samples (biodiesel not used in the calibration) is recommended for this study to verify the robustness of the model under any type of biodiesel. Biodiesel from soybean, palm, and rapeseed were used in this study. Wide biodiesel type is recommended to perform this study, considering the sources of biodiesel are increasing with the passage of the time. The use of mid infrared range for this application is recommended because of its higher sensitivity when compared to the near infrared range. The developed models were selected based on the best statistical performance of few used pre-treatments. Calibration based on identification of compounds using their absorption bands is the recommended method. In addition, the use of the first derivative of the spectrum is not recommended for the interpretation or calibration because the pattern of peaks and valleys of the first derivative spectra does not correspond to the pattern of the original spectra, Shenk et al. (2007) Finally, the most common method to determine fatty acid profile in biodiesel is gas chromatography. As already discussed, this method is suitable and accurate, but it is time-consuming, and requires well-trained personnel to perform the analysis, and a few expensive reagents. Additionally, it is not an online measure that could be used for real-time monitoring processes.

2. 3 Quantification of Impurities in Biodiesel Samples

The attention for this kind of analysis has focused on the impurities that came from the production method. Several methods to detect methanol content, water content, free glycerol, and triglycerides have been proposed. Bondioli and Bella (2005) proposed a method to determine free glycerol in biodiesel using a spectrophotometric measurement at 410 nm. This procedure is based on periodate oxidation of glycerol, following the preparation of formaldehyde that later will react with acetylacetone. Even when this method showed good results to predict a free glycerol, detailed process and sample preparation are required.

Felizardo et al. (2007) developed a method using near infrared spectroscopy, PCA and PLS to determine water and methanol content at the same time in industrial and laboratory scale biodiesel samples. The researcher chose the region between 9000 – 4500 cm^{-1} of the spectra to perform the water content method and the results showed good performance, for calibration process R^2 was 0.990 and the best RMSEP was 87 mg/kg. In the case of methanol detection method, the region used was from 4800 to 5050 cm^{-1}. Again the result showed good performance, for calibration process R^2 was 0.997 and the best RMSEP was 70 mg/kg. First order Savitsky-Golay derivative with filter width of fifteen or thirty-three data points and the third-order polynomial was used as data pretreatment. Biodiesel from soybean, palm, and rapeseed were used as calibration samples. The results present excellent agreement between measured and predicted values. However, free glycerol or triglycerides analysis was not performed in this study.

Oliveira et al. (2007) presented a study to predict the concentration of triglycerides in a blend of diesel-biodiesel-triglycerides using FT-NIR spectroscopy and FT-Raman spectroscopy. The performance of PLS, PCR, and artificial neural network (ANN) methods were evaluated. When FT-NIR and FT-Raman were compared using PLS and PCR, the best value for RMSEP was 0.238 % (w/w) for the model based in FT-NIR and PLS method. The authors also mentioned that the PLS and PCR / FT-Raman models are not able to detect concentration of triglycerides in diesel-biodiesel-triglycerides blend. However, when FT-NIR and FT- Raman were compared using ANN, the best RMSEP was 0.092 % (w/w) for the model based in FT-Raman spectroscopy. Additionally, a comparison between PCR, PLS, and ANN was presented. No significant difference was found between the listed methods when FT-NIR spectra were used.

But, for the case of FT-Raman the model based in ANN presented the better accuracy in prediction. The authors concluded that FT-NIR as well FT-Raman spectroscopy combined to PCR, PLS, and ANN can be used to predict concentration of triglycerides accurately, in diesel-biodiesel-triglycerides blend, when the concentration of triglycerides ranged from 0 to 5 % (w/w). Even though these model present accurate results, the effectiveness of the model predicting triglycerides over the range required by the ASTM 6751 11b standard, had not been demonstrated.

The study of Soares et al. (2008) was developed to predict the triglyceride content in biodiesel using FT-MIR spectroscopy and PLS method. The level of triglycerides in biodiesel was ranged from 1 to 40 % (v/v). The spectra region chosen to develop the models was from 2760 to 1800 cm^{-1}. RMSEP of developed models ranged from 0.65 to 1.39 % (v/v) when the models were tested using one type of biodiesel. Results of the models when the three different types of biodiesel were used showed the RMSEP of 2.09 % (V/V). The authors concluded that FT-MIR method using PLS is able to predict the triglyceride concentration in biodiesel with good accuracy, when the range of the triglyceride content is ranged from 0 to 40 % (v/v). This study presents the same limitation of the Oliveira et al. study; the range of impurities is a lot larger than the requirement of ASTM standard.

Pisarello et al. (2010) presented a volumetric method to determine free and total glycerin in biodiesel. This method used the standard glycerin titration based on its oxidation by sodium periodate. Good results were shown, but the main drawback of this method is its manual execution and no information that can be used for live monitoring or control in production processes.

Dorado et al. (2011) determined methanol and glycerol traces in biodiesel using visible and NIR ranges and modified partial least square (MPLS) method. First derivative of (log (1/R), where R is reflectance) was used as pre-treatment of the data. The samples for this study ranged with methanol from 0.0003% to 0.433% (w/w) and with glycerol from 0.005% to 0.050 % (w/w) in two separate sets to meet the requirement of the EN 14214 standard. The accuracy of calibration was determined by the ratio of performance to deviation (RPD). If the RPD is > 3, the calibration model results are considered acceptable for analytical purposes, according to the authors. The RPD was 10 for the methanol model and 2.5 for the glycerol model, respectively. The authors concluded that NIR and visible ranges are able to detect methanol and glycerol

traces in biodiesel, but recommended additional work to improve the performance of the model to detect glycerol in biodiesel samples. The main limitation found in this study was that interaction of two impurities (methanol and glycerol) in the models was not evaluated. Work including more than one impurity is highly recommended.

Gaydou et al. (2011) reported the prediction of concentration of triglycerides in a blend of diesel-biodiesel- triglycerides. In the study biodiesel ranged from 0 to 10 % (w/w), vegetable oil from 0 to 30 % (w/w), and petroleum diesel from 60 to 100 % (w/w). The models were developed using serial-PLS and hierarchical-PLS, both of them a particular variation of the regular PLS method. Several pre-treatments of the data were evaluated including; base line correction, standard normal deviation, derivative, and mean normalization. For NIR range the best regression was obtained in the model without pre-treatment, RMSEP was 0.363% (w/w). For the model using MIR range the best performance was observed in the models where mean-normalization was used as pre-treatment; RMSEP was 1.939 % (w/w). The authors concluded that the developed models predicted triglycerides concentration with good accuracy when the concentration of triglycerides ranged from 0 to 30 % (w/w) in a blend of diesel-biodiesel-triglycerides. Again, therange of triglycerides concentration observed in the developed models is a lot greater than the requirement of ASTM standard.

No study was found predicting more than two impurities in biodiesel at the same time. Considering more than two impurities could be present in the biodiesel sample at the same time, they can affect the model performance. A practical model to evaluate specific impurities from production or distribution processes should include the effects of other possible impurities (methanol, water, triglycerides, and glycerol).

2.4 Derivative Spectroscopy Technique

Derivative spectra, using first or higher order, are frequently used to correct the baseline, , reduce the scattering effect, and perform band separation. Several studies over the year were found related to this technique.

Morrey (1968) developed a method to determine spectral peak position from composite spectra. The theoretical section of the study was presented based on spectrum of several overlapping constituent absorption band using information from Gaussian, Student T, and Lorentzian shape for each case separately. But, experimental section of the analysis was

developed using a spectrum of several overlapping (15) constituent absorption bands, each being of Lorentzian shape but of different widths, strengths, and degree of overlap. Derivatives of the spectra from the first to the fourth were calculated with respect to wavenumber. The objective was to find where the derivative is zero for all wavelengths of the all basic constituents of the spectra. This condition was matched at the third derivative. Using the results of the experiments most of the peaks from the basic constituent were clearly assigned, but a few of them were not specifically identified. The author concluded that when peaks are very close the separation of the peak is not possible. Huguenin and Jones (1986) presented an algorithm to perform the band separation from a combination band in reflectance spectra using derivative analysis. This analysis used a spectrum of six overlapping constituent absorption band, using Gaussian shape with different widths, strengths, and degree of overlap.

The experiment developed by Huguenin and Jones (1986) is graphically presented in Figure 2-1.

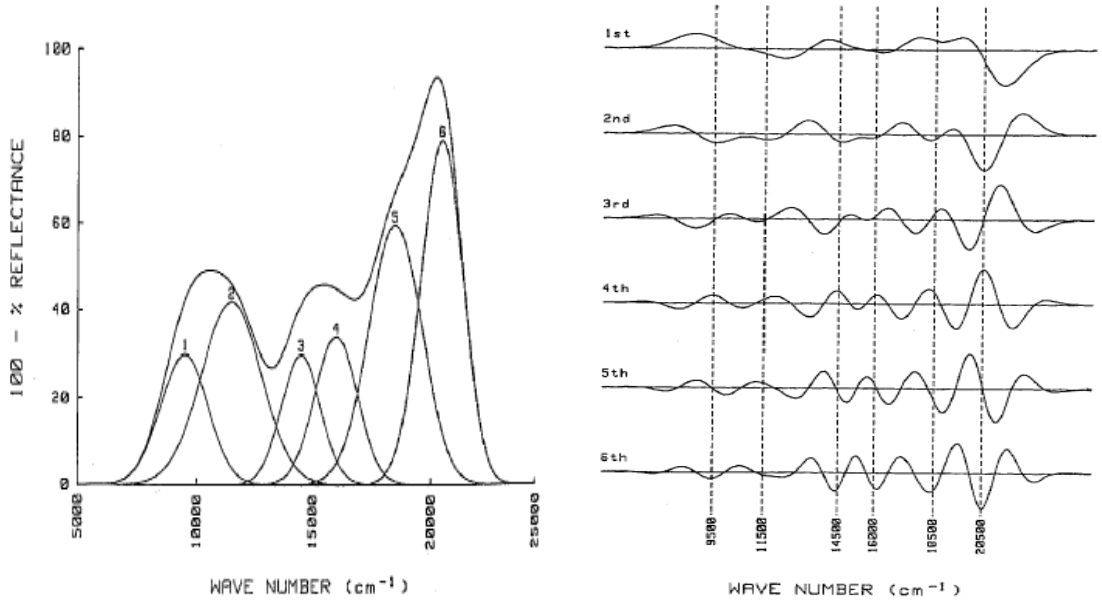

Figure 2-1 Synthetic spectrum composed of six Gaussian constituents (left), from the first to the sixth derivative spectra of the composed spectrum (right).

Derivatives of the spectra were calculated from the coefficient of a sixth-order polynomial with respect to wavenumber. Similar to the experiment of Morrey (1968), the Huguenin and Jones (1986) analysis was based on finding where a derivative is zero for all

wavelengths of the six basic constituents of the spectra. For this case the condition was matched for the fifth derivative. In both cases (Morrey and Huguenin), the authors released similar conclusions related to the error caused by an adjacent peak. Additional work is recommended to achieve a separation peak when they are too close.

Tsai, F. & Philpot, W. (1998) made a replication of the study of Huguenin and Jones (1986) and found similar conclusions related to the adjacent peak. The Tsai experiments also confirmed that performance of this tool has a strong influence on the selected parameter of the derivative method. The authors also recommend that parameters of the derivative must be selected considering each particular set of spectra and purpose of the analysis.

An estimation method for fatty acid composition in oil using NIR was also developed by Sato (2002). In this study spectra of pure fatty acid were obtained, second derivative was calculated to identify the corresponding peak for C16:0, C18:0, C18:1, C18:2, C18:3, and C22:1. An iterative process was developed examining the moving average (MA), the size of the derivative segments (SEG,) and the gap between derivative segments (GAP). The band for C18:3, C18;2, C18:1, C18:0, C16:0, and C22:1 was listed in the second derivative. The absorption bands were identified at 1708, 1712, 1724, 1730, 1728, and 1726 nm, respectively, when the MA= 4 nm, the SEG=12nm, and Gap=12 nm. The method was validated using the fatty acid profile of a known sample of rapeseed oil. Sato concludes that with this method it is possible to estimate the fatty acid profile roughly, simply, and rapidly.

Derivative spectra can be expressed by considering the derivative order, gaps between points to calculate the difference, and the number of data points used to pretreat the data. A first derivative of a spectrum is a curve that contains a peak and a valley, which correspond to the inflection point of the original spectra. It is not frequently used because the peak and the valley of the first derivative do not follow the pattern of the raw spectra Shenk et al. (2007). Second order derivative spectra are commonly used because the band intensity and peak location are maintained as in the original spectra. The third order derivative of a spectrum is not frequently used because it presents the same limitations of the first derivative, complicating the situation due to the fact that more peaks are present. The fourth order derivative is promising in spectra interpretation because with the correct set of the gap and number of data points, several peaks can be observed which display similar characteristics to the second order derivative.

The revised work of derivative technique suggests this method could be useful for extracting information from combination bands on the NIR spectra of biodiesel.

2.5 References

ASTM Standards. D7371-07. 2007. Standard test method for determination of biodiesel (Fatty acid methyl ester) content in diesel fuel oil using mid infrared spectroscopy (FTIR-ATR-PLS method), West Conshohocken, PA.: ASTM.

Baptista, P., Felizardo, P., Menezes, J. C., & Correia, M. J. N. (2008). Multivariate near infrared spectroscopy models for predicting the methyl esters content in biodiesel. Analytica Chimica Acta, 607(2), 153-159.

Bondioli, P., & Della Bella, L. (2005). An alternative spectrophotometric method for the determination of free glycerol in biodiesel. European Journal of Lipid Science and Technology, 107(3), 153-157.

Felizardo, P., Baptista, P., Uva, M. S., Menezes, J. C., & Correia, M. J. N. (2007). Monitoring biodiesel fuel quality by near infrared spectroscopy. Journal of Near Infrared Spectroscopy, 15(2), 97-105.

Gaydou, V., Kister, J., & Dupuy, N. (2011). Evaluation of multiblock NIR/MIR PLS predictive models to detect adulteration of diesel/biodiesel blends by vegetal oil. Chemometrics and Intelligent Laboratory Systems, 106(2), 190-197.

Gonzalez-Martin, I., Gonzalez-Perez, C., Hernandez-Mendez, J., & Alvarez-Garcia, N. (2003). Determination of fatty acids in the subcutaneous fat of iberian breed swine by near infrared spectroscopy (NIRS) with a fibre-optic probe. Meat Science, 65(2), 713-719.

Hugenin, R., & Jones, J. (1986). Intelligent information extraction from reflectance spectra - absorption-band positions. Journal of Geophysical Research-Solid Earth and Planets, 91(B9), 9585-9598.

Knothe, G. 2001. Determining the blend level of mixtures of biodiesel with conventional diesel fuel by fiber-optic near-infrared spectroscopy and 1H nuclear magnetic resonance spectroscopy. J. American Oil Chemists Soc. 78(10): 1025-1028.

Morrey, J. R. (1968). On determining spectral peak positions from composite spectra with a digital computer. Analytical Chemistry, 40(6), 905.

Oliveira, F., Brandao, C., Ramalho, H., da Costa, L., Suarez, P., & Rubim, 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(2), 194.

Pacheco Filho, J. G., M. Pimentel, L. Stragevitch, L. S. G. Teixeira, G. M. G. Ribeiro, and R. S. Cruz. 2006. Determination of the blend level of mixtures of biodiesel with mineral diesel fuel using near infrared spectroscopy. Bioenergy I. Tomar, Portugal. Engineering International Conferences.

Pilar Dorado, M., Pinzi, S., de Haro, A., Font, R., & Garcia Olmo, J. (2011). Visible and NIR spectroscopy to assess biodiesel quality: Determination of alcohol and glycerol traces. Fuel, 90(6), 2321-2325.

Sato, T. (2002). New estimation method for fatty acid composition in oil using near infrared spectroscopy. Bioscience Biotechnology and Biochemistry, 66(12), 2543-2548.

Shenk, J. S., Workman, J. J., Jr., Westerhaus, M. O. (2008) Application of NIR Spectroscopy to Agricultural Products. In D. Burns and E. Ciurczak (Eds.), Handbook of Near-Infrared Analysis (pp.347-386). Boca Raton, FL: Taylor &Francis Group.

Soares, I. P., Rezende, T. F., Silva, R. C., Castro, E. V. R., & Fortes, I. C. P. (2008). Multivariate calibration by variable selection for blends of raw soybean oil/biodiesel from different sources using fourier transform infrared spectroscopy (FTIR) spectra data. Energy & Fuels, 22(3), 2079-2083.

Soyeurt, H., Dardenne, P., Dehareng, F., Lognay, G., Veselko, D., Marlier, M., et al. (2006). Estimating fatty acid content in cow milk using mid-infrared spectrometry. Journal of Dairy Science, 89(9), 3690-3695.

Tsai, F. (1998). Derivative analysis of hyperspectral data. Remote Sensing of Environment, 66(1), 41.

Zawadzki, A. (2009). Biodiesel feedstock and blend level sensing using visible light spectra and neural network. Transactions of the ASABE, 52(2), 539.

Chapter 3 - Predicting the Concentration and Specific Gravity of Biodiesel-Diesel Blends Using Near-Infrared Spectroscopy

3.1 Abstract

Biodiesel made from different source materials usually has different physical and chemical properties and the concentration of biodiesel in biodiesel-diesel blends varies from pump to pump and from user to user; all these factors have significant effects on performance and efficiency of engines fueled with biodiesel. To address these challenges, regressions based on near-infrared spectroscopy were developed for relatively inexpensive and rapid on-line measurement of the concentration and specific gravity of biodiesel-diesel blends. Methyl esters of five different oils—soybean oil, canola oil, palm oil, waste cooking oil, and coconut oil—and two different brands of commercial-grade No. 2 on-highway diesel and one brand of off-road No. 2 diesel were used in the calibration and validation processes. The predicted concentration and specific gravity of the biodiesel-diesel blends were compared with the actual values. The maximum and average root-mean-square errors of prediction (RMSEP) of biodiesel concentration were 5.2% and 2.9%, respectively, from the biodiesel type-specific regression. For the general regression, the RMSEP were 3.2% and 0.2% for biodiesel concentration and specific gravity predictions, respectively.

3.2 Introduction

Biodiesel is a fuel composed of mono-alkyl esters of long-chain fatty acids derived from vegetable oils or animal fats. It is renewable, oxygenated, essentially sulfur-free, and biodegradable. Biodiesel is also the only alternative fuel that has passed the U.S. EPA required Tier I and Tier II health effects testing requirements of the Clean Air Act amendments of 1990 (Tyson, 2004). In the United States, biodiesel has been used mainly as 2% to 20% blends with petroleum diesel.

Elevated NO_x emissions have been considered as one of the major problems of biodiesel and biodiesel blends as compared to petroleum diesel in diesel engines (Choi and Reitz, 1999; Sharp et al., 2000; McCormick et al., 2001; Grimaldi et al., 2002; Hansen et al., 2006). Earlier

combustion that causes more rapid cylinder pressure rise and higher combustion temperature was believed to be one of the main causes (Tat and Van Gerpen, 2003; Yuan et al., 2005; Yuan et al., 2007). This suggests that NO_x emissions could be reduced by retarding the combustion timing of the fuel in diesel engines, which can be achieved by adjusting injection timing according to the concentration of biodiesel in petroleum diesel. Therefore, a means to detect the concentration of biodiesel in its diesel blends will be necessary.

Another problem of biodiesel is that biodiesel fuels made from different source oils usually have different physical and chemical properties (e.g., specific gravity and cetane number), which makes it difficult for engine manufacturers to optimize engine performance when biodiesel is used. Therefore, it is important that the means is able to determine the properties of any type of biodiesel fuel. Near-infrared (NIR) spectroscopy meets this requirement and also is suitable for relatively inexpensive and rapid on-line measurement. Although successful applications of NIR spectroscopy on predicting the oil fraction and some operating properties of diesel fuel (Sikora and Salacki, 1996) and on determining the concentration of a specific type of biodiesel in diesel fuel (Knothe, 2001; Pacheco et al., 2006) have been reported, at present, efforts to determine both the concentration and properties at the same time of various types of biodiesel fuels using NIR spectroscopy are limited. The objectives of this study were to develop (1) a regression for determining the concentration of biodiesel in biodiesel-diesel blends and (2) a regression for estimating the specific gravity of biodiesel-diesel blends.

3.3 Materials and Methods

3.3.1 Fuel samples

Biodiesel derived from five different oils were used in this study - soybean oil methyl ester (SME), canola oil methyl ester (CME), coconut oil methyl ester (CCME), waste cooking oil methyl ester (WCME), and palm oil methyl ester (PME). The food-grade soybean oil and canola oil were purchased from local grocery stores. The virgin coconut oil and palm oil were obtained from Tropical Traditions, Inc. (Springville, Calif.). The waste cooking oil was collected from a local restaurant. All biodiesel samples were freshly made through a standard base-catalyzed transesterification process followed by repeated water-wash and drying. The fatty acid profiles of the biodiesel fuels are shown in Table 3-1. The five biodiesel fuels chosen cover a wide range of fatty acids; CCME is rich in short-chain saturated fatty acids (C8:0 to C14:0), and PME is

abundant in C16:0, whereas SME, CME, and WCME are rich in long-chain unsaturated fatty acids such as C18:1 and C18:2 and even some C18:3. These are the major fatty acids present in natural oils. Therefore, the biodiesel samples we chose can represent a general type of biodiesel.

Table 3-1 Relative weight composition of fatty acid methyl ester of the biodiesel samples.

	SME[a]	CCME[b]	PME[b]	CME[b]	WCME[a]
C8:0	0.0002	0.092	0	0	0
C10:0	0	0.064	0	0	0
C12:0	0	0.487	0	0	0
C14:0	0.0008	0.170	0	0	0.008
C16:0	0.1049	0.077	0.406	0.042	0.222
C16:1	0.0012	0	0	0	0.004
C18:0	0.0427	0.022	0.051	0.017	0.042
C18:1	0.2420	0.054	0.428	0.568	0.542
C18:2	0.5136	0.022	0.110	0.217	0.133
C18:3	0.0748	0	0.005	0.157	0.008
C20:0	0.0036	0	0	0	0.012
C20:1	0.0028	0	0	0	0
C22:0	0.0040	0	0	0	0
C22:1	0.0007	0	0	0	0
C24:0	0.0014	0	0	0	0

[a] Analyzed by the Kansas Lipidomics Research Center at Kansas State University (Manhattan, Kans.).

[b] Analyzed by American Analytical Chemistry Laboratories (Champaign, Ill.).

Three commercial-grade No. 2 diesel fuels, a highway Phillips diesel (D2HWP), a highway Cenex diesel (D2HWC), and an off-road Cenex diesel (D2ORC), were used to blend with each biodiesel fuel to prepare the 90 calibration samples. The volume-based concentration of biodiesel in these blends ranged from 0% up to 100% at steps of 20%.

The validation set consisted of 15 randomly coupled biodiesel-diesel blends from the same three diesel and five biodiesel fuels used in the calibration process. These samples covered 5% up to 95% at steps of 5% in the blends without replications of 20%, 40%, 60%, and 80% blends used in the calibration. The validation samples are shown in Table 3-2.

Table 3-2 Validation samples.

Biodiesel	Diesel	Biodiesel Concentration (%)
CME	D2ORC	5
PME	D2HWP	10
WCME	D2HWC	15
CCME	D2HWC	25
SME	D2ORC	30
CME	D2HWP	35
WCME	D2ORC	45
WCME	D2HWC	50
CCME	D2HWC	55
SME	D2HWC	65
SME	D2HWP	70
CME	D2HWC	75
CCME	D2ORC	85
SME	D2HWP	90
SME	D2ORC	95

3.3.2 Specific gravity measurement

The specific gravities of the samples were measured at room temperature (22°C to 24°C) using a Fisherbrand hydrometer (size 0.795-0.910, accuracy 0.001, Thermo Fisher Scientific, Waltham, Mass.). The measurement was performed three times for each sample. The hydrometer was calibrated at the reference temperature of 60°F (15.56°C) by the manufacturer. Following

ASTM D1298-99e2 standard (2003), the observed hydrometer readings at temperatures other than the reference temperature were corrected to the reference temperature of 60°F (15.56°C) and converted to specific gravity by using the ASTM-IP D1250 petroleum measurement tables (1953).

3.3.3 NIR spectra collection

All the samples were scanned at room temperature (22°C to 24°C) on an NIR QualitySpec Pro spectrometer (ASD Inc., Boulder, Colo.). The spectrometer measures absorbance from 350 to 2500 nm using silicon and indium-gallium-arsenide sensors. A Micropack HL-2000 halogen light source (Micropack, Ostfildern, Germany) was used for illumination. The spectrometer was optimized, and a baseline was collected using RS3 software (Version 3.1, ASD Inc., Boulder, Colo.). The samples were placed in a Fisherbrand Suprasil 300 quartz cuvette (10-mm path length, Thermo Fisher Scientific, Waltham, Mass.), which was connected to the spectrometer through a multi-use fiberoptic fixture (ASD Inc., Boulder, Colo.). The spectrometer and the cuvette used in these experiments are shown in Figure 3-1. A fiber-optic probe was used to illuminate the cuvette and carry the transmitted energy to the spectrometer. Twenty spectra were collected for each sample, and the average spectrum was converted to ASCII format using ASD ViewSpecPro (ASD Inc., Boulder, CO).

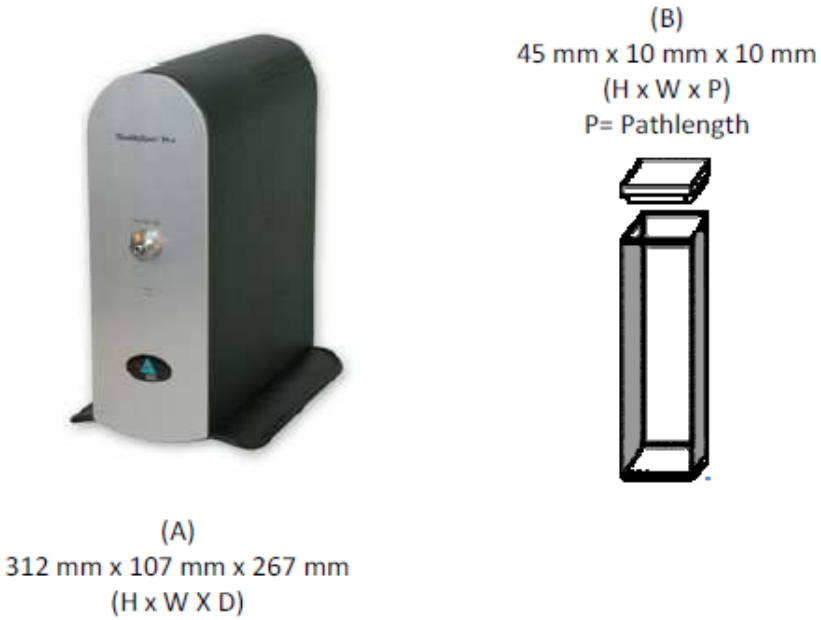

Figure 3-1 A-QualitySpec Pro spectrometer ASD, B- Quartz cuvette used in the experiments

3.3.4 Prediction method

The multi-linear regression (MLR) as shown by equation 1 was used for the prediction:

$$y = b_0 + b_1(A_{\lambda 1}) + b_2(A_{\lambda 2}) + b_3(A_{\lambda 3}) \quad (1)$$

Where y is biodiesel concentration or specific gravity; $A_{\lambda 1}$, $A_{\lambda 2}$, and $A_{\lambda 3}$ are absorbance values at wavelengths λ_1, λ_2, and λ_3, respectively. The coefficients b_0, b_1, b_2, and b_3 and the three best wavelengths λ_1, λ_2, and λ_3 were determined by the multiple linear regression method through the Sesame software version 3.1 using the calibration spectra. The fitness of the calibration scores to the regression line is represented by standard error of estimate (SEE) as shown in equation 2:

$$SEE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - y_{est})^2}{n_c}} \quad (2)$$

Where y_i and y_{est} are the actual and projected value of each calibration sample and n_c is size of the calibration samples. The three wavelengths were selected in the range of 2080 to 2200 nm by minimizing SEE through the Sesame software version 3.1.Using more than three wavelengths slightly improved the estimation (smaller SEE), however, computational times were significantly increased, and thus three wavelengths were used in this study. The regression was used to predict biodiesel concentration and specific gravity of biodiesel-diesel blends when the absorbance values of the fuel at three designated wavelengths (λ_1, λ_2, and λ_3) are known. The accuracy of predictions was measured by the root-mean-square error of predictions (RMSEP):

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n}(y_{pred} - y_i)^2}{n_v}} \quad (3)$$

Where y_{pred} is the predicted value of each validation sample from the regression equation, y_i is the actual value of the validation object, and n_v is the size of the validation samples. RPD value was also used to test the calibration models. RPD value is the ratio of the standard deviation (SD) of the reference data divided by standard error of estimate (SEE).

$$RPD = \frac{SD}{SEE} \tag{4}$$

3.4 Results and Discussion

Figure 3-2 shows the absorbance curves of SME and its blends with D2HWP in the wavelength range of 2080 to 2200 nm. At around 2145-nm wavelength, the 100% SME has the highest absorbance value, the D2HWP has the lowest, and the blends are intermediate. When the spectra for the other biodiesels and their blends with various diesel fuels in the range of 2080 to 2200 nm were plotted, the patterns of the curves were similar, although small variations in peak absorption intensity and related wavelength were observed. It is well known that the range of 2100 to 2200 nm is assigned to straight carbon chains and cis double bonds that reflect fatty acid moieties in fat molecules (Sato, 1994). Information about fatty acid compositions is demonstrated in this range through the in-saturation degree of the carbon chains. Therefore, the range of 2080 to 2200 nm was chosen as the range from which Sesame software would select the three best wavelengths. This range is also close to or in the middle of the NIR ranges used by some other researchers for similar purposes (Knothe 2001; Welch et al., 2006).

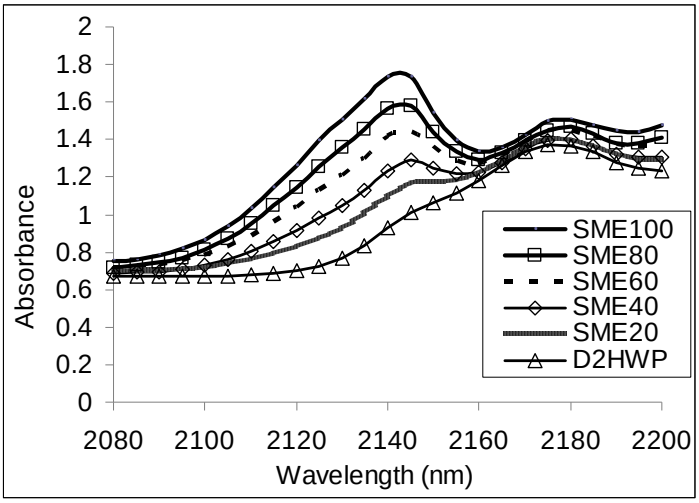

Figure 3-2 Spectra of soybean oil methyl ester and its blends with highway Philips diesel fuel**.**

3.4.1 Regressions for type-specific biodiesel

A calibration was developed for each type of biodiesel and its blends with the three types of diesel fuels. When the source oil of biodiesel is known, a biodiesel type-specific regression can be used to predict the concentration of biodiesel in the blends. The coefficients b_0, b_1, b_2, and b_3 and the three best wavelengths λ_1, λ_2, and λ_3 for each type of biodiesel are shown in Table 3-3. The negative coefficient indicates inverse proportional relation between the contributor and the predicted value. For the prediction of all five types of biodiesel, the multiple correlation coefficients (R^2) were greater than 0.999, and SEE were smaller than 1.2%. The concentrations of biodiesel in the validation samples were determined by using the regressions developed. The RMSEP and RPD values are shown in Table 3-4. The maximum RMSEP was 5.2% for BCA, the average RMSEP was 2.9%, and the minimum RPD value was 13.34, indicating that the regression was reasonably accurate in predicting biodiesel concentration. Williams (2001) suggested that RPD from 5 to 6.4 is suitable for quality control application, while a RPD of 8 or higher is excellent and the calibration can be used for any application.

Table 3-3 Regression for a specific type of biodiesel and its blends with the three diesel fuels.

Type of Biodiesel	b_0, b_1, b_2, and b_3	λ_1, λ_2, and λ_3	R^2	RPD	SEE	RMSEP
CCME	-0.08985, -2.616130, 3.674808, -0.800008	2120, 2129, 2150	0.999	20.95	0.008	0.015
CME	0.01967, -0.586301, 2.953473, -2.272920	2103, 2141, 2150	0.999	13.34	0.012	0.052
SME	-0.09132, -5.282706, 6.249927, -0.729645	2115, 2123, 2145	0.999	31.42	0.006	0.026
WCME	0.10454, -1.926984, 2.859055, -0.939792	2105, 2129, 2147	0.999	41.90	0.009	0.037
PME	-0.11840, -2.962160, 3.997019, -0.739531	2115, 2127, 2150	0.999	27.93	0.006	0.017
					Average	0.029

3.4.2 Regression for a general type of biodiesel

When the biodiesel type is unknown, the type-specific regression cannot be used to predict blending levels; a general regression is needed. All 90 calibration samples were used to determine the regression coefficients using the MLR method. The regression coefficients b_0, b_1, b_2, and b_3 and the three best wavelengths λ_1, λ_2, and λ_3 are shown in Table 3.4. The R^2 and SEE of the regression were 0.997% and 2.2%, respectively.

Table 3-4 Regression coefficients for predicting the concentration of a general type of biodiesel in its diesel blends.

b_0, b_1,b_2, and b_3	λ_1, λ_2, and λ_3
-0.01303, -2.340221, 2.929997, -0.482668	2100, 2122, 2146

The 15 validation samples were used to test the general regression. The RMSEP and RPD value were 3.2% and 11.42, respectively, which indicates an accurate prediction. Figure 3-3 shows the predicted biodiesel concentrations compared with the actual values. The maximum absolute prediction error (the difference between predicted and actual biodiesel concentration) was 7.5%, which was found on the sample of 75% CME blended with D2HWC. The average absolute prediction error of concentration was 2.6%. Figure 3-3 seems to indicate that the regression is tending to underestimate biodiesel concentration at higher concentration levels. This is not true but because CME and SME were randomly selected as the validation samples at the higher concentration levels (75%, 90%, and 95%). CME and SME happened to have the highest absorbance values among all the biodiesel fuels at the three selected wavelengths,

therefore, when the absorbance values were "averaged" in the regression, they were most under-predicted as shown in Figure 3-3.

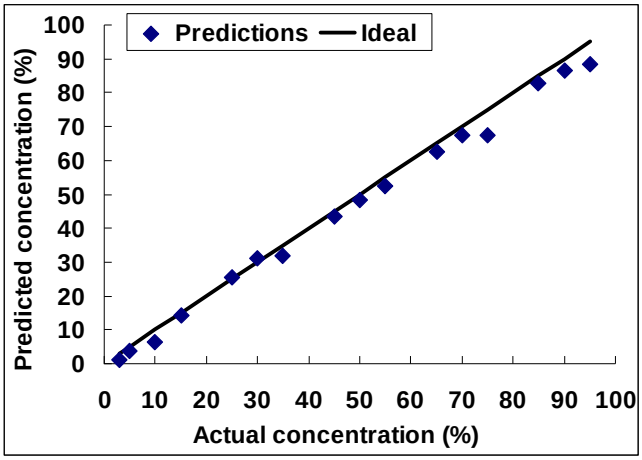

Figure 3-3 Predicted vs. actual biodiesel concentrations in biodiesel-diesel blends using the general regression.

3.4.3 Specific gravity prediction regression

Using the same spectra as in the study of biodiesel concentration, the specific gravities of biodiesel-diesel blends were predicted using the MLR method. The regression coefficients b0, b1, b2, and b3 and the three best wavelengths λ_1, λ_1, and λ_3 are shown in Table 3-5. The R^2 and SEE of the regression were 0.992 and 0.016, respectively.

Table 3-5 Regression coefficients for predicting the specific gravity of general type of biodiesel-diesel blends.

b_0, b_1,b_2, and b_3	λ_1, λ_2, and λ_3
0.82294, -0.049188, -0.073550, 0.128086	2100, 2121, 2130

The same set of validation samples were used to test the regression, and the RMSEP was 0.002. Figure 3-4 shows the predicted specific gravities compared with the actual values. The maximum absolute prediction error (the difference between predicted and actual specific gravity of the blends) was 0.005, and the average absolute prediction error of specific gravity was 0.002.

The over-predicted points in Figure 3-4 are not important enough to indicate that the regression over-predicts specific gravity, considering that all prediction errors were very small (<0.6%).

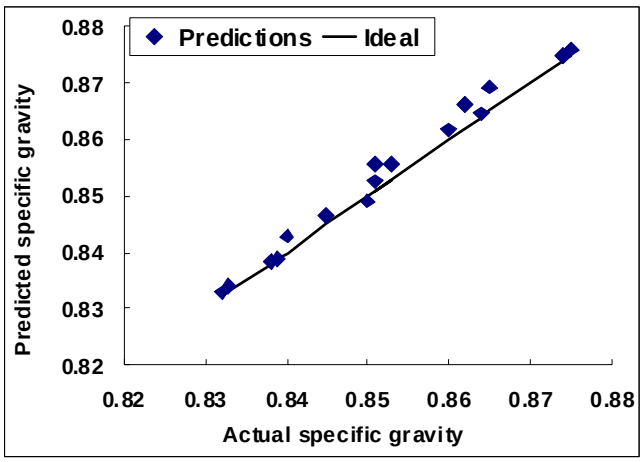

Figure 3-4 Predicted vs. actual specific gravities of biodiesel-diesel blends using the general regression.

Using the same calibration set but coupled with the partial least square regression (PLSR) method, regressions for predicting biodiesel concentration and specific gravity were also developed. The RMSEP was 0.026 and 0.002 for biodiesel concentration and specific gravity regressions, respectively. Compared with the RMSEP of 0.032 and 0.002 using the MLR method for concentration and specific gravity, respectively, the improvement in predictions was slight, and the computation time was significantly longer.

Although only five biodiesel and three diesel fuels were used in development and validation of the regressions, we expect that the regressions could be applied to other types of biodiesel and diesel fuels because the biodiesel and diesel fuels used are representative of a general biodiesel and diesel fuel. By using the regressions developed from this study, users may use this method for other different biodiesel-diesel blends. They can simply scan the samples at the designated wavelengths to obtain the absorbance values and use the coefficients provided to calculate the biodiesel concentration and specific gravity of the blends. Such method can be utilized by biodiesel retailers/distributors to measure biodiesel concentration in the blends, and

by engine manufacturers to detect biodiesel concentration in the fuel tank to adjust fuel injection timing.

3.5 Conclusions

Regressions based on NIR spectroscopy were developed for relatively inexpensive and rapid on-line measurement of the concentration and specific gravity of biodiesel-diesel blends. The NIR range of 2080 to 2200 nm was found suitable for the predictions regardless of biodiesel or diesel fuel type. The maximum and average RMSEP of biodiesel concentration in the blends were 5.2% and 2.9%, respectively, for the biodiesel type-specific regression. For the general regression, the RMSEP was 3.2%. The specific gravity prediction regression had an RMSEP of 0.002. The PLSR method was also used to develop the regressions; with this method, the improvement in predictions was slight and the computation time was significantly longer. The regressions developed can be used to predict the biodiesel concentration and specific gravity of biodiesel-diesel blends when the absorbance values at three designated wavelengths are known.

3.6 Acknowledgements

We thank Ms. Elizabeth Maghirang at the USDA ARS, Center for Grain and Animal Health Research, Engineering and Wind Erosion Research Unit, and Dr. David Wetzel, Department of Grain Science and Industry, Kansas State University, for assistance with the project. This work is supported by Kansas Agricultural Experiment Station (Contribution no. 09-110-J).

3.7 References

ASTM-IP Petroleum Measurement Tables. 1953. Prepared jointly by American Society for Testing Materials and the Institute of Petroleum. ASTM Designation: D1250, IP Designation: 200. West Conshohocken, Pa.: ASTM.

ASTM Standards. D1298-99e2. 2003. Standard test method for density, relative density (specific gravity), or API gravity of crude petroleum and liquid petroleum products by hydrometer method. West Conshohocken, Pa.: ASTM.

Choi, C. Y., and R. D. Reitz. 1999. A numerical analysis of the emissions characteristics of biodiesel blended fuels. Trans. ASME, J. Eng. for Gas Turbines and Power 121(1): 31-37.

Grimaldi, C. N., L. Postrioti, M. Battistoni, and F. Millo. 2002. Common rail HSDI diesel engine combustion and emissions with fossil/bio-derived fuel blends. SAE Tech. Paper 2002-01-6085. Warrendale, Pa.: SAE.

Hansen, A. C., M. R. Gratton, and W. Yuan. 2006. Diesel engine performance and NOx emissions from oxygenated biofuels and blends with eiesel. Trans. ASABE 49(3): 589-595.

Knothe, G. 2001. Determining the blend level of mixtures of biodiesel with conventional diesel fuel by fiber-optic near-infrared spectroscopy and 1H nuclear magnetic resonance spectroscopy. J. American Oil Chemists Soc. 78(10): 1025-1028.

McCormick, R. L., M. S. Graboski, A. M. Herring, and T. L. Alleman. 2001. Impact of biodiesel source material and chemical structure on emissions of criteria pollutants from a heavy-duty engine. Environ. Sci. & Tech. 35(9): 1742-1747.

Pacheco Filho, J. G., M. Pimentel, L. Stragevitch, L. S. G. Teixeira, G. M. G. Ribeiro, and R. S. Cruz. 2006. Determination of the blend level of mixtures of biodiesel with mineral diesel fuel using near infrared spectroscopy. Bioenergy I. Tomar, Portugal. Engineering International Conferences.

Sato, T. 1994. Application of principal-component analysis on near-infrared spectroscopic data of vegetable-oils for their classification. J. American Oil Chemist's Soc. 71(3): 293-298.

Sharp, C. A., S. A. Howell, and J. Jobe. 2000. The effect of biodiesel fuels on transient emissions from modern diesel engines: Part I. Regulated emissions and performance. SAE Paper No. 2000-01-1967. Warrendale, Pa.: SAE.

Sikora, Z., and W. Salacki. 1996. Use of near-infrared (NIR) spectroscopy to predict several physical and operating properties of oil fractions and diesel fuels. Petroleum and Coal 38(1): 65-68.

Tat, M., and J. H. Van Gerpen. 2003. Fuel property effects on biodiesel. ASAE Paper No. 036034. St. Joseph, Mich.: ASAE.

Tyson, K. S. 2004. Biodiesel handling and use guidelines. U.S. Department of Energy report DOE/GO-102994-1999.

Welch, W. T., R. R., Bledsoe Jr., and B. K. Wilt. 2006. Method and apparatus for analysis of relative levels of biodiesel in fuels by near-infrared spectroscopy. United patent application publication #US 2006/0213554 A1.

Williams, P. C. (2001) Chapter 8 Implementation of Near-Infrared Technology. In P. Williams, and K. Norris (Eds.), *Near- Infrared Technology* (pp.145-169). USA: American Association of Cereal Chemists, Inc.

Yuan, W., A. C. Hansen, M. E. Tat, J. H. Van Gerpen, and Z. Tan. 2005. Spray, ignition and combustion modeling of biodiesel fuels in a DI diesel engine. Trans. ASAE 48(3): 933-940.

Yuan, W., A. C. Hansen, and Q. Zhang. 2007. Computational modeling of NOx emissions from biodiesel combustion. Intl. J. Vehicle Design 45(1/2): 12-32.

Chapter 4 - Determining the Fatty Acid Profile of Biodiesel Fuels Using Fourier-Transformed Near- and Mid-infrared spectroscopy

4.1 Abstract

Biodiesel is an oxygenated, sulfur-free, biodegradable, non-toxic renewable fuel that can be derived from vegetable oils or animal fats. The quality and properties of biodiesel are directly related to their fatty acid compositions. The standard method of measuring fatty acid composition of biodiesel is gas chromatography. Even though this method is suitable and accurate, it requires a time-consuming procedure, well-trained personnel, and expensive reagents. Because of these reasons, near infrared (NIR) and mid infrared (MIR) spectroscopy were used for relatively inexpensive and rapid on-line measurement of the concentration of main fatty acid methyl esters present in biodiesel fuels, such as palmitic (C16:0), stearic (C18:0), oleic (C18:1), linoleic (C18:2), and linolenic (C18:3). Models were developed using four different biodiesels (MOD-4N/M) from palm, corn, canola, and flaxseed oil. After checking the accuracy of prediction of these models, using external samples, a second set of models were proposed and developed. In this case six biodiesel samples were used (MOD-6N/M), with biodiesel from animal fat and coconut being added to the previous set. The samples were scanned on a FTIR NIR/MIR Perkin Elmer spectrophotometer using transmittance for the NIR range and for the MIR range an ATR accessory was used. The partial least squares regression (PLSR) method was used to develop prediction models for each fatty acid methyl ester. Predicted concentrations of each methyl were compared with the actual values for each set in the model. The maximum root-mean-square error of prediction (RMSEP) was 1.59 % mass for the MIR range on the model for C18:2. Because of low performance of general models (MOD-6N/M) set for C16:0 and C18:0, additional work was developed to improve these results. Both NIR and MIR were found suitable for the prediction of concentration of C16:0, C18:0, C18:1, C18:2, and C18:3 in biodiesel samples.

4.2 Introduction

Biodiesel is a renewable alternative that can be used to replace significant amounts of diesel made from petroleum. It can be derived from vegetable oil or animal fat. In recent years new sources of oil or fats have been used to produce biodiesel. This large variety of raw

materials makes quality control of the fuel a challenging task. However, the fatty acid profile of fuel is the most important factor in the biodiesel properties evaluation. Consequently, a facility to determine the fatty acid profile in biodiesel will be an important tool in the quality control monitoring process. Although biodiesel has gained acceptance as a clean fuel, some technical problems have been reported with its use, such as an increase in the concentration of NOx exhaust emission and oxidative stability. The NOx level is mentioned in the literature as related to the cetane number and the cetane number is related to the fatty acid profile of the fuel, Knothe (2008). The oxidative stability of biodiesel has been related to the double bond present in the fatty acids, Falk, (2004). The five main fatty acids present in the biodiesel are palmitic (C16:0), stearic (C18:0), oleic (C18:1), linoleic (C18;2), and linolenic (C18:3). ASTM standard D 6751 has no direct regulations about the fatty acid profile of the fuel, but the European standard (EN14214) limits the concentration of linolenic acid methyl ester for biodiesel to 12 % (m/m). The most common method to determine a fatty acid profile in a substance is the gas-liquid chromatography. Even though it is reliable, this method presents some limitations. It requires well trained personnel to interpret the results, expensive reagents, and a time consuming procedure.

Infrared spectroscopy (IR) is successful when applied to determine the fatty acid profile of substances using both near infrared (NIR) and the mid infrared (MIR) regions, as can be seen in the following studies. I. Gonzales et al. (2003) and H. Soyeurt et al. (2006) Only one work that reported the use of NIR to predict methyl ester content in biodiesel was found, P. Baptista (2008).

Using NIR, Gonzales et al. (2003) presented a successful determination of the fatty acid profile in the fat of swine. Two different approaches were presented, one using extracted samples of sub-cutaneous fat and the other using the spectra obtained by direct application of a fiber-optic probe on samples of fat. The result showed standard error of calibration (SEC) from 0.09 % to 0.74% using extracted samples and from 0.09 % to 0.97% using direct application of fiber-optic probe case. The models were developed for C14:0, C16:0, C18:0, C18:1, C18:2, C18:3, and C20:1 fatty acid. The validation showed values of standard error of prediction (SEP) from 0.11% to 1.1 % in both cases concluding that the method presented comparable results to the reference method used. H. Soyeurt et al. (2006) developed a relatively successful estimation of fatty acid content in cow milk using MIR spectroscopy. The used regions were from 1736 to 1805 cm^{-1} and

between 2823 and 3016cm^{-1} the estimation of the efficiency of calibration was evaluated using the standard error of cross validation (SECV). The obtained SECV were from 0.01 to 0.18 g/dL of milk. The authors concluded that fatty acids profile in cow milk can be predicted using MIR spectroscopy when they are present in high concentrations; however the models presented poor performance when concentrations of fatty acids were low. P. Batista et al. (2008) developed a multivariate calibration to determine ester content in biodiesel samples using the near-infrared range between 9000 – 4500 cm^{-1}, and between 6102 – 5880 cm^{-1} in two different models. The root mean square errors of prediction (RMSEP) were 0.9 % (m/m) and 1.0 % (m/m), respectively. Additionally, the content of C18:3, C14:0, C16:0, C18:0, C18:1, and C18:2 were determined using the range between 9000 – 4500 cm^{-1}. The RMSEP were 0.18 % (m/m), 0.02 % (m/m), 0.79 % (m/m), 0.22 % (m/m), 1.79 % (m/m), and 2.5 % (m/m) respectively. The spectra were pretreated using the first order Savitsky-Golay derivative, and the models were developed using PLS methods. For this work, the validation was developed using samples from the same types of biodiesel used in the calibration (internal samples). Previous works show feasibility of NIR and MIR to predict a fatty acid profile in some substances and the viability of NIR to predict a fatty acid profile in biodiesel when internal samples were used. Considering that no report can be found on predicting fatty acid composition of biodiesel using MIR spectroscopy methods, and that the effectiveness of the NIR method has not been demonstrated when predicting the fatty acid profile on biodiesel, using external samples for validation, the objectives of this study were to develop predicting models for fatty acid composition of biodiesel using FTIR – NIR/MIR spectroscopy.

4.3 Materials and Methods

4.3.1 Samples preparation

The biodiesels used in this study were prepared from food grade canola oil, corn oil, flaxseed oil, animal fat, coconut oil, peanut oil, olive oil, and a mix of peanut, olive and soybean oil purchased from local grocery store, and palm oil purchased from Country Soap Shack (Missouri, USA). All biodiesel samples were freshly produced using a standard base-catalyzed transterification process followed by recurrent water washing and drying. The fatty acid profiles of the nine biodiesel fuels are shown in Table 4-1, which shows that the chosen biodiesel samples include a broad range of fatty acids. The palm methyl ester (PAME), and animal fat

methyl ester (AFME) present high content of C16:0, coconut methyl ester (CCME) is abundant in C12:0, canola methyl ester (CAME,) peanut methyl ester (PEME,) and olive methyl ester (OLME) are rich in C18:1, corn methyl ester (COME,) and flaxseed methyl ester (FXME) were selected for their high content of C18:2, and C18:3 respectively. To increase the variety of fatty acid content a mix of olive, peanut, and soybean oil was used to prepare mixed methyl ester (MXME).

Table 4-1 Fatty acid profile (% mass) of biodiesel samples prepared for this work.

FAME(m.%)	C8:0	C10:0	C12:0	C14:0	C16:0	C18:0	C18:1	C18:2	C18:3
COME	0.02	0.00	0.00	0.07	10.88	2.27	27.63	53.95	2.36
FXME	0.00	0.00	0.08	0.07	5.61	3.09	14.79	15.48	57.67
PAME	0.02	0.03	0.34	1.23	44.57	4.39	40.96	8.56	0.17
CAME	0.00	0.01	0.01	0.08	4.14	1.84	66.99	17.59	6.56
PEME	0.01	0.01	0.12	0.09	9.82	2.75	55.97	21.77	0.21
OLME	0.00	0.00	0.10	0.05	14.07	2.82	65.07	12.31	0.61
MXME	0.01	0.00	0.08	0.20	14.90	3.23	49.04	15.65	13.26
CCME	7.91	6.34	46.11	17.58	8.85	2.60	6.61	1.75	0.03
AFME	0.01	0.10	0.10	1.52	24.50	17.42	36.34	15.41	0.67

For the initial calibration set, combinations of four biodiesels were used (COME, CAME, FXME, and PAME.) The concentration (mass %) of methyl palmitate (C16:0), methyl stearate (C18:0), methyl oleate (C18:1), methyl linoleate (C18:2), and methyl linolenate (18:3) in the samples ranged from: 5.35 to 44.57, 1.92 to 4.39, 14.80 to 66.21, 8.56 to 52.21, and 0.18 to 57.67, respectively. The fatty acid profile of each sample was calculated using the fatty acid profile of four original biodiesel samples, obtained from gas chromatography (GC) analysis. A total of eighty samples were prepared.

Two sets of validation samples were prepared. The first set of eight samples were from the combination of the biodiesel used in the calibration (COME, CAME, FXME, PAME,) and the second set of eight samples were from the combination of biodiesel not used in the calibration (PEME, OLME, MXME, CCME, and AFME). The reason to use these two different sets was to verify the robustness of the models to predict the fatty acid profile of unknown biodiesel samples.

After analyzing the preliminary results, two additional biodiesels were used to increase the variability present in the calibration. In this case, combinations of six biodiesels were used, the four previously used plus AFME and CCME. These biodiesels were selected for their very different fatty acid profile when compared to the other four. AFME is rich in C18:0 and CCME is abundant in C12:0 and C14:0. The concentrations (% mass) of (C16:0), (C18:0), (C18:1),(C18:2), and (C18:3) in the samples were maintained in the same range as the previous experiments, but a significant increment in the sample numbers was obtained. Similarly to the previous case, the fatty acid profile of each sample was calculated using the fatty acid profile of six original biodiesel samples, obtained from gas chromatography (GC) analysis. A total of one hundred and thirty-seven samples were prepared. From this set, one hundred and twenty-six were used in the calibration and eleven samples were randomly selected to be used in the validation. To verify the robustness of the models under unknown biodiesel samples, the model was validated with a set of eight samples prepared from biodiesel not used in the calibration (PEME, OLME, MXME). The range of samples was kept within the calibration limit for each model.

4.3.2 Fatty acid profile measurement

All biodiesel samples were analyzed using gas chromatography in accordance with the following procedure: approximately 25 mg of biodiesel was dissolved in 4ml of benzene containing methyl-C13 internal standard. Samples were analyzed for fatty acid methyl esters using a HP 5890 GC with a FID detector and a SP-2560 capillary column (100m x .25mm x .2μ film, Supelco, Inc., Bellefonte, PA). Injection port and detector temperatures were 250^{o}C with a flow rate of 1 ml/min helium and a split ratio of 100:1. Injection volume was 1μl. Oven temperature began at 140^{o}C and increased at 2^{o}C/min to 200^{o}C then at 4^{o}C/min to 245^{o}C and held for seventeen minutes. All samples were scanned in The Ruminant Nutrition Lab, Department of Animal Science and Industry of Kansas State University.

4.3.3 FTIR- NIR/ATR spectroscopy scan

All the samples were scanned at a room temperature of 22-24^{o}C on a FT-IR/FT-NIR spectrometer (Perkin Elmer spectrum 400, Shelton, CT) for NIR and MIR ranges. The samples were placed for NIR (780 to 2500 nm) scan, in a quartz cuvette cell (Labomed Inc. Culver City, CA) of 5 mm pathlength for spectrophotometers. NIR spectra data were recorded as the absorbance in the wavelength range from 900 to 2500 nm at 1 cm^{-1} interval. The spectrometer and the cuvette used in these experiments are shown in Figure 4-1. For the case MIR (4000 to 400 cm^{-1}) scan, universal ATR accessory with Germanium (Ge) as crystal material (Perkin Elmer, Shelton, CT) was used. MIR spectra were recorded as the absorbance in the wavelength range from 4000 to 600 cm^{-1} at 1 cm^{-1} interval. All spectrums were recorded once for each sample, and were obtained as an average of thirty-two scans.

Figure 4-1 A- Perkin Elmer spectrometer, B- Quartz cuvette used in this experiments.

4.3.4 Calibration models

The regions on the NIR range used to estimate the concentration of C18:1, C18:2, and C18:3 in the biodiesel samples were located between 1600 and 1700 nm and between 2000 and 2200 nm. For the cases of C16:0, and C18:0 the used region was between 900 and 1400 nm. The

regions on the MIR range used to estimate the concentration of C18:1, C18:2, and C18:3 on biodiesel sample were located between 3028 and 2812 cm^{-1} and between 1598 and 793 cm^{-1}. For the cases of C16:0 and C18:0 models, the used region was between 2965 and 2807 cm^{-1} and between 1205 and 801 cm^{-1}. These regions were chosen based mainly on the absorbance band related to length of carbon chain, CH_2 and CH_3 ratio, and number of double bond (C=C) present in the sample, and optimized by Gram software version 6 using PLSR method. The performance of the models was assessed by the correlation coefficient (R^2), standard error of cross validation (SECV), the root mean square error of prediction (RMSEP), the absolute error of prediction (AEP), the average relative error (ARE), and RPD value.

4.4 Results and Discussion

4.4.1 Predicting using NIR spectroscopy

4.4.1.1 Near infrared spectra

It is well known that bands in the near infrared region are difficult to assign to specific compounds. This is because a single band in this region is the result of several possible combinations of fundamental bands and overtones. Nowadays, current advances in spectroscopy instruments conjugated with new chemometric software have made it possible to identify some chemical structures based on some characteristics of combination bands and overtones.

Figure 4-2 shows the ranges of NIR spectrum of biodiesel used in this study. The principal regions on the NIR range used to estimate the concentration of methyl oleate (C18:1), methyl linoleate (C18:2), and methyl linolenate (C18:3) in biodiesel samples were located between 1600 and 1700 nm and between 2000 and 2200 nm. Specifically, the band of 1620 nm is associated with C-H stretch first overtone with the $=CH_2$ structure. Additionally, the band of 2170 nm is assigned to the C-H stretch and C-H deformation combination with the HC=CH structure. J. Shenk et al. (2008) For the cases of methyl palmitate (C16:0) and methyl stearate (C18:0) the used region was between 900 and 1400 nm. This region is assigned to the C-H (CH_3, CH_2) stretching 2nd overtone, combination stretching band, and bending vibration. V. Sablinskas et al. (2003).

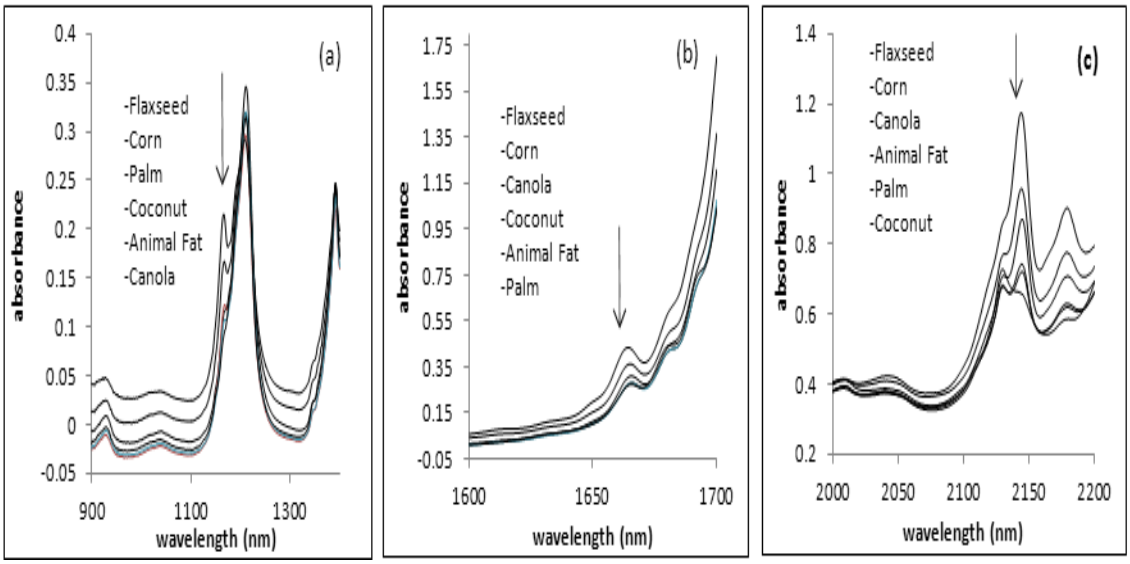

Figure 4-2 Typical spectra of biodiesel used in this study in the NIR range.

4.4.1.2 Predicting fatty acid composition, models using 4 types of biodiesel (MOD-4N).

The NIR spectroscopy data and Partial least squares regression (PLSR) methods were used for the development of a calibration model for each fatty acid methyl ester (FAME). PLSR is the technique most widely used in chemometric analysis. It is an extension of multiple regression analysis used to represent on a response variable the effects of linear combination of several predictors. PLSR is recommended when the number of predictor variables is higher than the numbers of observations and also when the predictors are highly correlated. L. Carrascal, (2009) Both situations are strongly present in our data.

The calibration results for MOD-4N are shown in Table 4-2. It is worth emphasizing that pre-treatment data was not used, nor were outliers detected.

Table 4-2 General results of calibration (MOD-4N) for the prediction of C16:0, C18:0, C18:1, C18:2, and C18:3 using PLSR method and NIR Spectra.

Type of Methyl Ester Models I	Multiple Correlation Coefficient (R^2) Calibration	Standard Error Cross Validation (SECV) % mass	Number of Factors	RPD Values
Palmitate	0.996	0.751	4	15.97
Stearate	0.995	0.041	4	14.52
Oleate	0.998	0.653	5	26.41
Linoleate	0.997	0.656	5	20.06
Linolenate	0.999	0.511	4	35.45

The maximum SECV was 0.656 % mass for methyl linoleate and the minimum SECV was 0.041 % mass for the methyl stearate model, all R^2 values were over 0.99. The minimum RPD value was 14.52 presented in the model for methyl stearate. Figure 4-3 presents the relationship between predicted and measured values for each fatty acid. The maximum absolute error was 1.73 % mass with an average relative error of 2.53% presented in the model for C18:2. The maximum RMSEP was 0.81 % mass presented in the same model for C18:2. The accuracy of these models is similar to those reported by P. Batista. (2008) Validation results, using samples from the combination of biodiesel used in calibration (internal samples), show there is an excellent agreement between fatty acid concentration of the samples analyzed by GC chromatography and the ones predicted using the data from NIR spectroscopy.

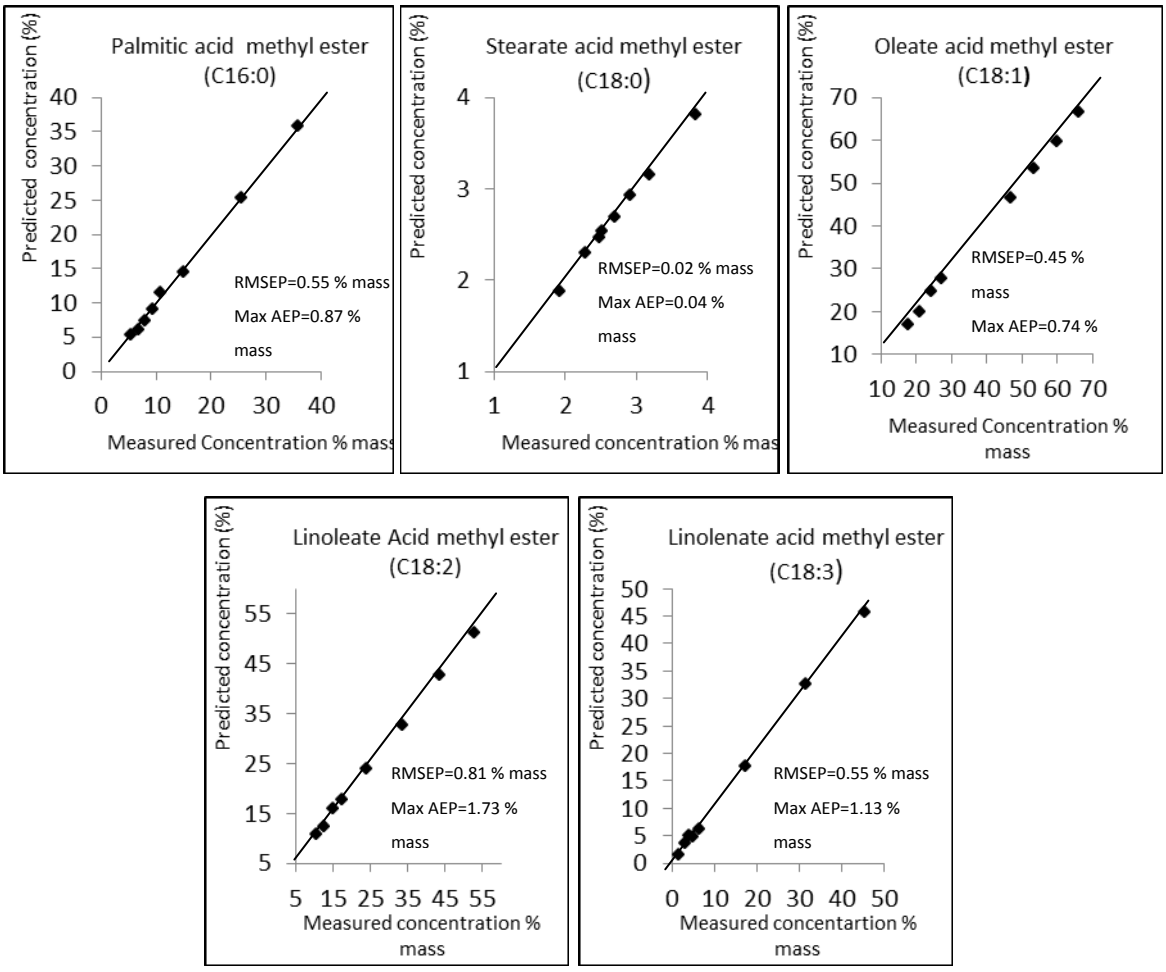

Figure 4-3 Results of validation set for each type of Methyl, using NIR spectroscopy models and internal samples MOD-4N.

The results presented above clearly indicate the calibration models developed to quantify concentration of the main fatty acid present in biodiesel from NIR spectra perform very well when the validation set is from the same type of biodiesel used in the calibration model; thus suggesting the fatty acid profile of biodiesel may be easily determined by utilizing NIR spectroscopy methods when the source of the biodiesel is known. This tool can be used as a fast and accurate method for monitoring biodiesel quality in specific production plants.

Regression coefficient for the calibration models are shown in Figure 4-4. Beta coefficients (Bc's) indicate that the calibration models are based on absorption bands related to biodiesel.

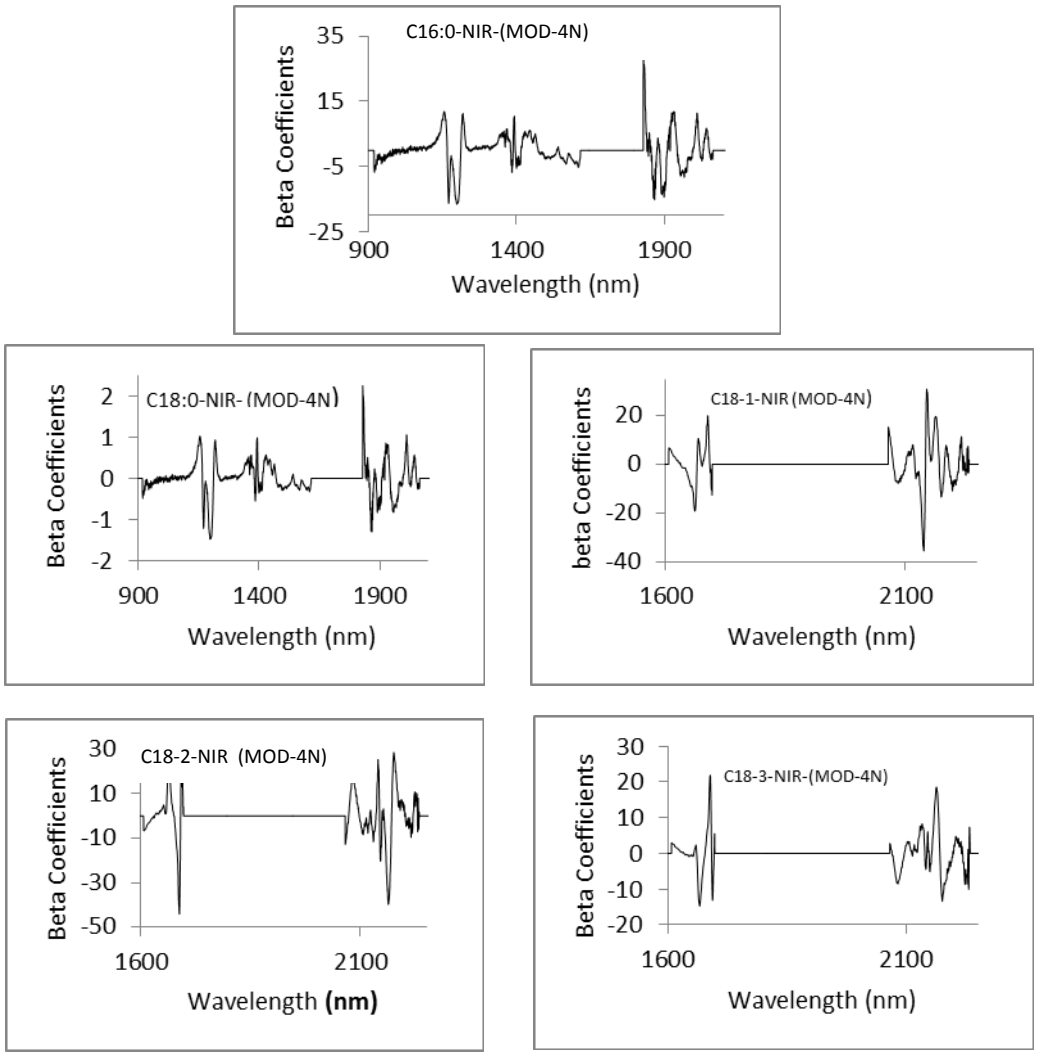

Figure 4-4 Regression coefficients for the (MOD-4N) models

The models for C16:0 and C18:0 show large positive and negative coefficients near 1190 nm, which are related to -CH_3 stretching 2nd overtone and near 1395 nm, which are related to the -CH combination band. The models for C18:1, C18:2, and C18:3 show positive and negative bands near 1695, which are related to the –CH stretch first overtone, near 2140 nm, which are related to the -CH/C=O stretching, combination or symmetric deformation and near 2170 nm, which are related to the –CH stretching combination or asymmetric deformation. The bands listed above all correspond to biodiesel absorption bands.

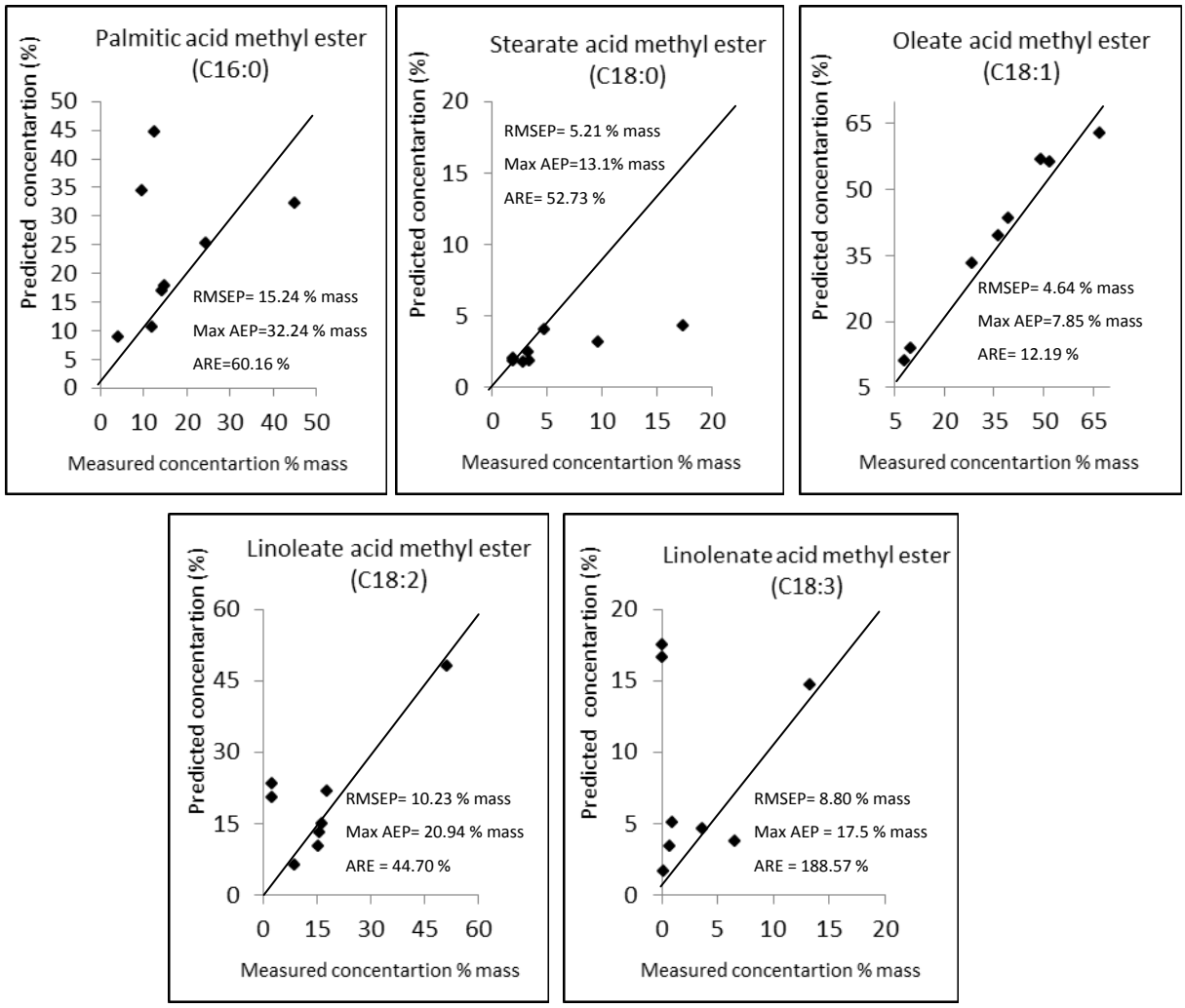

Figure 4-5 Results of validation set for each type of Methyl, using NIR spectroscopy models and external samples MOD-4N.

Figure 4-5 shows results when the validation set of external samples was used. The minimum absolute error was 0.04 % mass with an average relative error of 12.19 % presented in the model for C18:1. The minimum RMSEP was 4.64 % mass presented in the same model for C18:1. The validation results of the models, using the samples from combination of biodiesels not used in the calibration (external samples), were very disappointing.

Increasing the variability of the types of biodiesel present in the calibration could improve the performance of the models for any type of biodiesel. To prove this hypothesis, models using two additional biodiesels (AFME, and CCME) were proposed.

4.4.1.3 Predicting fatty acid composition, models using 6 types of biodiesel (MOD-6N).

Using the NIR spectroscopy data, models using combinations of six different biodiesels were developed to predict C16:0, C18:0, C18:1, C18:2, and C18:3 fatty acid. Table 4-3 shows the results for MOD-6N. In this case pre-treatment data was not used, but some outliers were detected in the model for methyl palmitate and methyl stearate. The samples used in the validation were from the combination of six biodiesels used for the calibration, plus some types of biodiesel that were not present in the calibration to evaluate the robustness of the model over any type of biodiesel.

Table 4-3 General results of calibration (MOD-6N) for the prediction of C16:0, C18:0, C18:1, C18:2, and C18:3 using PLSR method and NIR Spectra.

Type of Methyl Ester MOD-6N	Multiple Correlation Coefficient (R^2) Calibration	Standard Error Cross Validation (SECV) % mass	Number of Factors for Calibration	RPD Values
Palmitate	0.901	3.22	5	3.18
Stearate	0.788	1.82	5	2.26
Oleate	0.996	0.965	5	16.71
Linoleate	0.994	0.713	5	17.54
Linolenate	0.998	0.611	5	27.34

The higher SECV were 3.22 % mass for methyl palmitate model, and the minimum correlation coefficient (R^2) was 0.79 detected in the model for methyl stearate. The RPD values for methyl palmitate and methyl stearate were 3.18 and 2.26, respectively. Williams (2001), suggest that RPD value from 0.0 to 2.3 is poor and it is not recommended for any application, a value from 3.1 to 4.9 is fair and it can be used for screening application, and a value of 8.1 or higher is excellent and it can be used for any application. According with these results the calibration models for C18:1, C18:2, and C18:3 can be used, but the models for C16:0 and C18:0 require additional work. The results for the validation set of internal sample are shown in Figure 4-6. The maximum absolute error was 5.48 % mass with an average relative error of 16.15 %

presented in the model for C16:0. The maximum RMSEP was 2.62 presented in the same model for C16:0.

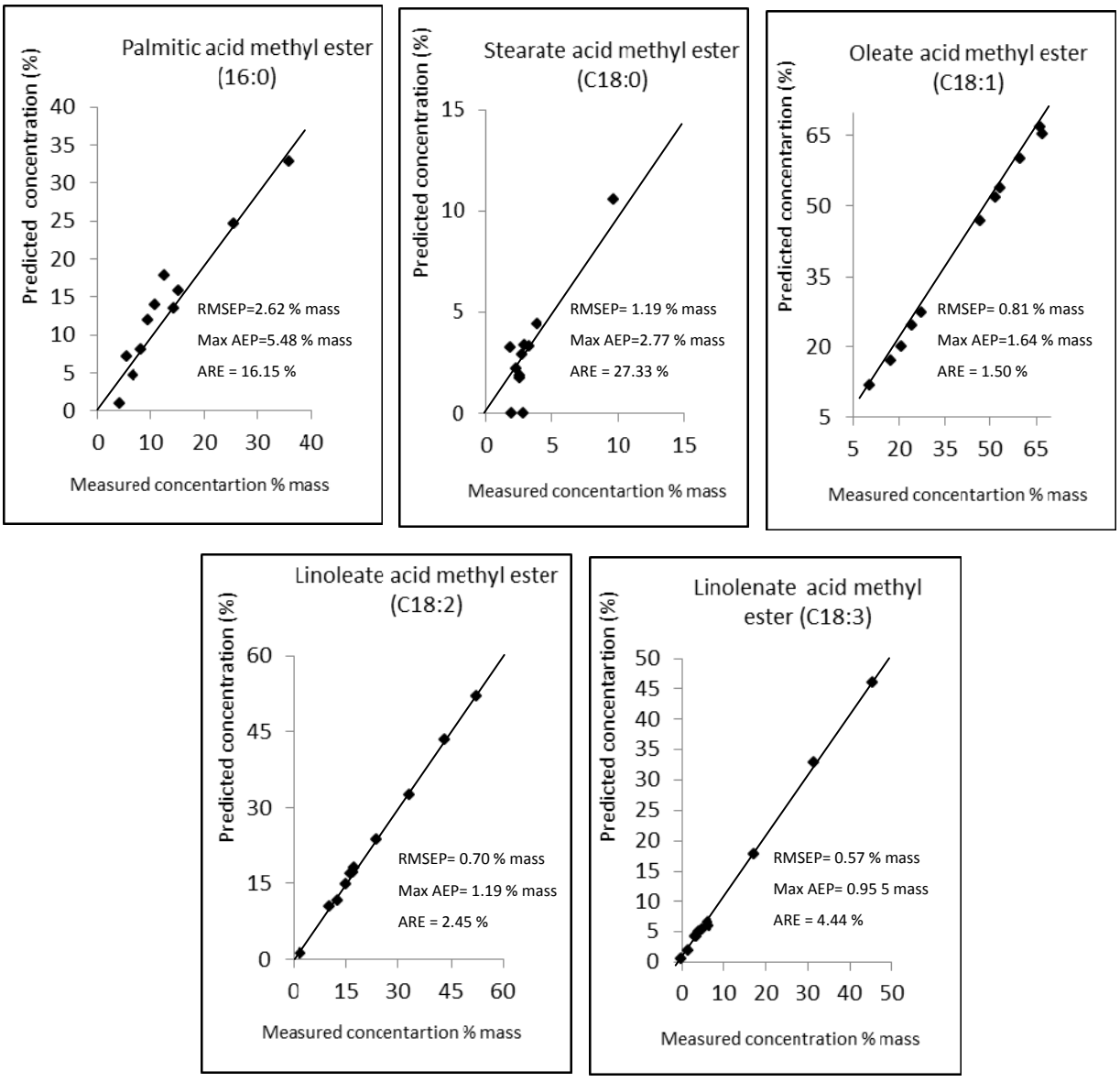

Figure 4-6 Results of validation set for each type of Methyl, using NIR spectroscopy model and internal samples MOD-6N.

Regression coefficients for the calibration models (MOD-6N) are shown in Figure 4-7. Similar behavior was found related to the regression coefficient between models MOD-4N and

MOD-6N, indicating that the calibration models are based on fatty acid methyl ester absorption bands.

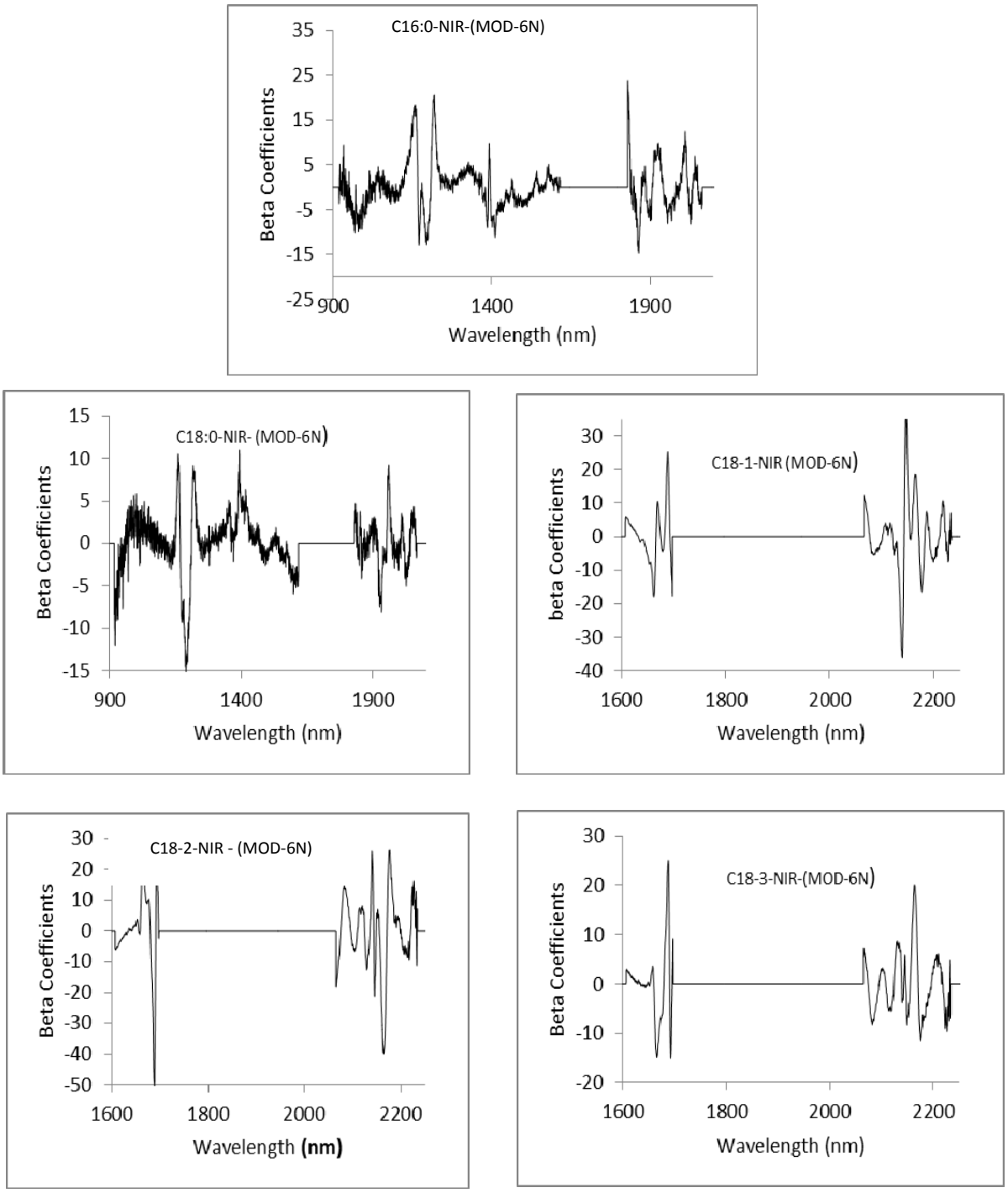

Figure 4-7 Regression coefficients for the MOD-6N models.

To compare the performance of the models MOD-6N and MOD-4N over the same condition, Figure 4-8 shows the results for validation set using external samples. The maximum absolute error was 8.52 % mass with an average relative error of 13.67 % presented in the model for C16:0. The maximum RMSEP was 3.41 presented in the same model for C16:0.

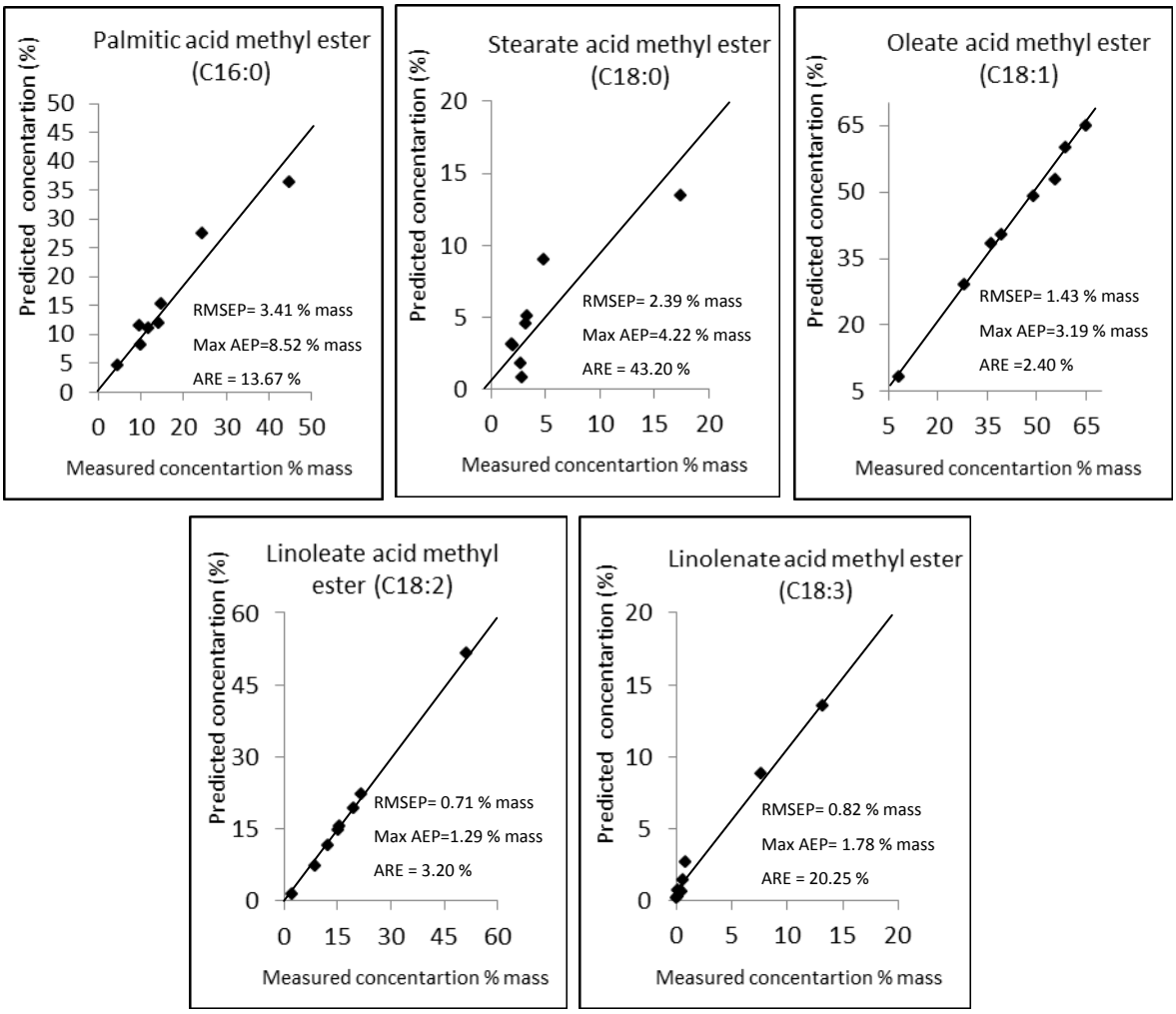

Figure 4-8 Results of validation set for each type of Methyl, using NIR spectroscopy model and external samples MOD-6N.

Considering that errors of the validation set using internal and external samples are comparable for the MOD-6N model, all statistical parameters used to verify the robustness of the

model were re-calculated using all validation samples available. Figure 4-9 shows the comparison between predicted and measured fatty acid concentration for each model, when a combination of internal and external sample sets were used. The maximum absolute error was 8.52 % mass with an average relative error of 14.97% presented in the model for C16:0. The maximum RMSEP was 2.22 % mass presented in the same model for C16:0. Validation results show there is good agreement between fatty acid concentration of the samples analyzed by GC chromatography and the ones predicted using the data from NIR spectroscopy , with respect to cases of methyl oleate (C18:1), methyl linoleate (C18:2), and methyl linolenate (C18:3); however, not for cases of methyl palmitate (C16:0), and methyl stearate(C18:0) models.

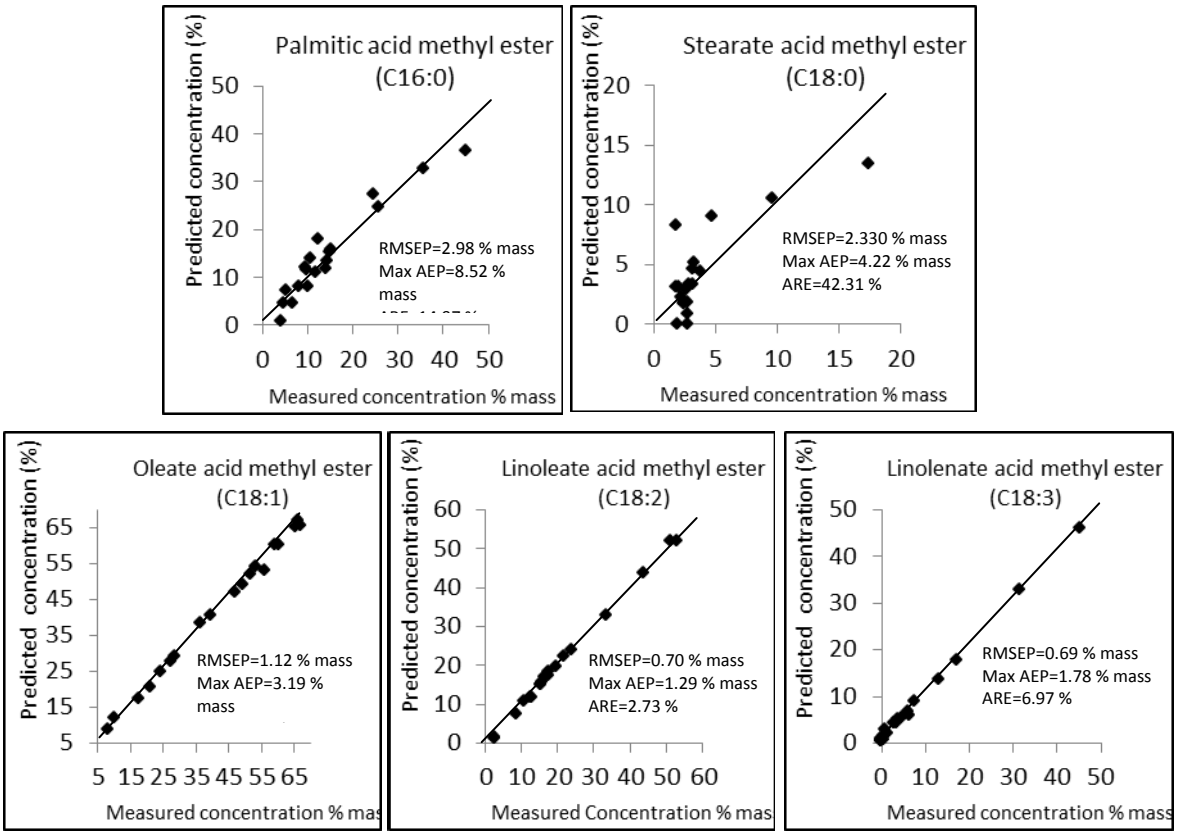

Figure 4-9 Results of validation set for each type of Methyl, using NIR spectroscopy model and all validation samples (internal and external) MOD-6N.

The results presented above indicate that the calibration models developed to quantify the concentration of methyl oleate, methyl linoleate, and methyl linolenate present in biodiesel from

NIR spectra perform well. However, models developed for methyl palmitate, and methyl stearate presented some deficiencies related to accuracy of predictions. These results suggested the fatty acid concentration of C18:1, C18:2, and C18:3 in biodiesel can be determined using the NIR spectroscopy method, with reasonable accuracy, even when the source of the biodiesel is unknown. Contrarily, models developed to predict C16:0 and C18:0 cannot be used to predict these fatty acids when the source of biodiesel is unknown. These models should be optimized to improve the accuracy of prediction. After analyzing the results of the models for C16:0 and C18:0, their behavior suggests that developing models for specific ranges, which are shown in Table 4-4, could improve their performance. In order to demonstrate this hypothesis three sub-models for specific ranges of concentration (low, medium, and high) were developed to predict the concentration of C16:0 and C18:0.

Table 4-4 Defined ranges for the sub-models low, medium, and high to determine C16:0, C18:0

Model For:	Low range of Concentration % mass	Medium range of concentration % mass	High range of concentration % mass
C16:0	0 - 12.99	13.00 - 24.99	25.00 – 45.00
C18:0	0 – 3.99	4.00 – 9.99	10.00 – 16.00

To select the correct sub-model, the general model is first used to predict the concentration of C16:0 and C18:0, and then the adequate model was chosen using this result. Figure 4-10 shows the results using the sub models.

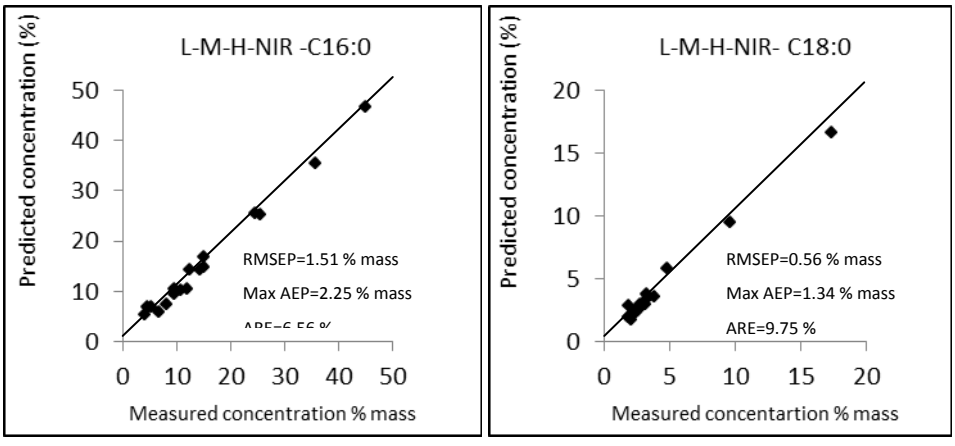

Figure 4-10 Results of validation set for methyl palmitate (16:0), and methyl stearate (18:0) using sub-models (lower, medium, and high) over NIR range.

The obtained results using the sub-models (low, medium, and high) improved the prediction accuracy of the models (MOD-6N) for methyl palmitate and methyl stearate. The maximum absolute error was 1.34 % mass with an average absolute error of 9.75 % mass presented in the model for C18:0. The RMSEP is reduced from 2.22 % mass to 1.51% mass for the case of C16:0 and from 2.30 % mass to 0.57 % mass for the case of C18:0. Considering the results of two step models to predict C16:0 and C18:0, these models can be used to predict the fatty acid present in biodiesel with reasonable accuracy, although the type of biodiesel is unknown.

4.4.2 Predicting using MIR spectroscopy

4.4.2.1 MID infrared spectra

It contrast with the near infrared (NIR) spectra, the mid infrared (MIR) spectra show a high degree of spectral resolution; consequently, a peak can be assigned to specific compounds except in the called fingerprint region (1400 – 900 cm^{-1}), where many chemical groups have frequencies. In other words, particular bands in this region can hardly be assigned to specific bonds. However, it is also well known that bands caused by specific complex compounds in this region are unique. Considering the previous advanced instruments and the new chemometric software it is now possible to identify complex chemical structures of our study.

Figure 4-11 shows the ranges of MIR spectrum of biodiesel used in this study. The main region on the MIR range used to estimate the concentration of C18:1, C18:2, and C18:3 in the biodiesel samples were located between 3028 and 2812 cm^{-1} and between 1598 and 793 cm^{-1}. Specifically, the band of 3025 cm^{-1} is associated with C-H stretching vibration with the =CH structure, and the band of 1418 cm^{-1} is assigned to the C-H bending vibration with the =CH structure; additionally, the bands of 914 and 968 cm^{-1} have been associated with -HC=CH- structure. B. Yaakob et al. (2010) For the cases of C16:0, C18:0 models the used region was between 2965 and 2807 cm^{-1} and between 1205 and 801 cm^{-1}. In these regions, the band 2953 cm^{-1} is assigned to the C-H asymmetric stretching, and the band 2853 cm^{-1} is assigned to the C-H symmetric stretching. Additionally, stretching and bending vibrations of CH_2 were noticed at 1163 cm^{-1} . B. Yaakob et al. (2010)

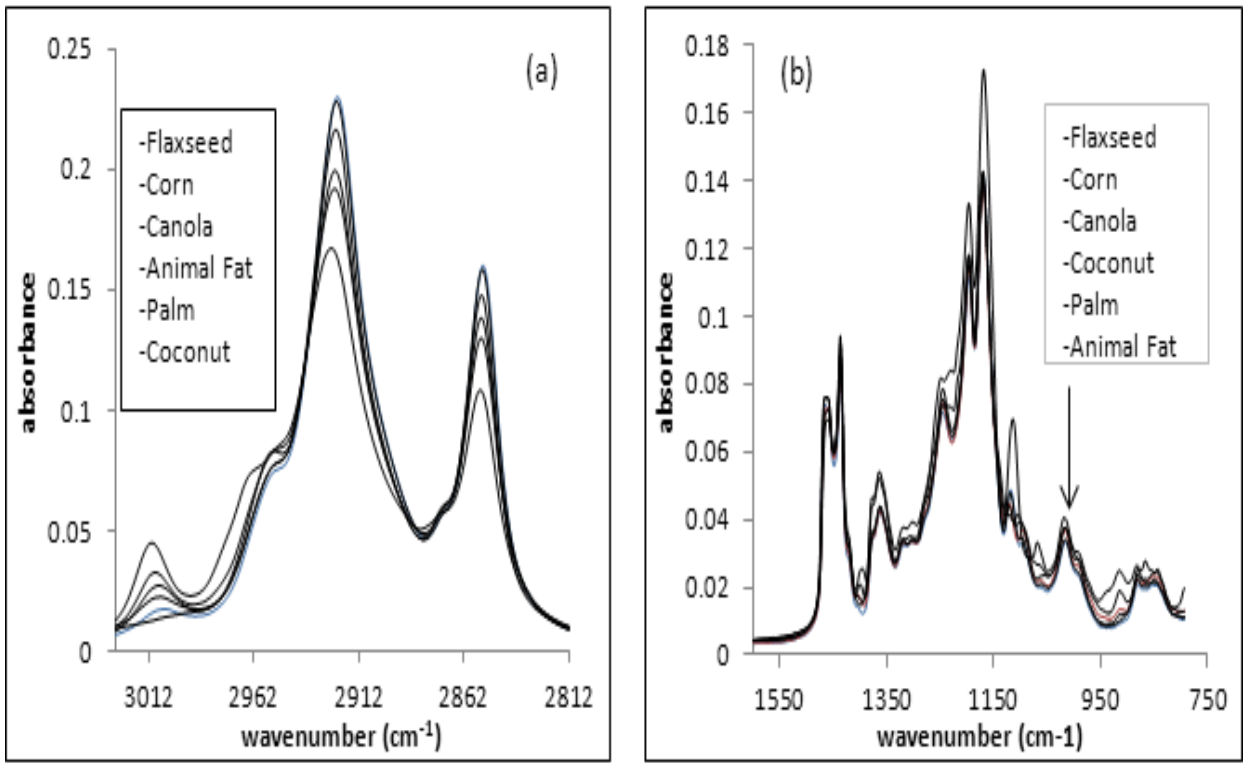

Figure 4-11 MIR spectra of biodiesel shown the used ranges in calibration

4.4.2.2 Predicting fatty acid composition, models using four types of biodiesel (MOD-4M).

The same eighty samples, from the combination of four biodiesels (COME, CAME, FXME, and PAME.) used in the MOD-4N for the NIR study, were used for the development of a calibration model for each fatty acid methyl ester in the MIR range (MOD-4M).

Table 4-5 shows the results for the models MOD-4M. In these models, pre-treatment data was not used, nor were outliers detected. Similarly to the NIR study, two set of samples were used in the validation, one set was from the combination of four biodiesels used on the calibration (internal samples,) and the other one was from combinations of biodiesels not used in the calibration (external samples). The following results are from the internal validation set.

Table 4-5 General results of calibration of MOD-4M for the prediction of C16:0, C18:0, C18:1, C18:2, and C18:3 using PLSR method and MIR Spectra.

Type of Methyl Ester Models I	Multiple Correlation Coefficient (R^2) Calibration	Standard Error Cross Validation (SECV) % mass	Number of Factor	RPD Values
Palmitate	0.996	0.729	4	16.45
Stearate	0.994	0.043	4	13.84
Oleate	0.997	0.921	5	18.72
Linoleate	0.998	0.544	5	24.20
Linolenate	0.998	0.773	5	23.44

The maximum SECV was 0.921 % mass for methyl oleate, and the minimum SECV was 0.043 % mass for the methyl stearate model, all R^2 values were over 0.99. The minimum RPD value was 13.84 in the models for methyl stearate. The maximum absolute error was 2.09 % mass with an average relative error of 2.94 % presented in the model for C18:3. The maximum RMSEP was 0.434 % mass presented in the model for C18:1. All errors were considered acceptable.

Regression coefficients for the calibration model (MOD-4M) are shown in Figure 4-12. The models for C16:0, C18:0, C18:1, C18:2, and C18:3 show positive and negative coefficients near 2854 cm^{-1} which are related to –CH symmetric stretching, near 2924 cm^{-1} , related to –CH asymmetric stretching, and near 1165 cm^{-1}, related to stretching and bending vibration. All bands listed above are related to fatty acid methyl ester absorption bands.

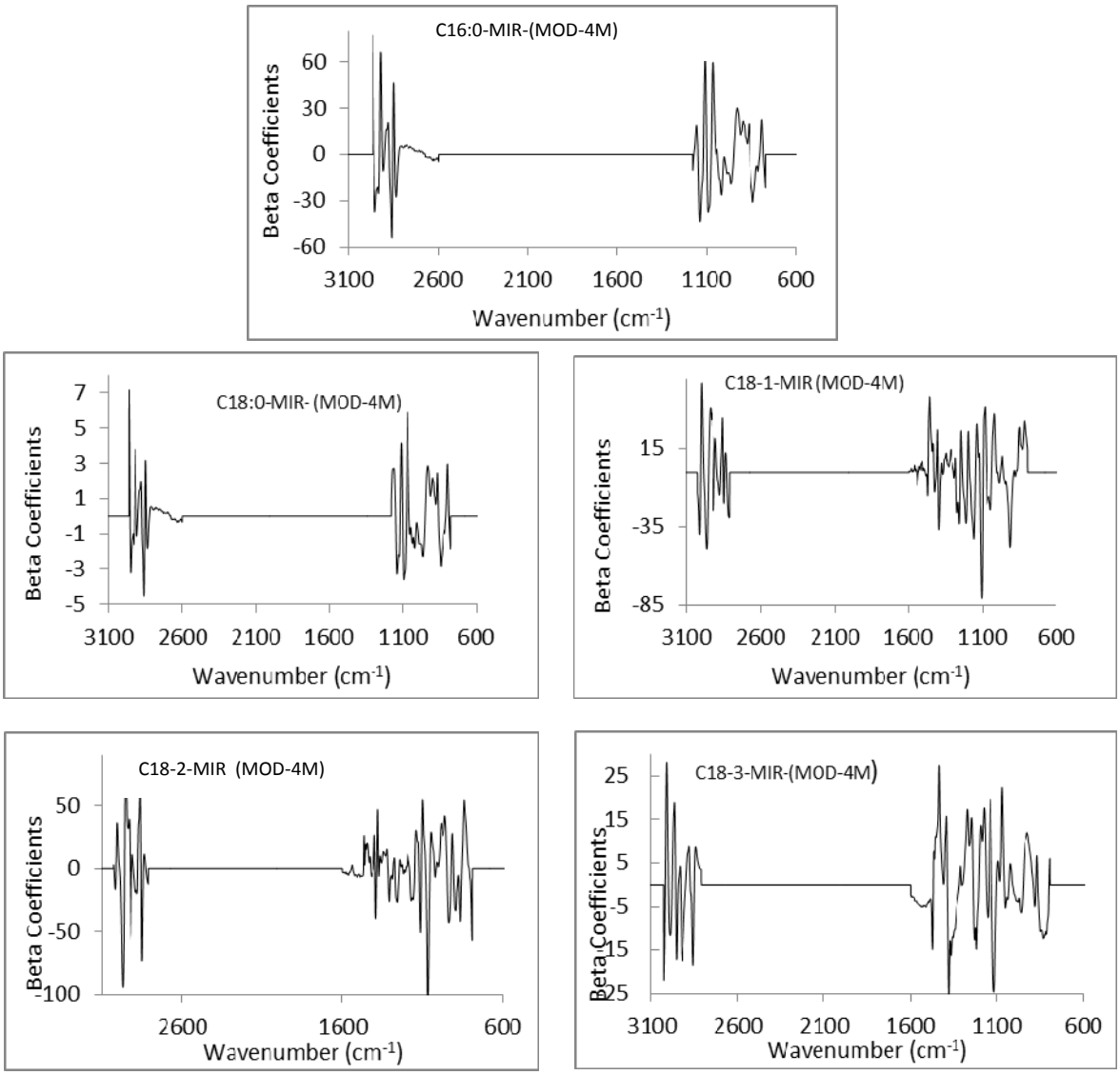

Figure 4-12 Regression coefficients for model MOD-4M

Figure 4-13 presents the relationship between predicted and measured values for each fatty acid when validation set of internal sample was used. The results when the validation set of external samples was used are shown in Figure 4-14. The minimum absolute error was 0.28 mass % with an average relative error of 20.96 % presented in the model for C18:1. The minimum RMSEP was 6.23 % mass presented in the model for C18:0. Validation results showed similar promising results with the NIR spectroscopy method when the internal validation set is used, and very disappointing results are given when the external validation set is used.

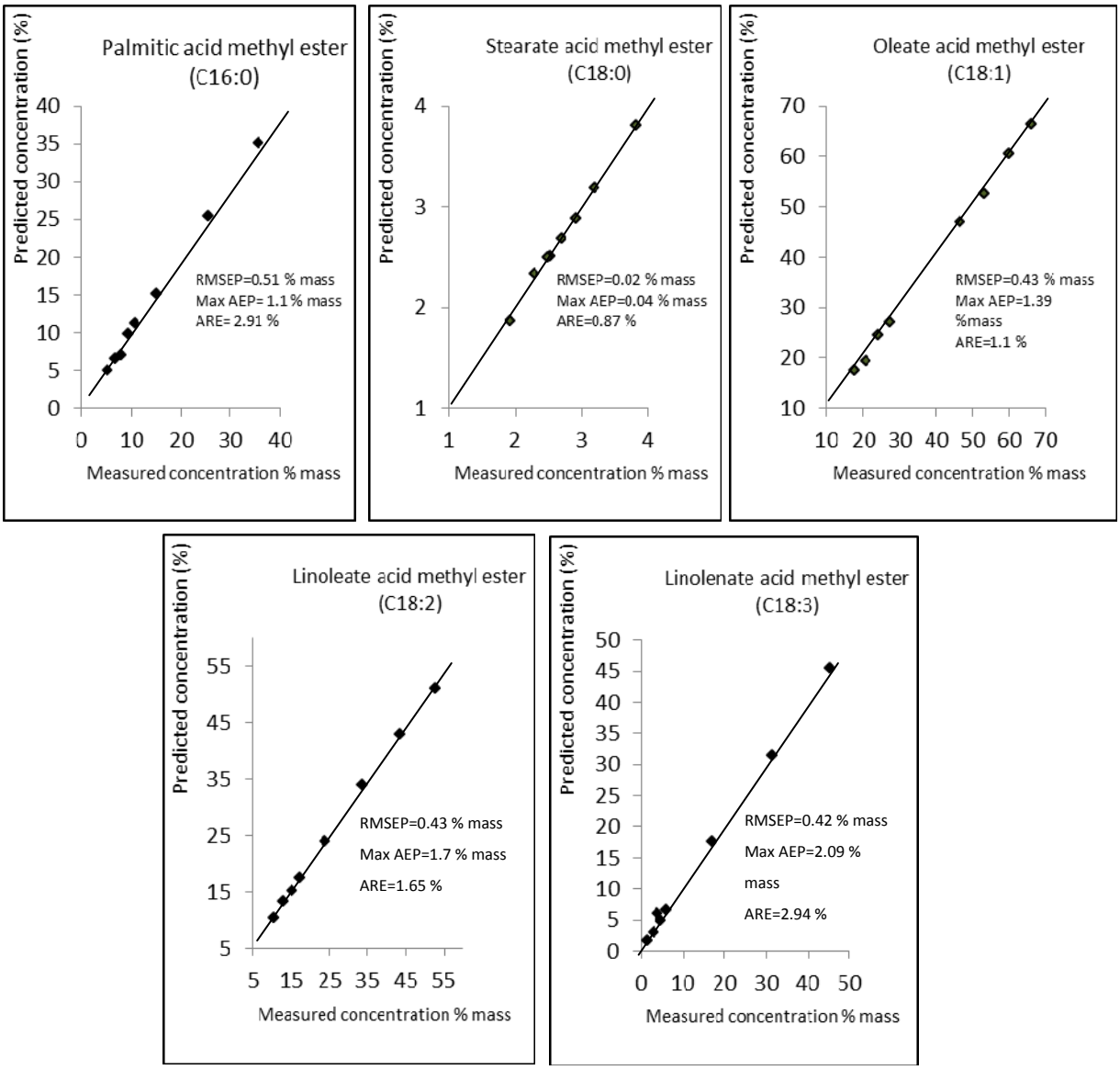

Figure 4-13 Results of validation set for each type of methyl, using MIR spectroscopy and internal samples MOD-4M.

Results obtained from the MIR methods (MOD-4M) indicate the calibration models developed to quantify concentration of the main fatty acids present in biodiesel, from the MIR spectra, performed accurately when a set of internal samples was used for validation, this means that the fatty acids profile of biodiesel can be determined using the MIR spectroscopy method, when the source of the biodiesel is known.

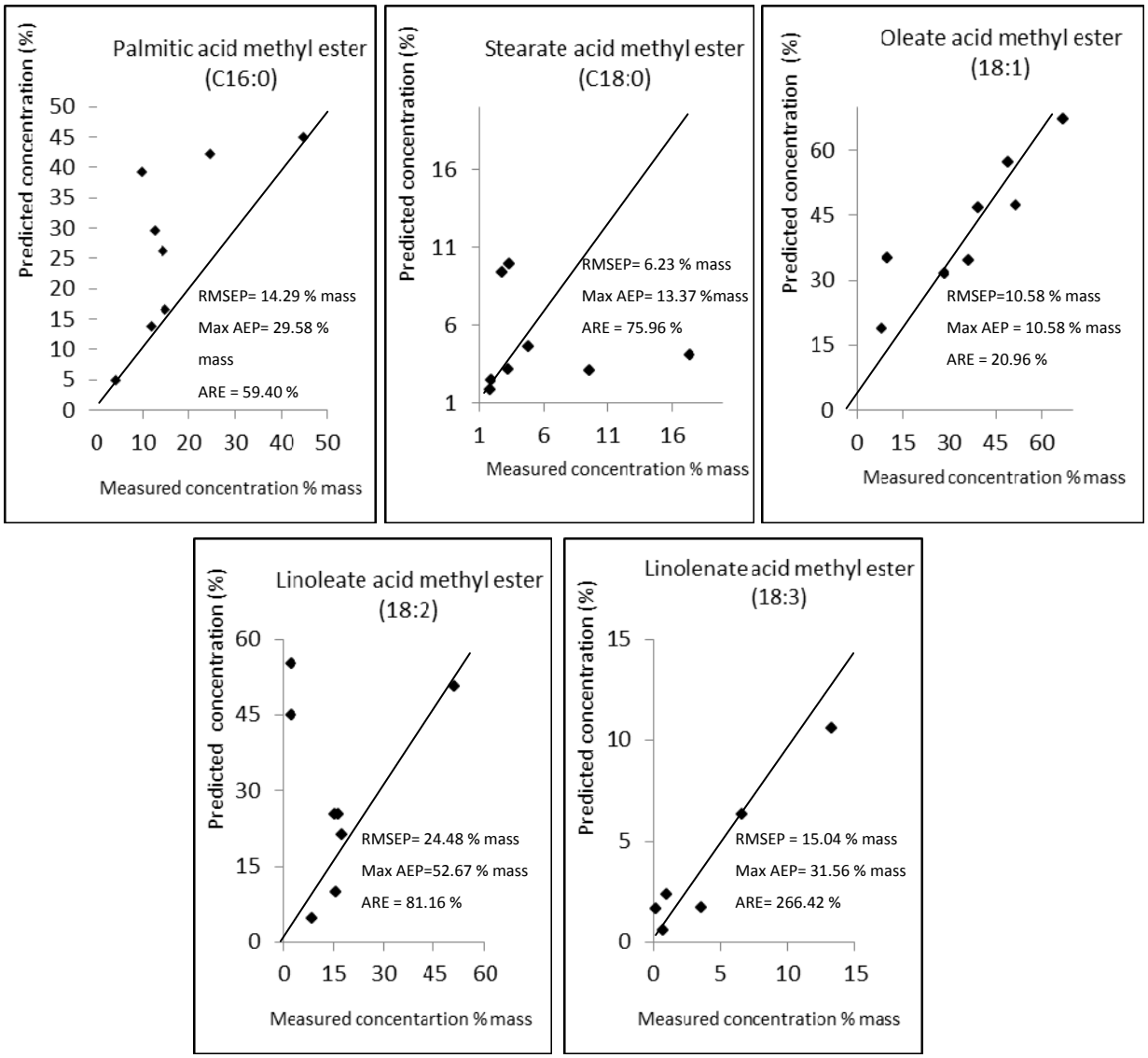

Figure 4-14 Results of validation set for each type of Methyl, using MIR spectroscopy and external samples MOD-4M.

Similarly to the approach presented for NIR method, two additional biodiesels (AFME, and CCME) were included to develop a new calibration.

4.4.2.3 Predicting fatty acid composition, models using six types of biodiesel (MOD-6M).

The same one hundred and twenty-six samples used for the MOD-6N in the NIR study were used to develop the MOD-6M in the MIR range to predict the main fatty acid methyl ester (FAME) present in biodiesel. Table 4-6 shows the results for MOD-6M in the MIR range.

Similar to the NIR models, pre-treatment data was not used, but some outliers were detected in the model for methyl palmitate and methyl stearate. The samples used in the validation were from the combination of six biodiesels used on the calibration, plus some types of biodiesel that were not present in the calibration.

Table 4-6 General results of MOD-6M calibration, for the prediction of C16:0, C18:0, C18:1, C18:2, C18:3, using the PLSR method and the MIR Spectra.

Type of Methyl Ester Models II	Multiple Correlation Coefficient (R^2) Calibration	Standard Error Cross Validation (SECV) % mass	Number of Factors	RPD Values
Palmitate	0.936	2.11	7	4.86
Stearate	0.944	0.709	8	5.82
Oleate	0.993	1.19	7	13.55
Linoleate	0.996	0.65	7	19.19
Linolenate	0.998	0.74	7	22.57

The higher SECV was 2.11 % mass for methyl palmitate model, and the minimum correlation coefficient (R^2) was 0.944, detected in the model for methyl stearate. Validation results using the internal set of samples are shown in Figure 4-15. The maximum absolute error was 4.31 % mass with an average relative error of 2.49 % presented in the model for C18:1. The maximum RMSEP was 1.53 % mass, which was presented in the same model for C18:1.

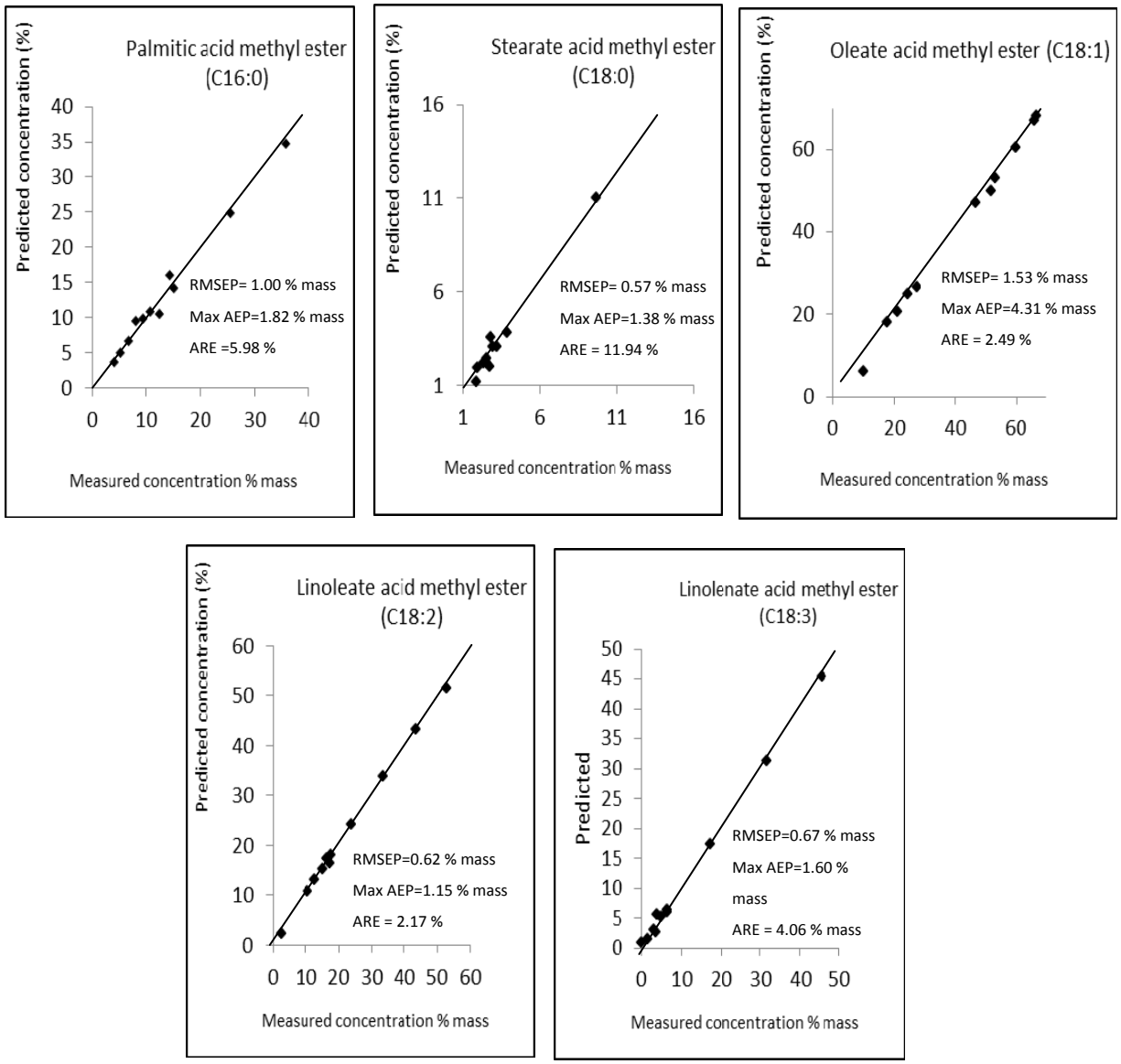

Figure 4-15 Results of validation set for each type of methyl, using MIR spectroscopy and internal samples MOD-6M model.

Regression coefficients for the calibration models (MOD-6M) are presented in Figure 4-16. Similar to the models of four biodiesel, the calibration models are based on absorption bands related to the fatty acid methyl esters.

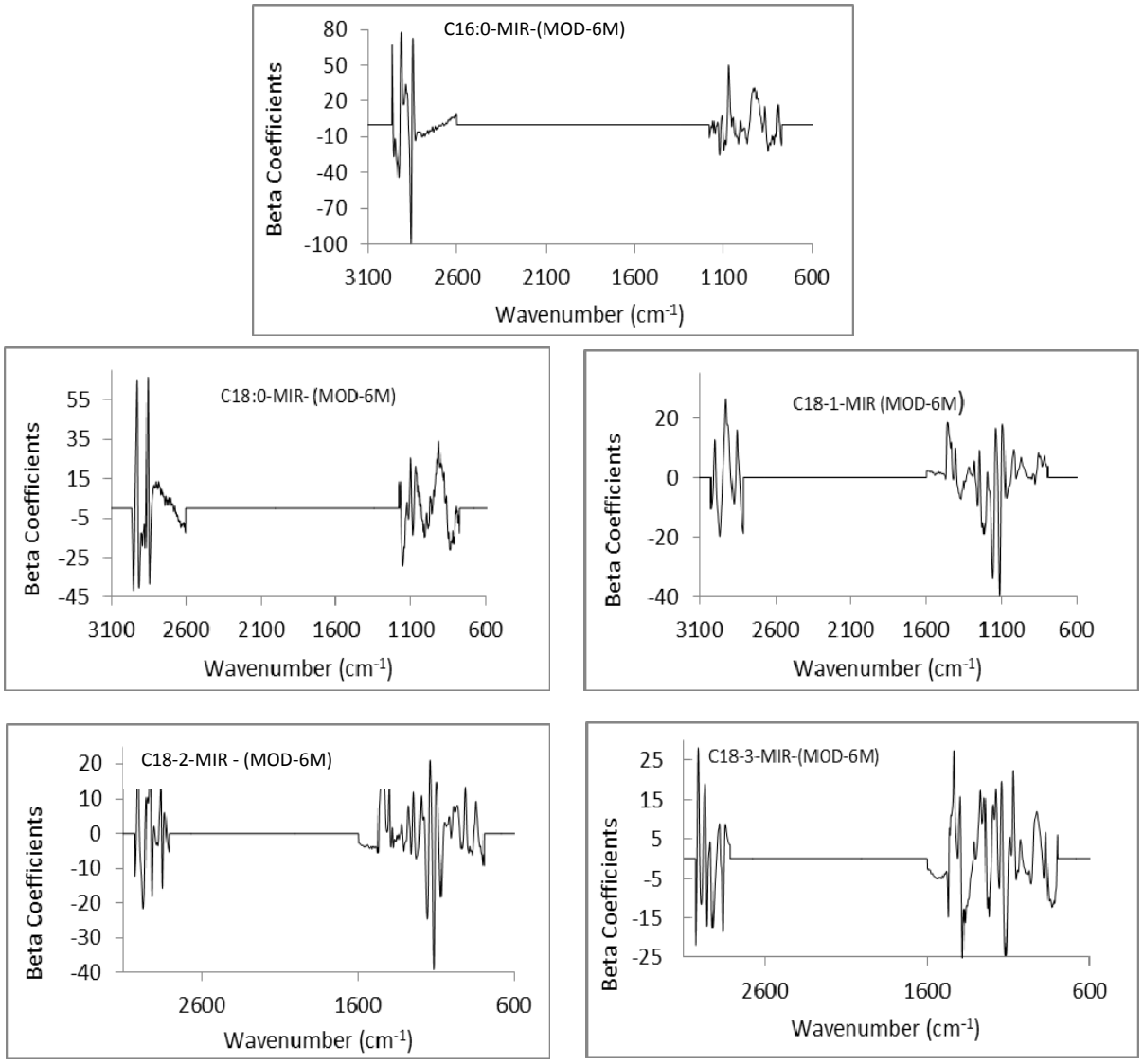

Figure 4-16 Regression coefficients for the models MOD-6M.

Similarly to the procedure developed for NIR method, to compare the performance of the models MOD-6N and MOD-4N over the same condition, the models were tested with a set of external samples. Figure 4-17 shows the results for the validation set using these external samples. The maximum absolute error was 6.27 % mass with an average relative error of 11.80 % presented in

the model for C16:0. The maximum RMSEP was 2.73 % mass presented in the same model for C16:0.

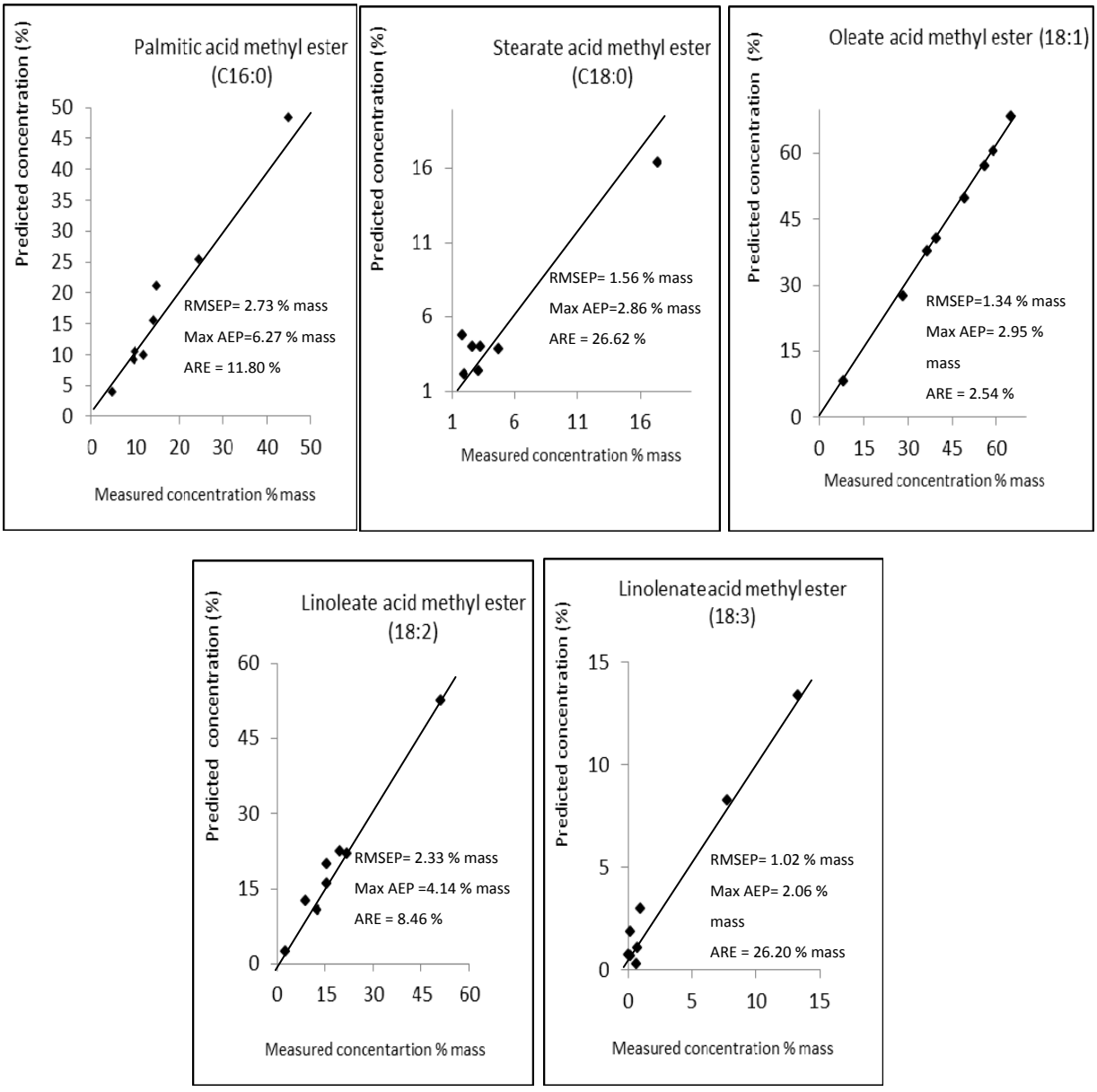

Figure 4-17 Results of validation set for each type of Methyl, using MIR spectroscopy and external samples MOD-6M model.

Similar results to the NIR method were found with the model of six different types of biodiesel. Considering the errors of the validation set using internal and external samples are

similar for the MOD-6N model, all statistic parameters used to verify the model were re-calculated using all validation samples as a one set. Figure 4-18 shows the comparison between predicted and measured fatty acid concentration for each model using the internal and external samples for validation. The maximum absolute error was 4.14 % mass with an average relative error of 4.93 % presented in the model for C18:2. The maximum RMSEP was 1.93 % mass, which was presented in the model for C16:0. These are very similar results to the ones from the NIR study. Validation results show accurate results for cases of methyl oleate, methyl linoleate, and methyl linolenate, but not for cases of methyl palmitate, and methyl stearate models.

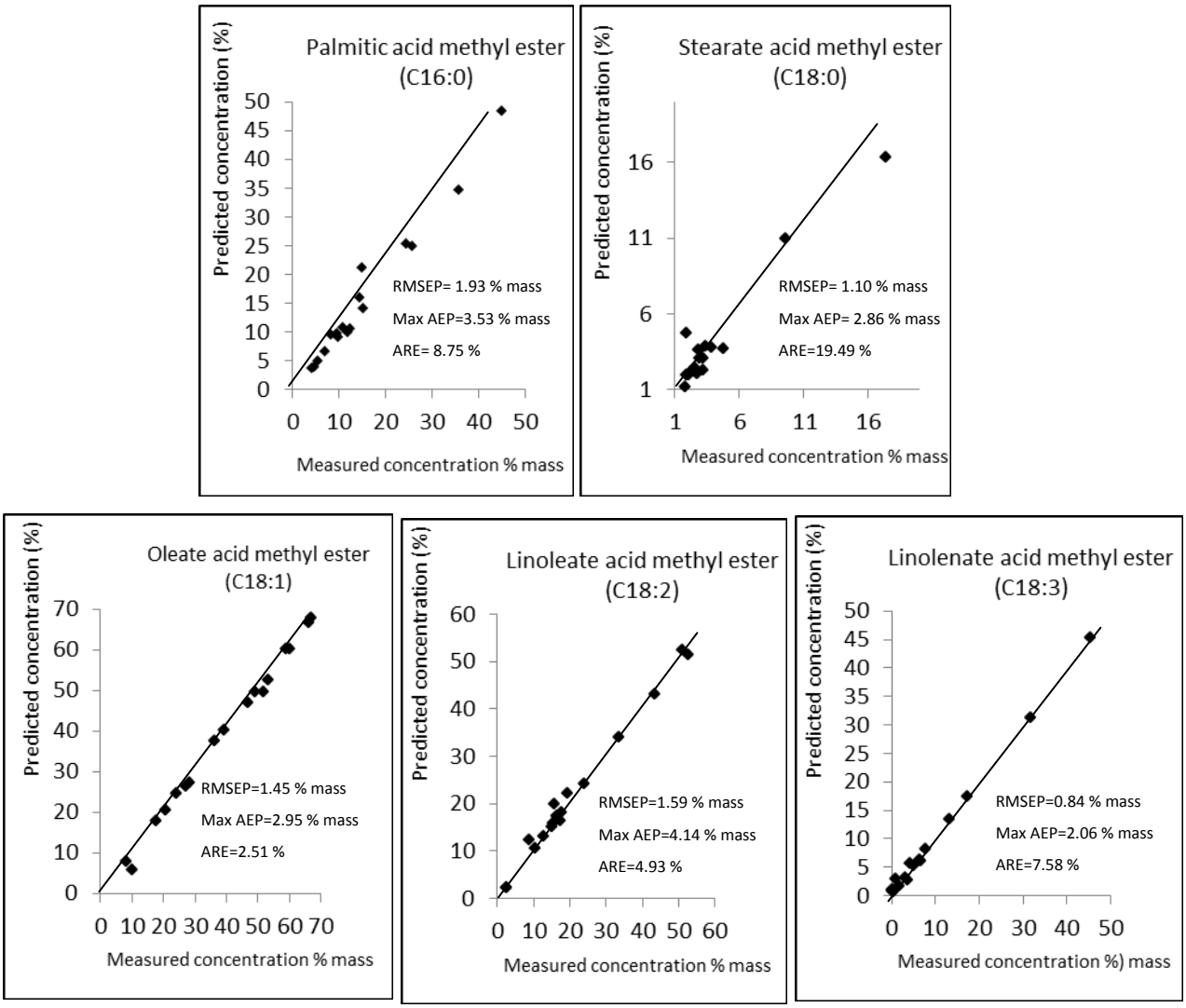

Figure 4-18 Results of validation set for each type of methyl, using MIR spectroscopy MOD-6M model.

The obtained results from the MIR method were very similar to the NIR study. Consequently, these results showed that the fatty acid concentration of 18:1, 18:2, and 18:3 in biodiesel can be determined using the MIR spectroscopy method, with reasonable accuracy, even when the source of the biodiesel is unknown. The models developed to predict 16:0 and 18:0, with the MIR methods present slightly better results than the models developed with NIR spectra. To complete the comparative study between the NIR and MIR methods, the three sub-models developed for the NIR method for the same specific range were developed for this case. Figure 4-19 shows the comparison of predicted and actual concentration of the validation set.

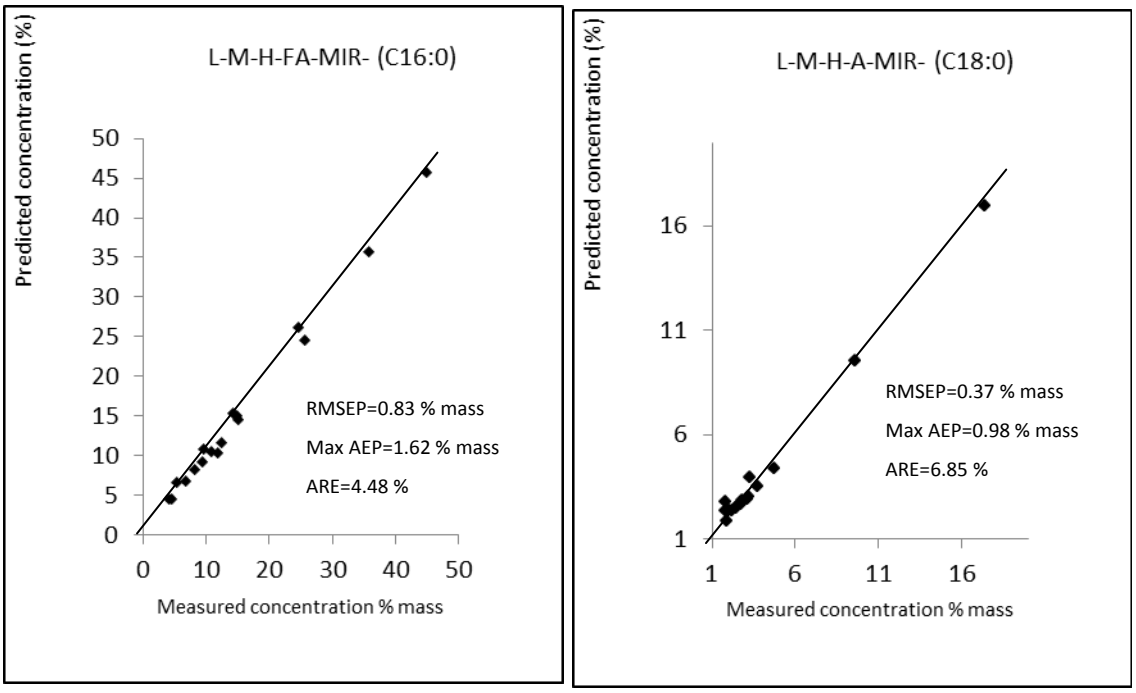

Figure 4-19 Results of validation set for methyl palmitate (C16:0), and methyl stearate (C18:0) using the sub-models (lower, medium, and high) over MIR range.

Similar to the NIR method, the obtained results using the sub-models (low, medium, and high) improved the prediction accuracy of the original models for methyl palmitate, and methyl stearate. The maximum absolute error was 1.62 % mass with an average relative error of 4.48 % presented in the model for C16:0. The maximum absolute error for C18:0 model was 0.98 % mass with an average relative error of 6.86 %. The RMSEP is reduced from 1.93 to 0.83 % mass

for the case of C16:0 and from 1.10 to 0.37 % mass for the case of C18:0. Considering results of the two step models to predict 16:0 and 18:0, the MIR method can be used to predict the fatty acid present in biodiesel with reasonable accuracy when the source of biodiesel is unknown.

Performance factor of the PLSR models that were developed for fatty acid prediction using NIR and MIR spectra are summarized in Table 4-7. The MIR spectrum models were slightly better than NIR at predicting C16:0 and C18:0, Contrarily, NIR present slightly better results when predicting C18:1, C18:2, and C18:3. Although, both methods were found reasonably accurate when predicting the concentrations of C16:0, C18:0, C18:1, C18:2, and C18:3 in biodiesel samples.

Table 4-7 Validation results for prediction of the main fatty acid present in biodiesel

	NIR Models					MIR models				
	C16:0*	C18:0*	C18:1	C18:2	C18:3	C16:0*	C18:0*	C18:1	C18:2	C18:3
R^2	0.92	0.98	0.99	0.99	0.99	0.96	0.98	0.99	0.99	0.99
SECV	0.53	0.05	0.96	0.76	0.66	0.48	0.04	1.49	0.71	0.79
RMSEP	1.51	0.56	1.12	0.70	0.69	0.83	0.37	1.45	1.59	0.84
Max AEP	2.25	1.34	3.19	1.29	1.78	1.62	0.98	2.95	4.14	2.06
ARE	6.56	9.75	1.89	2.73	6.97	4.48	6.85	2.51	4.93	7.58

* Data from model used in the validation (low, medium or high concentration) MOD- 6N/M

4.5 Conclusions

The selected ranges of NIR and MIR spectra allowed building reliable models to predict the concentration of C16:0, C18:0, C18:1, C18:2 and C18:3 fatty acids present in biodiesel samples, when the source of biodiesel is known (MOD-4N/M.)

The models developed using heterogonous types of biodiesel (MOD-6N/M) presented accurate predictions for C18:1, C18:2, and C18:3. The average relative error of prediction (ARE) was smaller than 6.97 % and 7.98 % for NIR and MIR spectroscopy methods respectively. For

these types of models additional work was necessary to obtain a reasonable accurate prediction for the cases of C16:0 and C18:0 fatty acid.

The results obtained using the MIR spectroscopy data were slightly better than those obtained with the NIR spectroscopy data for the C16:0 and C18:0 models. For the case of 16:0, the average relative errors of prediction were 4.48% and 6.56% for MIR and NIR spectroscopy methods respectively. However, the validation results also showed that both NIR and MIR spectroscopy methods can be used to predict the fatty acid profile in biodiesel, although the biodiesel type is unknown. NIR and MIR spectroscopy methods can be considered as a promising tool to effectively monitor the biodiesel quality over the production and distribution chain.

4.6 References

Allen, . (1999). Predicting the viscosity of biodiesel fuels from their fatty acid ester composition. Fuel, 78(11), 1319.

Ambrozin, , Liao, , Ferreira, , Monteiro, , & , . (2008). Critical review on analytical methods for biodiesel characterization. Talanta, 77(2), 593.

Azizian, H., Kramer, J. K. G., & Winsborough, S. (2007). Factors influencing the fatty acid determination in fats and oils using fourier transform near-infrared spectroscopy. European Journal of Lipid Science and Technology, 109(9), 960-968.

Baptista, P., Felizardo, P., Menezes, J. C., & Correia, M. J. N. (2008). Multivariate near infrared spectroscopy models for predicting the methyl esters content in biodiesel. Analytica Chimica Acta, 607(2), 153-159.

Bellorini, S., Fumiere, O., Strathmann, S., Baeten, V., Fumire, O., Berben, G., et al. (2005). Discriminating animal fats and their origins: Assessing the potentials of fourier transform infrared spectroscopy, gas chromatography, immunoassay and polymerase chain reaction techniques. Analytical and Bioanalytical Chemistry, 382(4), 1073-1083.

Dardenne, . (2007). Prediction of fatty acid contents by mid-infrared spectrometry. Journal of Dairy Science, 90, 274.

De Marchi, M., Penasa, M., Cecchinato, A., Mele, M., & Secchiari, P. (2011). Effectiveness of mid-infrared spectroscopy to predict fatty acid composition of brown swiss bovine milk. Animal, 5(10), 1653-1658.

De Voort, , Simpson, , & Aryee, . (2009). FTIR determination of free fatty acids in fish oils intended for biodiesel production. Process Biochemistry, 44(4), 401.

Falk, . (2004). The effect of fatty acid composition on biodiesel oxidative stability. European Journal of Lipid Science and Technology, 106(12), 837.

Gonzalez-Martin, I., Gonzalez-Perez, C., Hernandez-Mendez, J., & Alvarez-Garcia, N. (2003). Determination of fatty acids in the subcutaneous fat of iberian breed swine by near infrared spectroscopy (NIRS) with a fibre-optic probe. Meat Science, 65(2), 713-719.

Guy, F., Prache, S., Thomas, A., Bauchart, D., & Andueza, D. (2011). Prediction of lamb meat fatty acid composition using near-infrared reflectance spectroscopy (NIRS). Food Chemistry, 127(3), 1280-1286.

Knothe, G. (2008). "Designer" biodiesel: Optimizing fatty ester (composition to improve fuel properties. Energy Fuels, 22(2), 1358.

Knothe, G. (2005). Dependence of biodiesel fuel properties on the structure of fatty acid alkyl esters. Fuel Processing Technology, 86(10), 1059-1070.

Koprna, R., Nerusil, P., Kolovrat, O., Kucera, V., & Kohoutek, A. (2006). Estimation of fatty acid content in intact seeds of oilseed rape (brassica napus L.) lines using near-infrared spectroscopy. Czech Journal of Genetics and Plant Breeding, 42(4), 132-136.

Mahamuni, N. N., & Adewuyi, Y. G. (2009). Fourier transform infrared spectroscopy (FTIR) method to monitor soy biodiesel and soybean oil in transesterification reactions,

Oliveira, F., Brandao, C., Ramalho, H., da Costa, L., Suarez, P., & Rubim, J. (2007).

Patil, A. G., Oak, M. D., Taware, S. P., Tamhankar, S. A., & Rao, V. S. (2010). Nondestructive estimation of fatty acid composition in soybean [glycine max (L.) merrill] seeds using near-infrared transmittance spectroscopy. Food Chemistry, 120(4), 1210-1217.

Prieto, N., Lopez Campos, O., McAllister, T. A., & Aalhus, J. L. (2012). Near infrared reflectance spectroscopy predicts the content of polyunsaturated fatty acids and

biohydrogenation products in the subcutaneous fat of beef cows fed flaxseed. Meat Science, 90(1), 43-51.

Sablinskas, V., Steiner, G., Hof, M. (2003) Chapter 6-Application. In G. Gauglitz, and T.Vo-Dinh (Eds.), Handbook of Spectroscopy (pp.89-168). Germany: WILEY-VCH.

Sato, T. (2002). New estimation method for fatty acid composition in oil using near infrared spectroscopy. Bioscience Biotechnology and Biochemistry, 66(12), 2543-2548.

Shenk, J. S., Workman, J. J., Jr., Westerhaus, M. O. (2008) Application of NIR Spectroscopy to Agricultural Products. In D. Burns and E. Ciurczak (Eds.), Handbook of Near-Infrared Analysis (pp.347-386). Boca Raton, FL: Taylor &Francis Group.

Soyeurt, H., Dardenne, P., Dehareng, F., Lognay, G., Veselko, D., Marlier, M., et al. (2006). Estimating fatty acid content in cow milk using mid-infrared spectrometry. Journal of Dairy Science, 89(9), 3690-3695.

Velasco, L., Mollers, C., & Becker, H. C. (1999). Estimation of seed weight, oil content and fatty acid composition in intact single seeds of rapeseed (brassica napus L.) by near-infrared reflectance spectroscopy. Euphytica, 106(1), 79-85.

Williams, P. C. (2001) Chapter 8 Implementation of Near-Infrared Technology. In P. Williams, and K. Norris (Eds.), *Near- Infrared Technology* (pp.145-169). USA: American Association of Cereal Chemists, Inc.

Yaakob, B. C. M., Syahariza, Z. A., Rohman, A. (2010) Chapter 1. Fourier Transform (FTIR) Spectroscopy: Development, Techniques, and Application in the Analyses of Fats and Oil . In O. J. Rees(Eds.), Fourier Transform Infrared Spectroscopy: Development, Techniques, and Application (pp. 1-26).New York, NY: Nova Science.

Chapter 5 - Quantifying Trace Biodiesel Impurities Using Fourier-Transformed Mid- and Near-Infrared Spectroscopy

5.1 Abstract

This work reports the use of mid- and near-infrared (MIR and NIR) spectroscopy to determine the content of free glycerin, triglycerides, water, and methanol in biodiesel. Partial least square regression (PLSR) modeling method was used for calibration based on six different types of biodiesel (methyl esters of palm oil, soybean oil, corn oil, peanut oil, olive oil, and canola oil) containing the four impurities above mentioned in certain ranges. All samples were scanned on an FT-IR spectrometer, and specific prediction models for methanol, water, free glycerin, and triglycerides were developed in two different approaches: when a single impurity was present and when all impurities were present. The results showed that the mid- and near-infrared spectroscopy were able to accurately and rapidly predict the amount of four impurities in the biodiesel when only a single impurity was present in the biodiesel sample. The root mean square error of prediction (RMSEP) for methanol, water, triglycerides, and glycerol models were 206, 68, 69, and 56 mg kg^{-1}, respectively in the MIR range. For the NIR range, the RMSEP were 125 mg kg^{-1} for methanol model, 49 mg kg^{-1} for water model, 647 mg kg^{-1} for triglycerides model, and 17 mg kg^{-1} for glycerol model. When all impurities were present in biodiesel, only the NIR method demonstrated certain capacities to predict these impurities. Pre-treatment of the data by Savitzky-Golay's second derivative was necessary to achieve reasonable accuracy. The root mean square error of prediction (RMSEP) for methanol, water, triglycerides, and glycerol models were 170, 90, 900, and 50 mg kg^{-1}, respectively.

5.2 Introduction

Biodiesel is mainly produced by a catalytic transterification reaction of vegetable oils or animal fat with an alcohol, such as methanol. It is considered an important alternative fuel for diesel engines, because it reduces greenhouse gases, sulphur emissions, and particle matter, Van Gerpen (2004).

Additionally, it is biodegradable and renewable. At the end of the reaction, crude glycerol is produced as a byproduct, and a water-washing process is usually necessary to reduce impurity levels in the biodiesel. However, some of the raw materials such as triglycerides and methanol, or products such as glycerol and water, could remain in the final product compromising the biodiesel quality, Banga & Varshney (2010). To guarantee the quality of biodiesel, some standards have been established. The best known are ASTM D6751 used in the United States and EN 14214 used in Europe. The maximum levels of methanol, water, triglycerides, and glycerol have been clearly established in these standards. With the rapid increase in biodiesel production and use worldwide, it is necessary to develop a fast, simple and precise method for monitoring biodiesel quality. The presence of contaminants in biodiesel which as methanol, residual glycerol, triglycerides (unreacted oil), and water can severely damage diesel engines. High methanol content in biodiesel promotes the deposits formation on the injectors. Furthermore, methanol content in large quantities can adversely affect some fuel properties, which as heating value and flash point, decreasing the overall performance of the fuel. Glycerol content in biodiesel has been reported as a cause of deposit in the injector tip as well as in the combustion chamber. Some emission problems have also been attributed to the presence of glycerol in the fuel. The water presence in biodiesel can cause corrosion problems in the engine, and can also react with triglycerides producing an undesirable compound. Finally, triglycerides in biodiesel can generate emissions of noxious pollutants from the combustion, Mittelbach et al. (2010).
The use of infrared spectroscopy in the analysis of biodiesel has been reported in literature in combination with multivariate techniques. Many authors have already developed reliable predictions of some biodiesel properties and have also published the use of these tools in quality control tasks. P. Felizardo et al. (2007) determined water and methanol content using NIR spectroscopy in biodiesel samples from soybean, palm oil, and waste frying oils. The best mean square error of prediction (RMSEP) achieved was 70 mg/kg in the region 4800-5050 cm^{-1} for methanol concentration model and 87 mg/kg in the region 9000 – 4500 cm^{-1} for water concentration model. Water content used in calibration ranged from 218 to 1859 mg/kg, and methanol content ranged from 2 to 1859 mg/kg. The authors concluded that results present excellent agreement between measured and predicted values, and also concluded that water content does not affect the calibration of methanol or vice versa. Using NIR/MIR spectroscopy V. Gaydou et al. (2011) reported the prediction of concentration of triglycerides in a blend of

diesel-biodiesel- triglycerides. The RMSEP was 2.11% (w/w) for the model using MIR range and 0.363 % (w/w) for the model for NIR range. The authors concluded that the developed models predicted triglycerides concentration with good accuracy when the concentration is ranged from 0 to 40 % (w/w) in a blend of diesel-biodiesel-triglycerides. F. Oliveira et al. (2007) presented similar studies to predict the concentration of triglycerides in a blend of diesel-biodiesel-triglycerides using FT-NIR spectroscopy and FT-Raman spectroscopy. The best values for RMSEP were 0.238 % (w/w) using FT-NIR and PLS method and 0.604 % (w/w) using FT-Raman and PCR method. The authors concluded that the developed method predicted concentration of triglycerides accurately, when the concentration of triglycerides ranged from 0 to 5 % (w/w). M.P.Dorado et al. (2011) determined methanol and glycerol traces in biodiesel using visible and NIR ranges. The accuracy of calibration was determined by RPD. The RPD was 10 for the methanol model and 2.5 for the glycerol model, respectively. The samples for this study were ranged with methanol from 0.0003% to 0.433% (w/w) and with glycerol from 0.005% to 0.050 % (w/w) in two separate sets. The authors concluded that NIR and visible ranges are able to detect methanol and glycerol traces in biodiesel, but recommend additional work to improve the method's accuracy to detect glycerol in biodiesel samples. I. P.Soares et al. (2008) presented a study to predict the triglyceride content in biodiesel using FT-IR. The range for RMSEP of developed models was from 0.65 to 2.09 % (v/v), concluding that FT-IR method using PLS is able to predict the triglyceride concentration in biodiesel with good accuracy, when the range of the triglyceride content is ranged from 0 to 40 % (v/v). The revised previous work showed the ability to predict methanol, glycerol, triglycerides, and water in biodiesel when a single impurity is present. The study of P. Felizardo was the only one that presented two impurities at the same time (methanol and water). Considering that the four listed impurities could be present in the biodiesel at the same time and no report can be found on predicting the concentration of these impurities under this condition, the objective of this work was to predict the concentration of methanol, triglycerides, glycerol, and water when all of these impurities are present in biodiesel, using both MIR and NIR spectroscopy. The PLSR method was used to develop calibration models for the listed impurities in biodiesel. The second derivative of Savitzky-Golay was used as pre-treatment of the data for NIR models.

5.3 Materials and Methods

5.3.1 Sample preparation

The biodiesels used in this work were prepared from food grade canola oil, corn oil, peanut oil, olive oil, and soybean oil, purchased from local grocery stores, and palm oil purchased from Country Soap Shack (Missouri, USA). All biodiesel samples were produced using a standard transterification process described elsewhere, Sagar et al. (2006). After separation, washing and drying processes were developed over the entire biodiesel to minimize the level of impurity. Later, all the biodiesel was analyzed to obtain the fatty acid profile by gas chromatograph (GC) to verify purity.

Table 5-1, shows the list of fatty acid methyl ester (FAME) used in this study, palm methyl ester (PAME), canola methyl ester (CAME), peanut methyl ester (PEME), olive methyl ester (OLME), corn methyl ester (COME), and soybean methyl ester (SOME).

Table 5-1 Fatty acids profile (mass %) of biodiesel samples prepared for this work.

FAME(mass%)	***C8:0***	***C10:0***	***C12:0***	***C14:0***	***C16:0***	***C18:0***	***C18:1***	***C18:2***	***C18:3***
COME	*0.02*	*0.00*	*0.00*	*0.07*	*10.88*	*2.27*	*27.63*	*53.95*	*2.36*
SOME	*0.00*	*0.00*	*0.08*	*0.10*	*10.40*	*4.01*	*20.92*	*50.24*	*7.31*
PAME	*0.02*	*0.03*	*0.34*	*1.23*	*44.57*	*4.39*	*40.96*	*8.56*	*0.17*
CAME	*0.00*	*0.01*	*0.01*	*0.08*	*4.14*	*1.84*	*66.99*	*17.59*	*6.56*
PEME	*0.01*	*0.01*	*0.12*	*0.09*	*9.82*	*2.75*	*55.97*	*21.77*	*0.21*
OLME	*0.00*	*0.00*	*0.10*	*0.05*	*14.07*	*2.82*	*65.07*	*12.31*	*0.61*

5.3.2 Impurities level

Materials used in preparing samples for the infrared scan were deionized water, glycerol (99.5%, G33-1, Fisher Scientific), methanol (99.8%, A412-1, Fisher Scientific), and corn oil as triglycerides. Using the biodiesel listed above, this study was developed in two steps. First, a set

of samples was prepared using a single impurity to verify the ability of the NIR and MIR spectroscopy to identify very small amounts of methanol, triglycerides, glycerol, and water in biodiesel. Next, a new set of samples was prepared for each impurity. However, at this time, the other three impurities were present at different amounts. Four different sets of fifty biodiesel samples, were each polluted with a single impurity for the first step: as well as glycerol, in a range from 10 to 470 *mg kg*$^{-1}$, methanol in a range from 110 to 5000 *mg kg*$^{-1}$ triglycerides, in a range from 110 to 4770 *mg kg*$^{-1}$, and water in a range from 20 to 940 *mg kg*$^{-1}$. Forty four samples were used in the calibration, and six samples were reserved for validation in this first step. For the second step, four other sets, each containing thirty biodiesel samples, were contaminated with methanol, glycerol, triglycerides, and water in the same ranges as previous ones. However, in this experiment the other three impurities had been present at the same time in different concentrations in each sample, as shown in Table 5-2. To develop the calibration, the first set of samples (44) and the second set of samples (23) were combined to be used in the models with all impurities. Sixty seven samples were used in the calibration, and thirteen samples were randomly selected from the entire set for the validation.

Table 5-2 Ranges of the concentration used for the impurities models (one impurity/all impurities)

Impurities Model For:	Max. Conc. %(w/w) ASTM D6751 Or EN 14214	Range for calibrations *mg kg*$^{-1}$ (one impurity) (all impurities)	Concentration *mg kg*$^{-1}$, Other impurities (All Impurities Models)			
			Methanol	Water	Triglycerides	Glycerol
Methanol	0.25	110 to 5000	----	200 500 700	500 700 200	700 200 500
Water	0.05 %(v/v)	20 to 940	200 500 700	----	500 700 200	700 200 500
Triglycerides	0.25	110 to 4770	200 500 700	500 700 200	----	700 200 500
Glycerol	0.02	10 to 470	200 500 700	500 700 200	700 200 500	----

To obtain the desired levels of impurity, every sample was prepared by weight using previously prepared biodiesel samples with each impurity. A dilution process was necessary to achieve the final level.

5.3.3 NIR and MIR spectra acquisition

To develop an impurities prediction model, the samples were scanned at a room temperature of 22-24^{o}C on a FT-IR/FT-NIR spectrometer (Perkin-Elmer spectrum 400, PerkinElmer, Inc., Shelton, CT). For the NIR scan, the samples were placed in a quartz standard cell of 1mm pathlength for spectrophotometers (Labomed Inc. Culver City, CA). The data was recorded as the absorbance in the wavelength range from 900 to 2500 nm with a resolution of 4 cm^{-1} at 1 cm^{-1} interval. For the MIR infrared scan, universal ATR accessory of Germanium (Ge) crystal with refractive index of 4.0 (PerkinElmer, Inc., Shelton, CT) was used. The spectra were recorded as the absorbance in the wavelength range from 4000 to 600 cm^{-1} with a resolution of 4 cm^{-1} at 1 cm^{-1} interval. All spectrums were recorded once for each sample and were obtained as an average of 32 scans.

5.3.4 Calibration models

5.3.4.1 Model with single impurity (S-IMP)

In the NIR range, the selected region to estimate the concentration of methanol, glycerol, and triglycerides in a biodiesel sample were located between 1100 and 2300 nm. For the case of water the used regions were between 1300 and 1500 nm and between 1800 and 2000 nm. The regions on the MIR range used to estimate the concentration of glycerol in biodiesel samples were located between 3925 and 840 cm^{-1}. For the cases of methanol, triglycerides, and water, the used regions were between 3650 and 600 cm^{-1}. These regions were chosen based on the absorbance band of the pure impurity in the NIR and MIR region.

5.3.4.2 Model with all impurities (A-IMP)

In the case of models for all impurities, the regions were chosen based on derivative technique analysis. The regions from 1100 to 1650 nm and from 1800 to 2150 nm were selected to develop the calibration model for methanol. The glycerol model was developed over the ranges from 1400 to 1650 nm and from 1800 to 2100 nm. The regions from 900 to 1250 nm and

from 1300 to 1500 nm were used to develop the calibration model for triglycerides. For the case of water, three regions were used, from 950 to 1150 nm, from 1400 to 1650 nm, and from 1800 to 2050 nm. Grams software version 6, using a PLSR method, was used to develop the models for both cases (single and all impurities). The performance of the models was evaluated by the correlation coefficient (R^2), standard error of cross validation (SECV), the root mean square error of prediction (RMSEP), the average relative error (AVRE), the absolute error of prediction (ABEP), and the RPD value. These values were calculated using the following equations:

$$\mathrm{SECV} = \sqrt{\frac{\sum_{i=1}^{n}(\mathrm{X_{ATi}} - X_{PTi})^2}{\mathrm{n}}} \tag{1}$$

$$\mathrm{RMSEP} = \sqrt{\frac{\sum_{i=1}^{N}(\mathrm{X_{AVi}} - X_{PVi})^2}{\mathrm{N}}} \tag{2}$$

$$\mathrm{ABEP} = \left|X_{AVi} - X_{PVi}\right| \tag{3}$$

$$\mathrm{AVRE} = \frac{RMSEP}{\overline{X}} \tag{4}$$

$$\mathrm{RPD} = \frac{SD}{SECV} \tag{5}$$

Where X_{PT} is the predicted value of each sample in the training set, X_{PV} is the predicted value of each sample in the validation set, X_{AT} is the measured or actual value of the sample in the training set, X_{AV} is the measured or actual value of the sample in the validation set, $\overline{X}$ is the mean, n is the number of samples in the training set, N is the number of samples in the validation set, and SD is the standard deviation.

5.4 Results and Discussion

5.4.1 Models with single impurity (S-IMP)

5.4.1.1 Prediction using MIR spectroscopy

5.4.1.1.1 MIR spectra

Raw spectra of methanol, triglycerides, glycerol, and water in the MIR range are shown in Figure 5-1. The most prominent bands around 3350, 1640, and 675 cm^{-1} are characteristics for water, J. B. Brubach et al. (2005). Methanol shows its distinctive absorption bands near to 3340, 2945, 2833, 1456, 1030, and 655 cm^{-1}, National Institute of Advanced Industrial Science and Technology, (2012). For glycerol, the absorption bands were identified around 3300, 2930, 1040, and 650 $cm^{-1,}$ National Institute of Advanced Industrial Science and Technology, (2012). The spectra of triglycerides presents its absorption band around 2932, 2880, 1740, 1416, 1043, 924, and 854 cm^{-1}, among others, Y. Che Man et al. (2010). The selected calibration ranges in the MIR range were based on the region spectra where the impurity has a strong absorption band. In the MIR region, models with only a single impurity were developed.

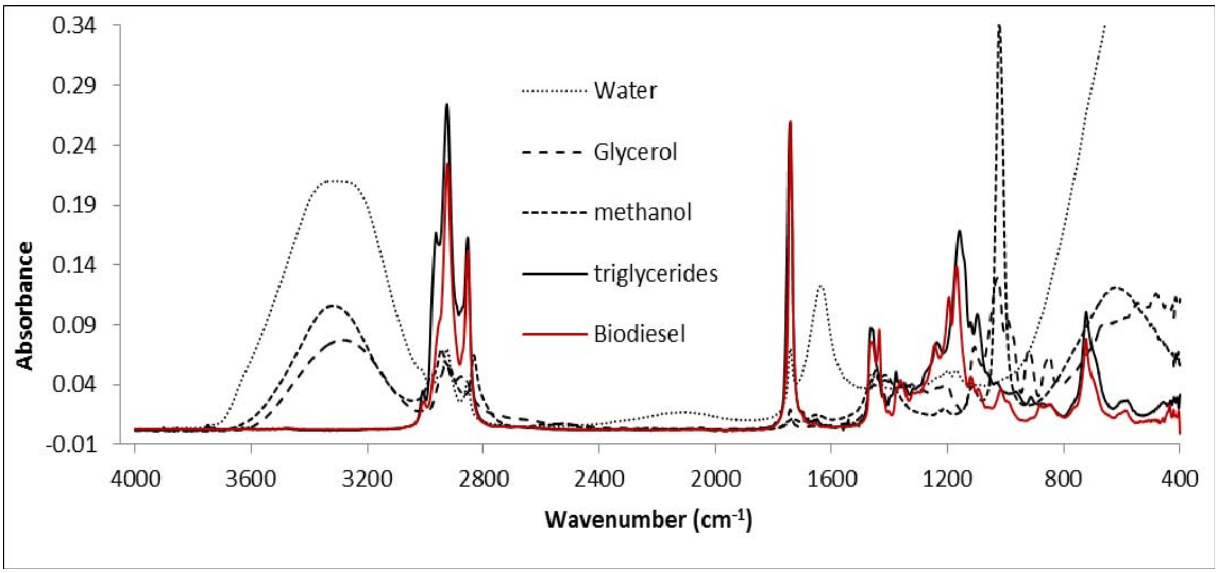

Figure 5-1 The mid infrared spectra of pure glycerol, methanol, triglycerides, and water

5.4.1.1.2 Methanol content prediction

The MIR spectral region from 3600 to 802 cm^{-1} was selected to develop the calibration model for methanol. This region includes almost all predominant bands for methanol, listed above. The PLSR model was tested with the validation set. The correlation coefficient R^2 = 0.977, and SECV = 210 mg kg^{-1} were obtained. The spectral region of 3600 to 802 cm^{-1} shows at 3340 cm^{-1} indications of O-H stretching, at 2945 cm^{-1} indications of C-H asymmetric stretching, at 2833 cm^{-1} indications of C-H symmetric stretching, at 1456 indications of asymmetric deformation with $–CH_3$ structure, and at 1030 cm^{-1} indications of C-O Stretching, V. Sablinskas et al. (2003). Figure 5-2 shows the comparison between predicted and measured methanol concentration on biodiesel. RMSEP = 206 mg kg^{-1}, maximum ABEP = 318 mg kg^{-1}, and AVRE = 9.63 % were obtained. The RPD value was 5.96, which is low but can be used for quality control application (Williams, 2001). The result showed reasonable accuracy and clearly demonstrated that the MIR spectra can be used to identify small concentrations of methanol in the biodiesel.

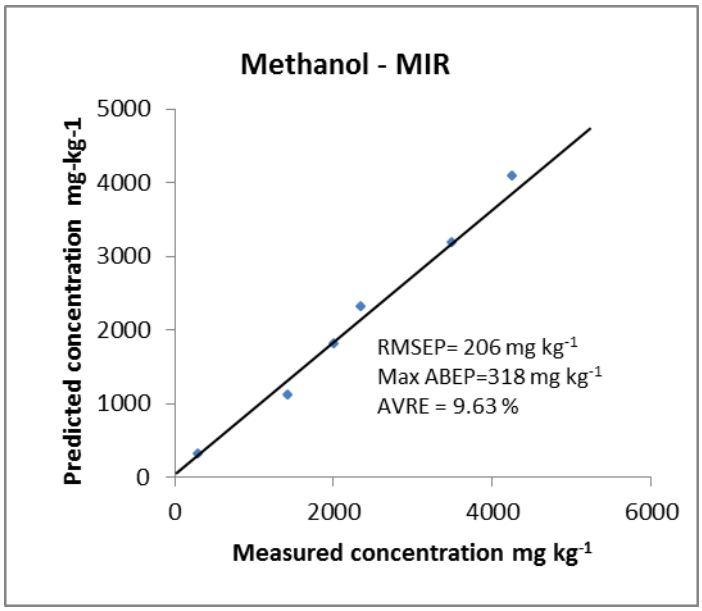

Figure 5-2 Predicted vs. measured methanol concentration in biodiesel, model by MIR

5.4.1.1.3 Water content prediction

The MIR spectral region from 3600 to 600 cm^{-1} was selected to develop the calibration model for water. The selected region includes all principal bands for water, listed above. The PLSR model was developed using the raw spectra. The correlation coefficient R^2 was 0.968, and the obtained value for SECV was 40 mg kg^{-1}. Spectral region of 3600 to 600 cm^{-1} shows at 3350 cm^{-1} information related to O-H stretching, at 1640 information about bending of H-O-H structure, and at 675 cm^{-1} information related to -OH out-of-plane bending, R. Silverstein et al. (2005). Figure 5-3 shows the relationship between predicted and measured water concentration in biodiesel. RMSEP = 68 mg kg^{-1}, maximum ABEP = 92 mg kg^{-1}, and AVRE = 12.08 % were obtained when the model was tested with the validation set. The RPD value was 6.69, which is sufficient for process control (Williams, 2001). The results demonstrated that MIR spectra can be used to identify small concentrations of water in the biodiesel.

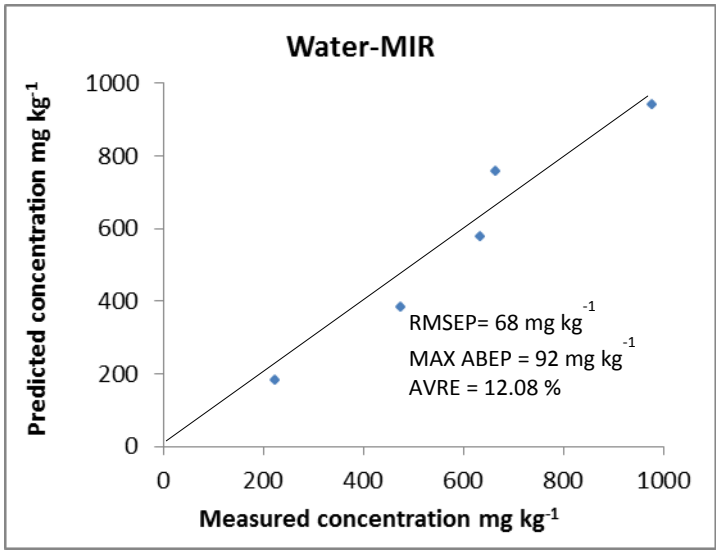

Figure 5-3 Predicted vs. measured water concentration in biodiesel, model by MIR.

5.4.1.1.4 Triglycerides content prediction

The entire MIR spectral region from 3900 to 650 cm^{-1} was used to develop the calibration model for triglycerides content in biodiesel. Since a triglycerides spectrum presents small

differences when it is compared with its respective methyl ester spectra, the PLSR model was used to develop the calibration. The correlation coefficient $R^2 = 0.994$, and SECV = 100 mg kg^{-1} were obtained. Spectral region of 3900 to 650 cm^{-1} shows at 2932 cm^{-1} information related to C-H asymmetric stretching with $–CH_2$ structure, at 2880 cm^{-1} information about C-H symmetric stretching with $–CH_3$ structure, at 1740 cm^{-1} information about C=O stretching, at 1416 information about bending vibration with =CH structure, and at 1043 cm^{-1} information related to -C-O Stretching, Yaakob et al. (2010). Figure 5-4 shows the comparison between predicted and measured triglycerides concentration in biodiesel. RMSEP = 69 mg kg^{-1}, maximum ABEP = 122 mg kg^{-1}, and AVRE = 2.93 % were obtained. The RPD value was 6.55 and the model can be used in process control (Williams, 2001). The result shows very good accuracy and clearly demonstrates that the MIR spectra can be used to identify small concentrations of triglycerides in biodiesel when only one impurity is present.

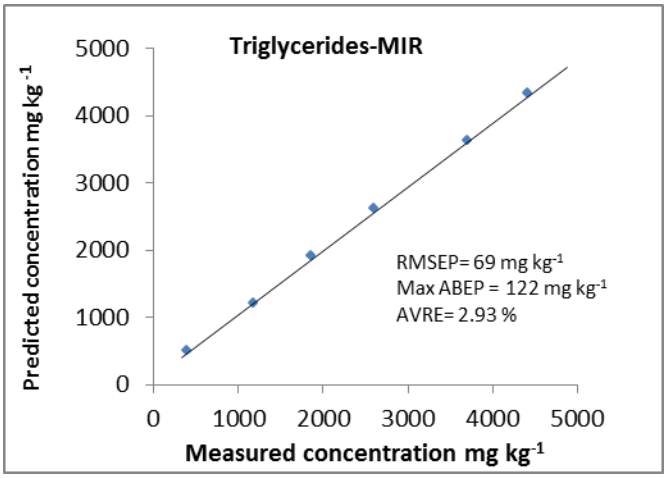

Figure 5-4 Predicted vs. measured triglycerides concentration in biodiesel, model by MIR

5.4.1.1.5 Glycerol content prediction

Considering that glycerol has absorbance bands in the same region as methanol, similar MIR regions from 3600 to 802 cm^{-1} were used to develop the calibration model to predict glycerol concentration in biodiesel, using a PLSR method. The correlation coefficient $R^2 = 0.865$,

and SECV = 56 mg kg^{-1} were obtained. Figure 5-5 shows the comparison between predicted and measured glycerol concentrations in the biodiesel. RMSEP = 56 mg kg^{-1}, maximum ABEP = 106 mg kg^{-1}, and AVRE = 25.66 % were obtained. The RPD value was 2.46, which is poor and the calibration only can be used for very rough screening (Williams, 2001). The result showed poor accuracy, thus demonstrating that the MIR spectra may not be used to identify small concentrations of glycerol in the biodiesel. When methanol and glycerol spectra were compared, glycerol spectra had many more absorption bands than the methanol spectra in the MIR range, especially in the fingerprint area (1400 to 900 cm^{-1}). In other words, too much information in a narrow region may represent a difficulty when collecting information about specific compounds.

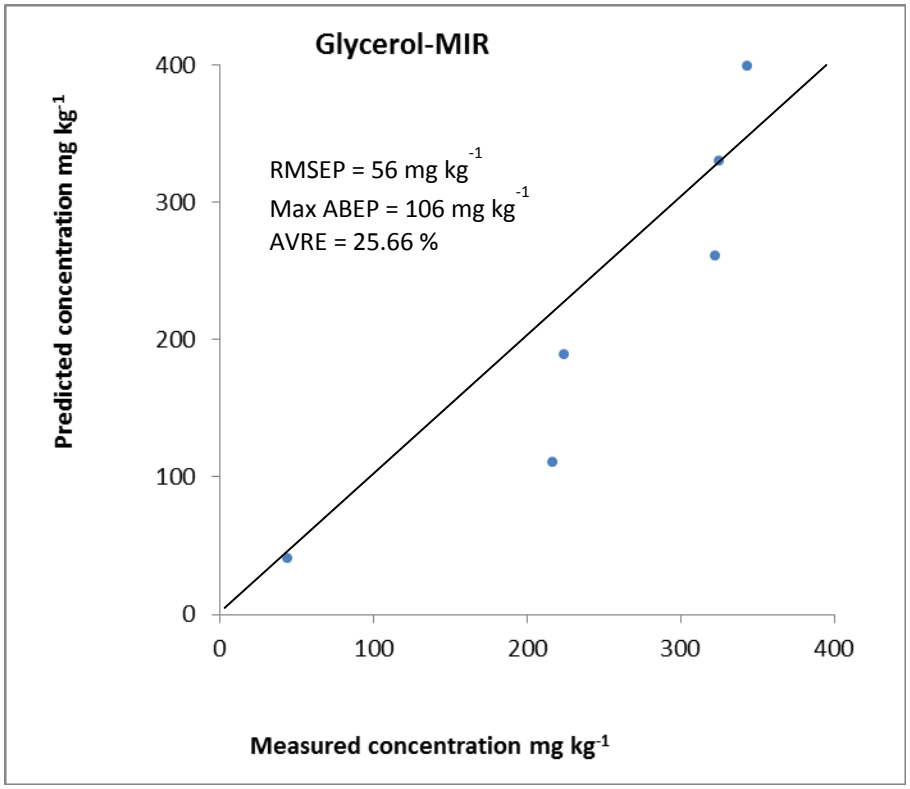

Figure 5-5 Predicted vs. measured glycerol concentration in biodiesel, model by MIR.

Regression coefficients for the calibration models are shown in Figure 5-6. Even though the plots look noisy, a few bands were clearly identified and assigned to the impurities. The

model for methanol shows coefficients near 2945 cm^{-1} and 1030 cm^{-1}, which are methanol absorption bands. The model for water shows coefficients near 1640 cm^{-1}, which are water absorption band. The model for triglycerides shows coefficients near 1740 cm^{-1} and 2932 cm^{-1}, which are triglycerides absorption bands. Finally the model for glycerol shows coefficients near 2930 cm^{-1}, which are glycerol absorption band.

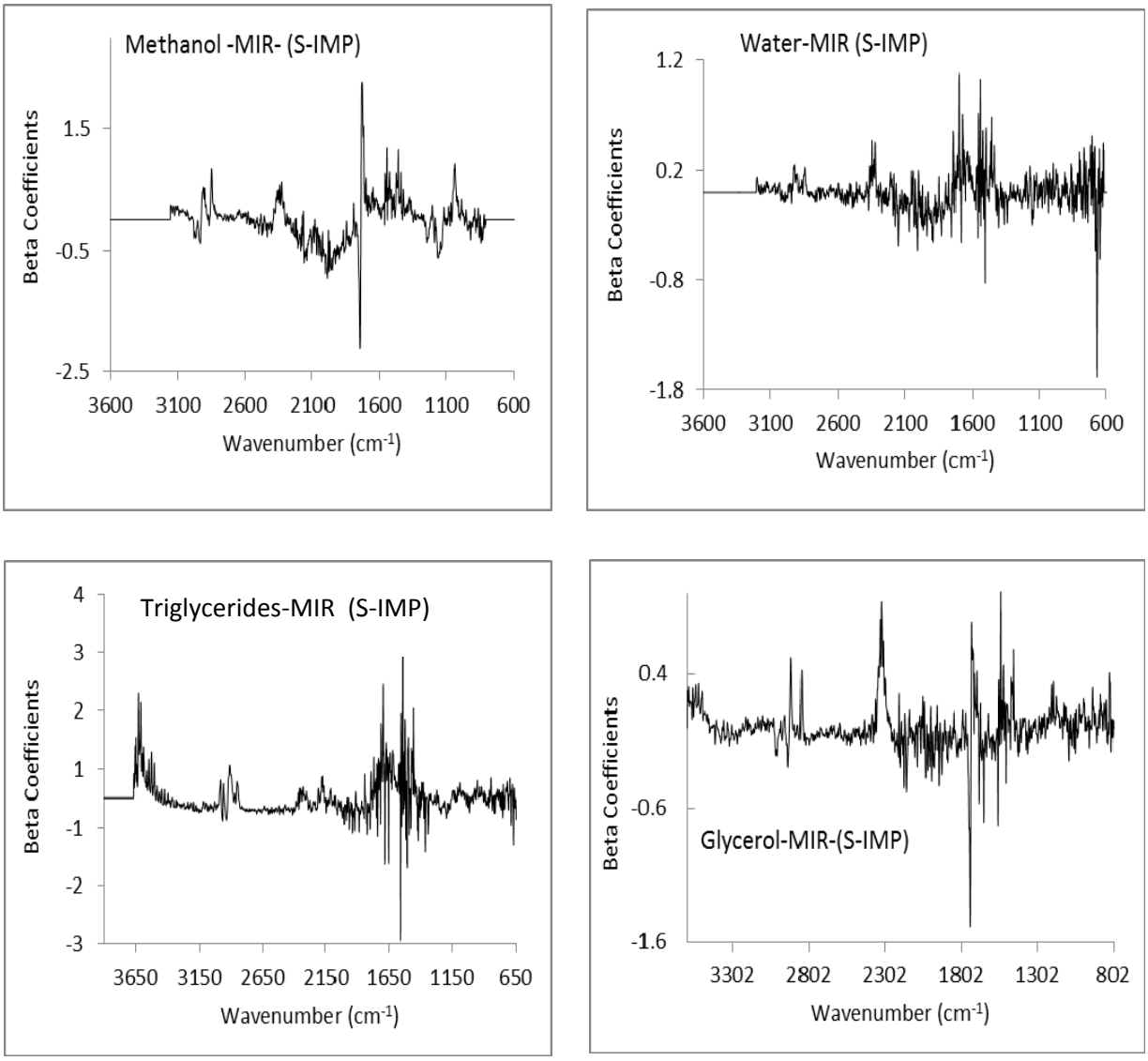

Figure 5-6 Regression coefficients for the models (S-IMP)

5.4.1.2 Prediction using NIR spectroscopy

5.4.1.2.1 Raw NIR spectra

Raw spectra of listed impurities are shown in Figure 5-7. Very distinctive peaks were observed around 1450 and 1940 nm which is characteristic of water bands. The typical absorption bands for methanol are shown near to 1582, 1698, 2068, and 2275 nm. Glycerol shows absorption bands around 1582, 1708, 2090, and 2280 nm. Triglycerides spectra shows absorption bands around 1209, 1724, 1762, and 2144 nm; in this particular case, small differences were observed between the triglycerides spectra and its correspondent methyl ester spectra. The selected calibration ranges were chosen based in spectra sections where the impurity has absorption bands with low or no absorption for biodiesel. The triglyceride model was the exception case.

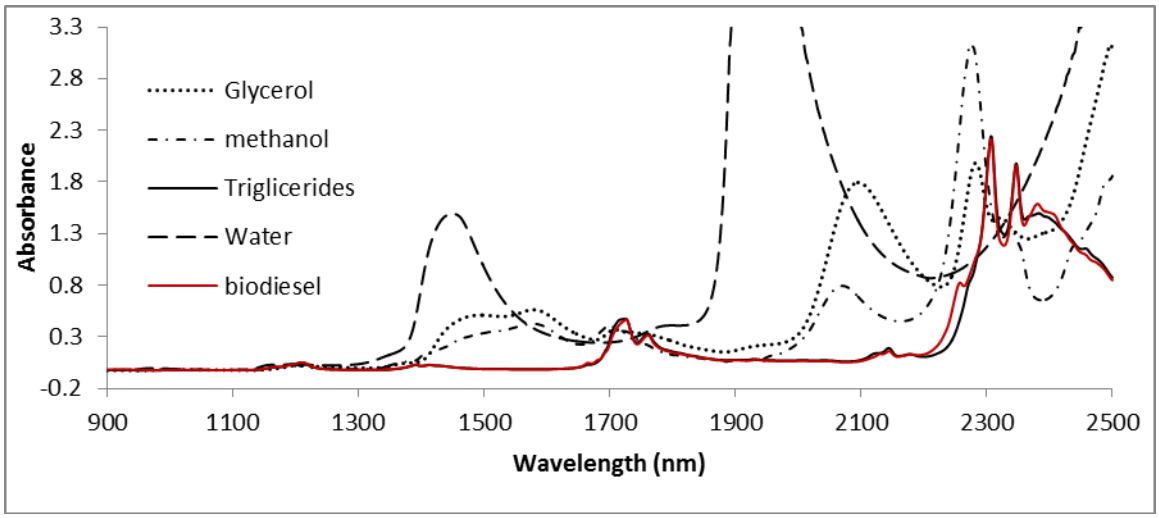

Figure 5-7 The near infrared spectra of pure glycerol, methanol, triglycerides, and water

5.4.1.2.2 Methanol content prediction

The NIR spectral region from 1100 to 2200 nm was selected to develop the calibration model for methanol. This region includes all predominant bands for methanol, listed above, with the exception of band at 2275 nm; due to the fact that biodiesel also has strong absorbance in this area. The PLSR model was developed using Grams software. The correlation coefficient R^2 = 0.960, and SECV = 250 mg kg^{-1} were obtained. No outliers were detected and no pre-treatment

data was necessary to obtain accurate results. Spectral region 1100 to 2200 nm shows at 1582 indications of O-H stretch first overtone, at 1698 nm indications of C-H stretch first overtone with $-CH_3$ structure, and at 2068 nm indications of O-H combination band, J. Shenk et al. (2008). Figure 5-8 shows the comparison between predicted and measured methanol concentration in biodiesel. RMSEP = 125 mg kg^{-1}, maximum ABEP = 233 mg kg^{-1}, and AVRE = 5.84% were obtained. The RPD value was 5.01, which is good and the calibration can be used in quality control application (Williams, 2001). The results obtained are comparable to those reported on the review studies. The NIR model performs slightly better than similar models developed using the MIR spectra.

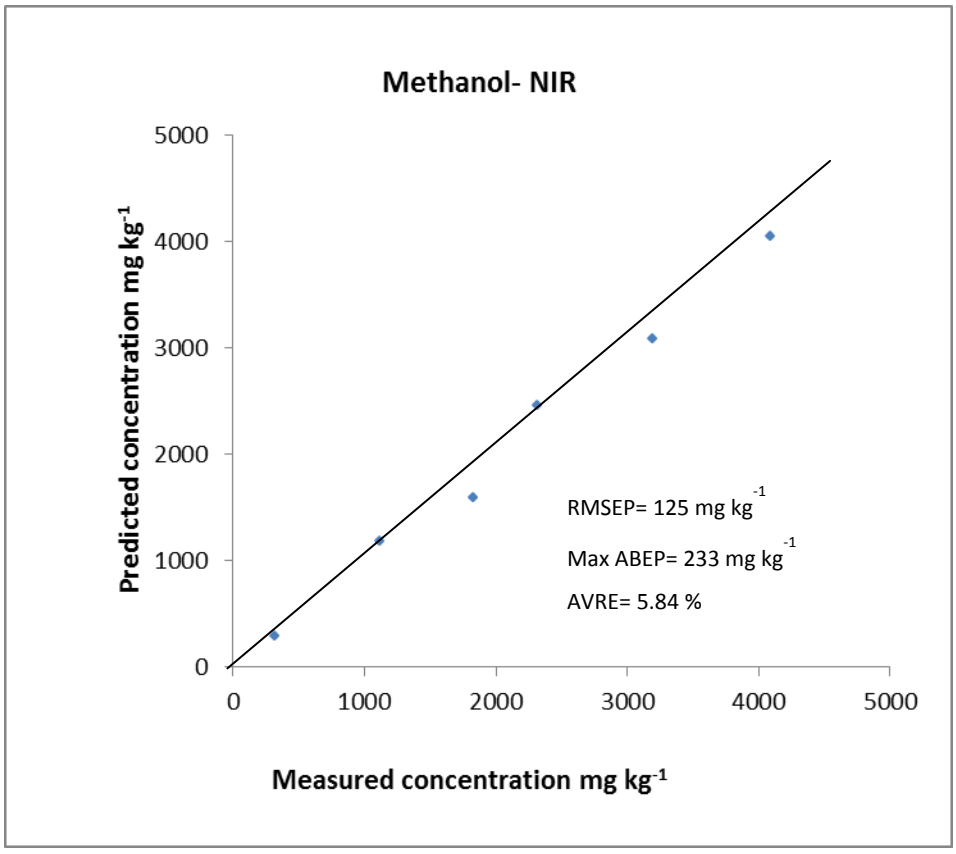

Figure 5-8 Predicted vs. measured methanol concentration in biodiesel, model by NIR

5.4.1.2.3 Water content prediction.

Spectral regions from 1300 to 1500 nm and from 1800 to 2000nm were selected to develop the calibration model for water in the NIR range. These regions include the predominant bands for water in the NIR range. The PLSR model was developed using the raw spectra. The correlation coefficient $R^2 = 0.972$, and SECV = 40 mg kg^{-1} were obtained. Between the mentioned spectral regions bands at 1450 gave indications of -OH stretching vibration (first overtone), and at 1940 nm indications of -OH second overtone of bending vibration, J. Shenk et al. (2008). Figure 5-9 shows the comparison between predicted and measured water concentration in biodiesel. RMSEP = 49 mg kg^{-1}, maximum ABEP = 98 mg kg^{-1}, and AVRE = 8.63 % were obtained. The RPD value was 6.69, which is reasonable and the calibration can be used in process control (Williams, 2001). The results obtained present good accuracy. The NIR model performs slightly better than similar models developed using the MIR spectra for water concentration in biodiesel.

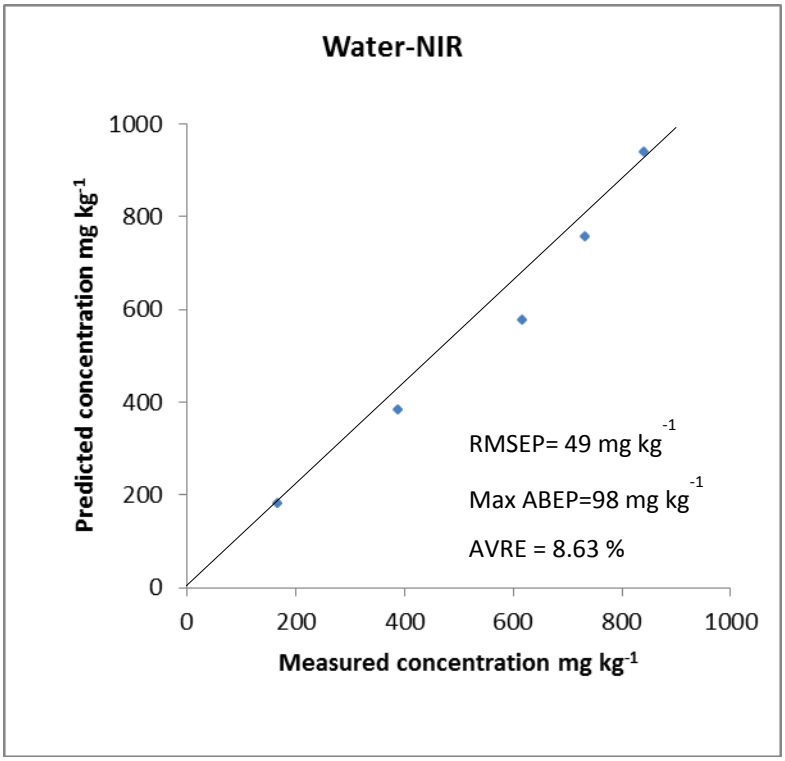

Figure 5-9 Predicted vs. measured water concentration in biodiesel, model by NIR

5.4.1.2.4 Triglycerides content prediction

Using the NIR spectral region from 1100 to 2200 nm, the calibration model for triglycerides was developed. This region includes all predominant bands for methanol, listed above, except the band at 2275 nm; due to the fact that biodiesel also has strong absorbance in this area. The PLSR model was developed using Grams software. The correlation coefficient R^2 = 0.960, and SECV = 200 mg kg^{-1} were obtained. Spectral region of 1100 to 2200 nm shows at 1209 indications of -CH stretch second overtone with $–CH_2$ structure, at 1724 nm indications of -CH stretch first overtone with $–CH_3$ structure, at 1762 nm indications of -CH stretch first overtone with $–CH_2$ structure, and at 2144 nm information could be related to -CH stretch/ C=O combination band, Silverstein et al. (2005). Figure 5-10 shows the comparison between predicted and measured Triglycerides concentration in biodiesel. RMSEP = 647 mg kg^{-1}, maximum ABEP = 1442 mg kg^{-1}, and AVRE = 27.32% were obtained. The results obtained show poor accuracy. However, considering the sample with the higher absolute error as an outlier, the RMSEP, ABEP, and AVRE can be reduced to 269 mg kg^{-1}, 441 mg kg^{-1}, and 12.71%, respectively. The RPD value was 6.55. . Then, results present reasonable accuracy on the prediction. For the triglycerides case, the MIR model performs better than the similar model developed using NIR spectra.

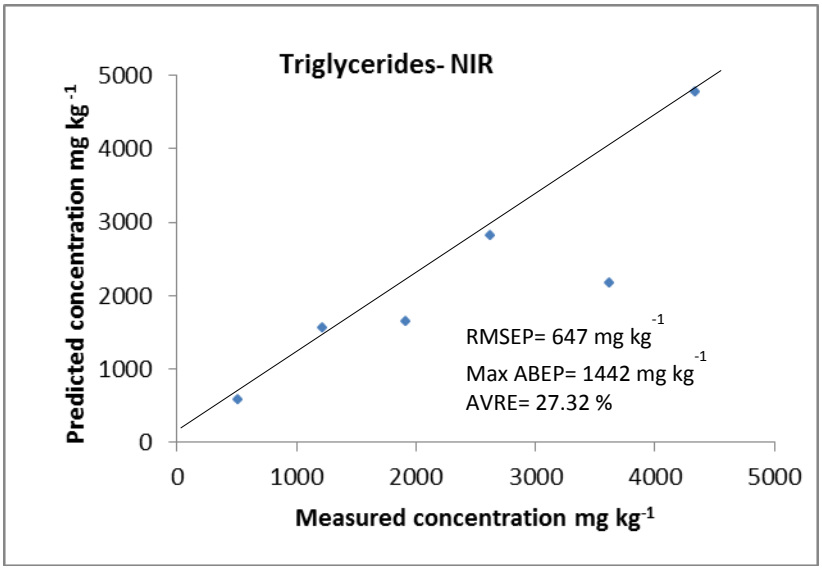

Figure 5-10 Predicted vs. measured triglycerides concentration in biodiesel, model by NIR

5.4.1.2.5 Glycerol content prediction

Similar to the other cases, the NIR spectral region from 1100 to 2200 nm was selected to develop the calibration model for glycerol. This region includes almost all predominant bands for glycerol, listed above. The PLSR model was developed using Grams software and the raw spectra. The correlation coefficient $R^2 = 0.988$, and SECV = 10 mg kg^{-1} were obtained. The RPD value was 11.40, which is excellent and the calibration can be used in quantification. The spectral region from 1100 to 2200 nm shows at 1582 indications of -OH stretch first overtone, at 1708 nm indications of C-H stretch first overtone with $–CH_3$ structure, and at 2090 nm indications of -OH combination band , Shenk et al (2008). Figure 5-11 shows the results obtained, which present good accuracy in prediction and are comparable to the results reported on the related review studies. Again, the NIR model performs better than similar models developed using MIR spectra.

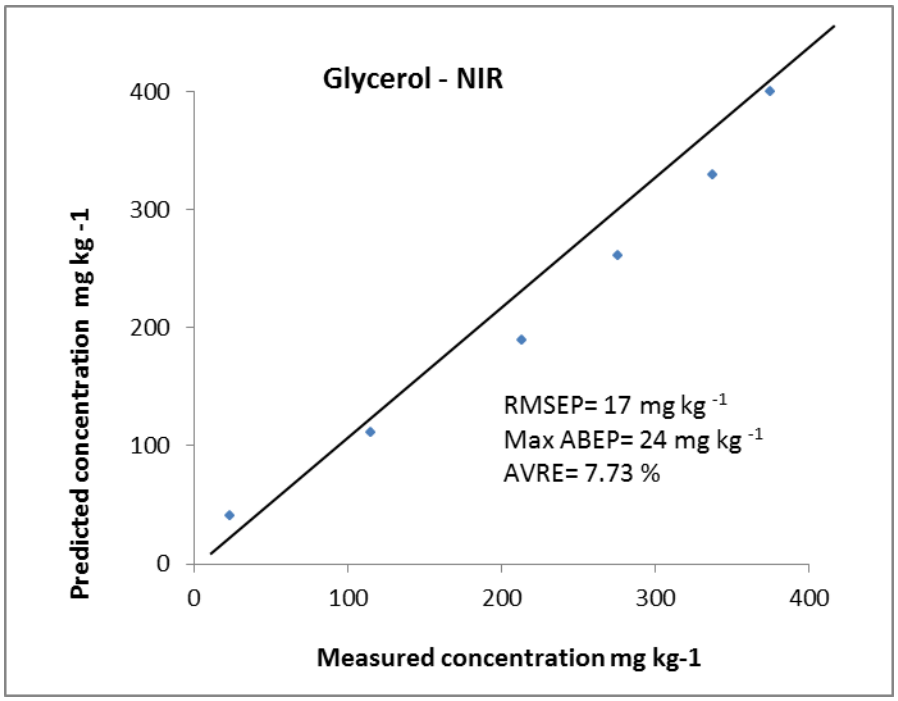

Figure 5-11 Predicted vs. measured triglycerides concentration in biodiesel, model by NIR

Regression coefficients for the models of S-IMP using NIR spectra are shown in Figure 5-12. Regression coefficients presented in the graphic indicated that the calibrations are based in absorption bands related to the impurity under study. For instance, band was observed near 2040 nm, which are related to methanol. Band observed near 1950 nm corresponded to water absorption band. Band observed near 1705nm can be assigned to the triglycerides absorption band, and the band observed near 1690 nm is related to the glycerol absorption band.

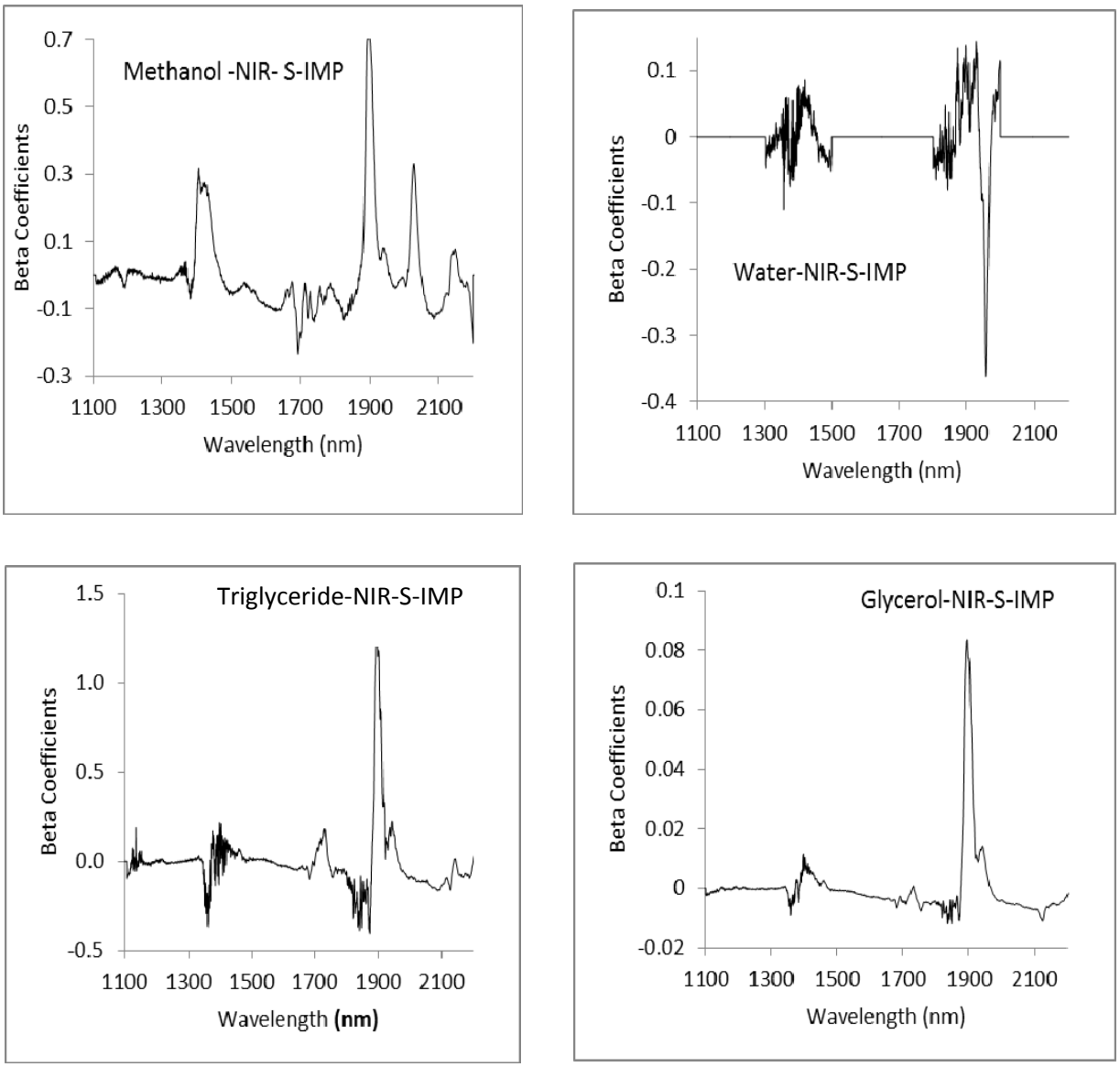

Figure 5-12 Regression coefficients for calibration of the S-IMP models using NIR spectra.

Performance factors of the developed model using MIR and NIR for impurities detection are summarized in Table 5-3. The only model that presented poor RPD value (2.46) was the model for glycerol using MIR spectra, the other models presented better RPD in both NIR and MIR ranges. NIR spectra yielded better results for methanol, water, and glycerol models, but MIR showed better results for the triglycerides model. However, both ranges NIR and MIR were found reasonably accurate for predicting a single impurity in biodiesel.

Table 5-3 Validation results for impurities prediction using models with a single impurity in the NIR and MIR ranges

	MIR				NIR			
	Methanol	Water	Triglycerides	Glycerol	Methanol	Water	Triglycerides	Glycerol
R^2	0.977	0.968	0.994	0.865	0.960	0.972	0.960	0.988
SECV $mg\ kg^{-1}$	210	40	100	50	250	40	200	10
RMSEP $mg\ kg^{-1}$	206	68	69	56	125	49	647	17
Max ABEP $mg\ kg^{-1}$	318	92	122	106	233	98	1442	24
AVRE %	9.63	12.08	2.93	25.66	5.84	8.63	27.32	7.73
RPD	5.96	6.69	13.22	2.46	5.01	6.69	6.55	11.40

5.4.2 Model with all impurities (A-IMP)

When all impurities were present and the raw spectra were used, the developed analysis for quantifying impurities in biodiesel showed poor accuracy in prediction for both NIR and MIR ranges. These results were expected considering that similar ranges in NIR and MIR spectra

were used to identify different impurities when a single impurity was present in biodiesel. For these reasons, narrow ranges for calibration and derivative techniques were pre-evaluated as a possible option to develop all the impurity models. Based on the fact that biodiesel spectra will be mainly constituted of combination bands when all impurities are present, the derivative technique using NIR spectra could be a better option to seek information about the major constituents. MIR spectra were not included in the model that was used to test for all impurities because the distinctive bands from methanol, glycerol, and water are very close in the raw MIR spectra, making it difficult to see the band separation using the derivative technique.

5.4.2.1 Prediction using NIR spectroscopy and second derivative approach

5.4.2.1.1 Second derivative spectra

Derivative spectra techniques can be used to perform the band separation, Huguenin et al. (1986). This can be first or higher order. The second order derivative spectra are commonly used, because the band intensity and peak location are maintained as in the original spectra. This characteristic could be very useful when identifying the needed original compounds in combination bands. To determine the appropriate range for calibration for each model derivative technique was developed, a raw spectrum of biodiesel containing all impurities and raw spectra of methanol, triglycerides, water, and glycerol in the NIR range are shown in Figure 5-13. The spectrum of biodiesel with all impurities was taken as a result of the overlapping of impurities and biodiesel (without impurities) absorption bands. Figure 5-13 shows the second derivative through the fourth derivative of the spectrum of biodiesel with all impurities. The spectra were calculated using Savitzky-Golay's second derivative with thirty-one point. Based on the results of the derivative analysis, the possible absorption bands for each impurity in the second derivative spectra were identified. The bands around 1200, 1410, and 1910 nm in the second derivative were related to methanol presence. For water, the selected bands were 1145, 1450, and 1975 nm. The bands selected for glycerol were around 1590 and 2090 nm, and for triglycerides the used bands were 1165, 1210, and 1400 nm. Using the information above, described ranges for calibration were specified for each model. Near infrared data, pre-treatment by Savitzky-Golay's second derivative, and partial least square regression (PLSR) method were used for the calibration model of each impurity.

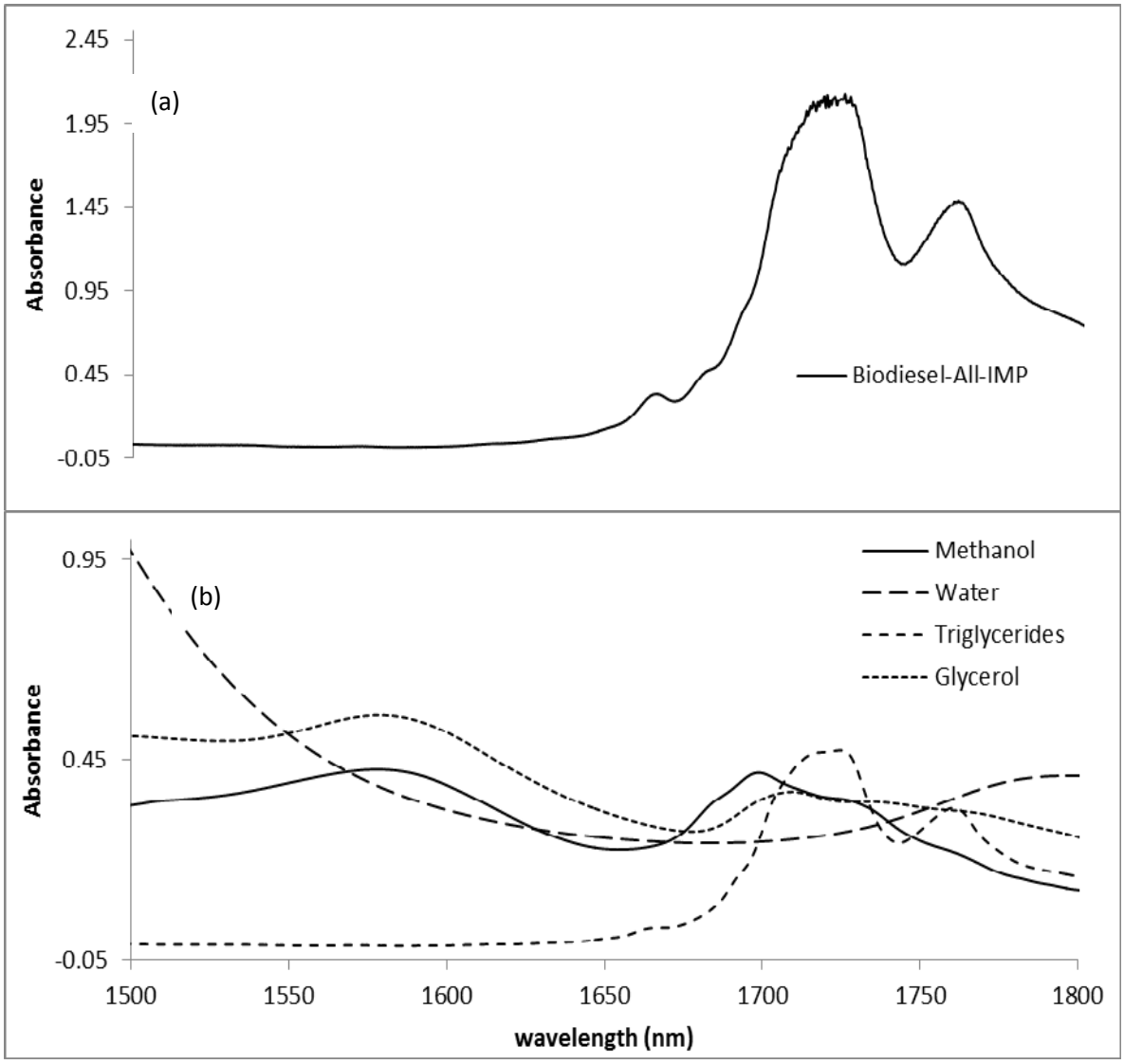

Figure 5-13 Raw spectra of (a) biodiesel with all impurities, and (b) methanol, water, triglycerides and glycerol.

The zero crossing in the 3rd derivative spectra detects the band center position in the transmittance spectrum of the original component of the band, Morrey (1968). The analysis to specify the calibration range for each impurity was developed over the entire NIR range.

However, in Figures 5-13 and 5-14 only, the range from 1400 to 1800 nm was presented to clarify information.

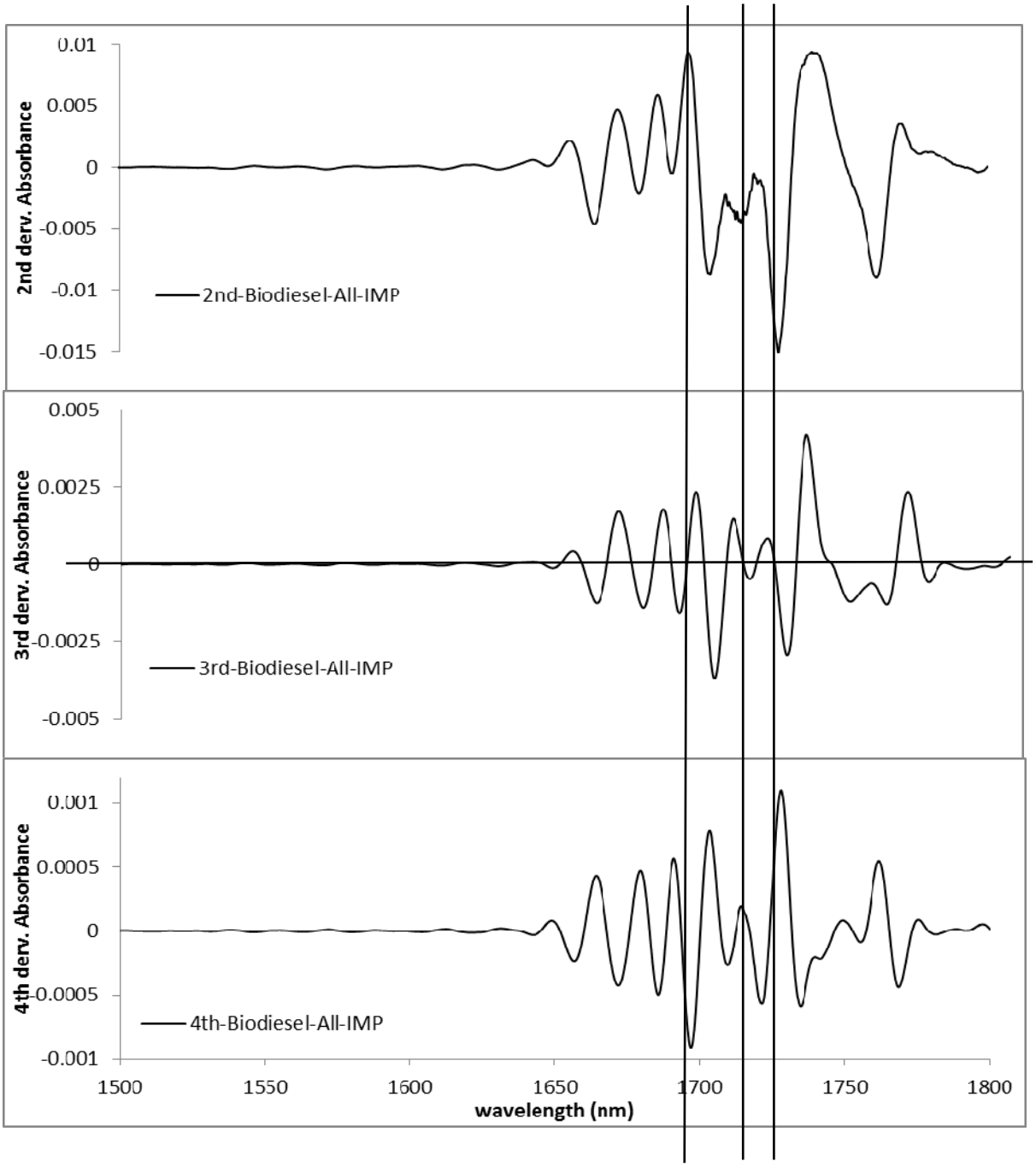

Figure 5-14 The second (top) through fourth (bottom) derivative spectra of biodiesel with all impurities. The vertical lines represent the absorption band central position of impurities.

5.4.2.1.2 Methanol content prediction.

Using the information from the derivative analysis the regions from 1100 to 1650 nm and from 1800 to 2150 nm were chosen to develop the calibration models using the PLSR method with second derivative of Savitzky-Golay with 31 point as pre-treatment. The correlation coefficient $R^2 = 0.934$, and SECV = 330 mg kg^{-1} were obtained with a number factor of = 5. Figure 5-15 shows the comparison between predicted and measured methanol concentration on biodiesel using A-IMP type model. RMSEP = 177 mg kg^{-1}, maximum ABEP = 410 mg kg^{-1}, and AVRE = 6.86 % were obtained. The results obtained present reasonable accuracy when they are compared with the results of models developed for a single impurity.

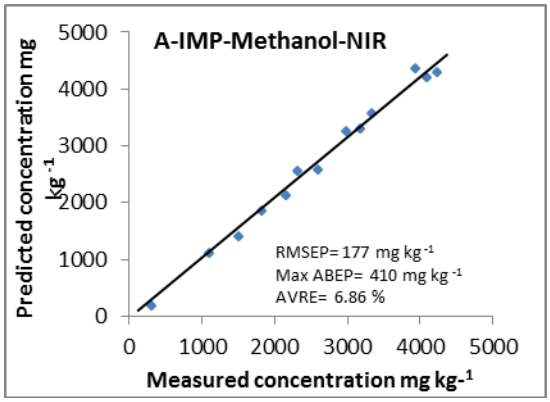

Figure 5-15 Predicted vs. measured methanol concentration in biodiesel, A-IMP model by NIR

5.4.2.1.3 Water content prediction.

Using the same procedure that was used to select the calibration ranges in the previous model; the regions from 950 to 1150 nm, from 1400 to 1650 nm, and 1800 - 2050 nm were chosen to develop the calibration models based in the PLSR method with a second derivative approach by Savitzky-Golay with 31 points as pre-treatment. The results of calibration showed the following values for number of factor, correlation coefficient (R^2) and SECV, 6, 0.918 and 100 mg kg^{-1} respectively. Figure 5-16 shows the comparison between predicted and measured water concentration in biodiesel using A-IMP type model. RMSEP = 93 mg kg^{-1}, maximum ABEP = 201 mg kg^{-1}, and AVRE = 12.06 % were obtained. The results obtained present

reasonable accuracy in prediction, when the model is compared with the model developed to detect water in biodiesel for a single impurity. Little influence from the other impurities could be implied.

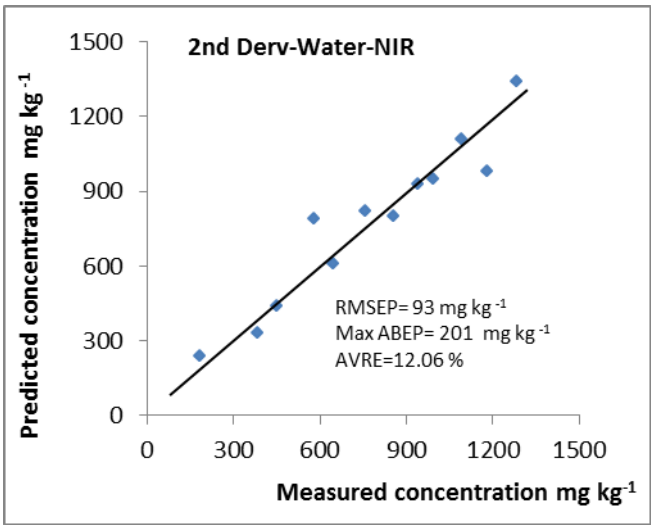

Figure 5-16 Predicted vs. measured water concentration in biodiesel, A-IMP model by NIR

5.4.2.1.4 Triglycerides content prediction

The regions from 900 to 1250 nm and from 1300 to1500 nm were chosen to develop the calibration models based on the PLSR method with second derivative approach by Savitzky-Golay with 31 points. Calibration results showed values for correlation coefficient $R^2 = 0.869$, SECV= 470 mg kg^{-1}, and for number of factors = 5. Figure 5-17 shows the comparison between predicted and measured triglycerides concentration in biodiesel using A-IMP type model. RMSEP = 932 mg kg^{-1}, maximum ABEP = 1680 mg kg^{-1}, and AVRE = 35.03 % were obtained. The poor accuracy in prediction presented for this model suggested that triglycerides may not be detected by the NIR and second derivative method when many impurities are present in the sample of biodiesel. When the second derivative spectra were analyzed in detail, strong influences from the other impurities were detected close to the bands used to detect triglycerides.

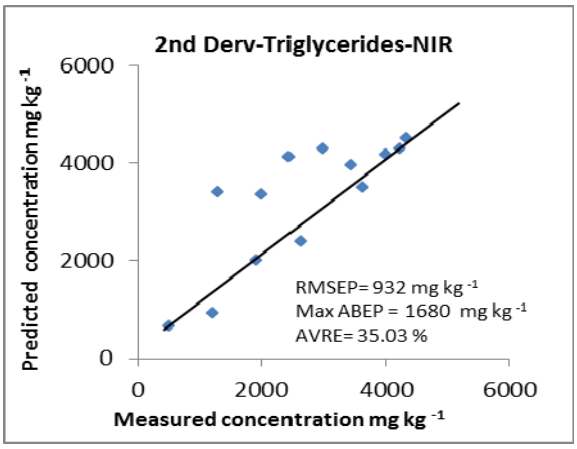

Figure 5-17 Predicted vs. measured triglycerides concentration in biodiesel, A-IMP model

5.4.2.1.5 Glycerol content prediction

Finally, the regions from 1400 to 1650 nm and from 1800 to 2100 nm were chosen to develop the calibration models for glycerol concentration in biodiesel using NIR spectra and the second derivative approach. The results for calibration were: number of factors = 6, correlation coefficient $R^2 = 0.845$, and SECV = 50 mg kg^{-1}. Figure 5-18 shows the comparison between predicted and measured water concentration in biodiesel using A-IMP type model. RMSEP = 54 mg kg^{-1}, maximum ABEP = 123 mg kg^{-1}, and AVRE = 20.04 % were obtained.

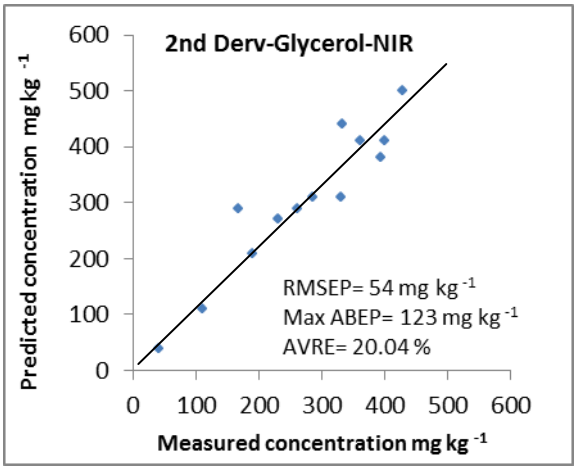

Figure 5-18 Predicted vs. measured glycerol concentration in biodiesel, A-IMP model.

Regression coefficients for the models of A-IMP using NIR spectra and 2nd derivative are shown in Figure 5-19. Regression coefficients presented in the graphic indicated that the calibrations are based in absorption bands related to the impurity under study. For instance, bands were observed near 1390 nm and 2025 nm, these bands can be assigned to methanol. Bands observed near 1430 nm and 1885 nm corresponded to water absorption bands. Bands observed near 1355 nm and 1393 nm can be assigned to the triglycerides absorption bands, and the band observed near 1650 nm is related to the glycerol absorption band.

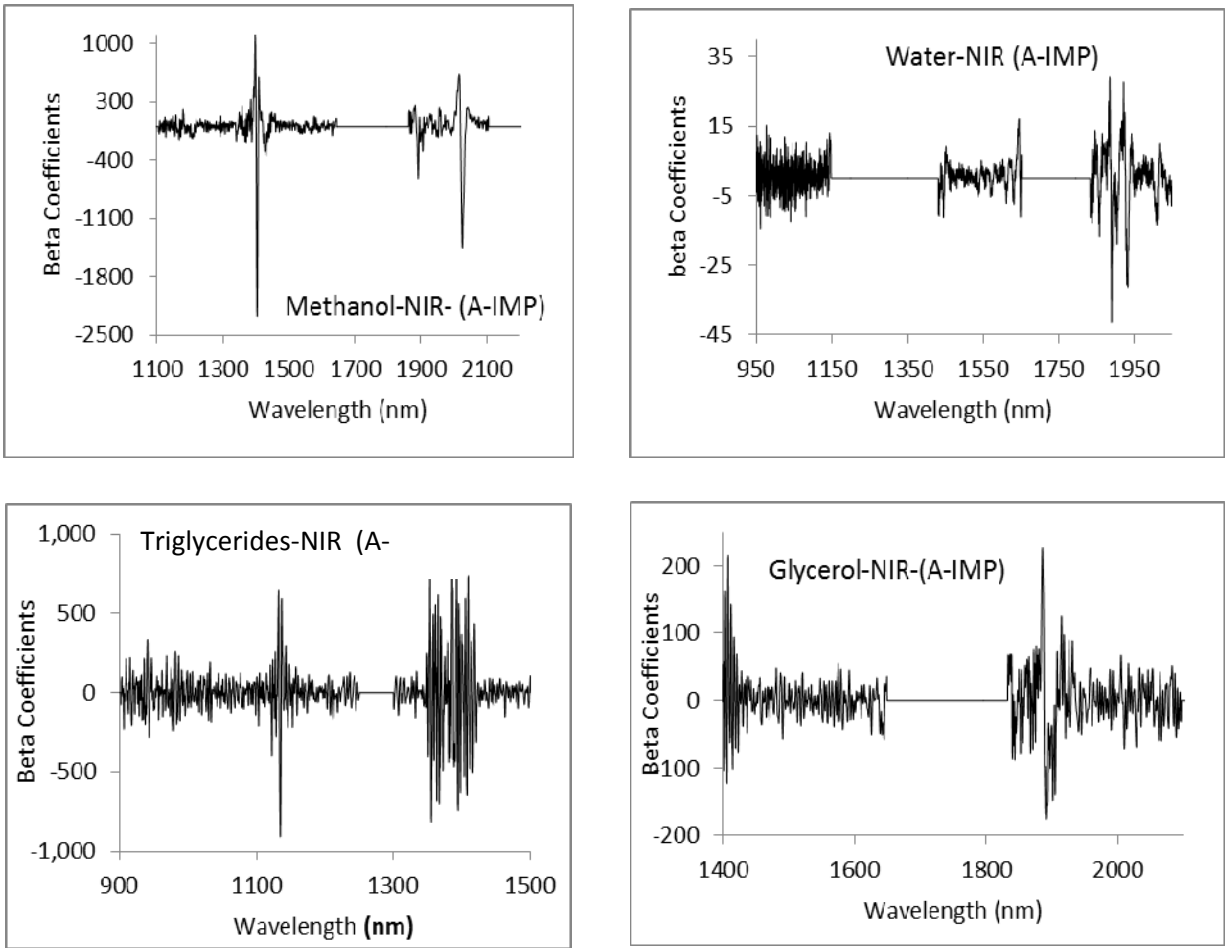

Figure 5-19 Regression coefficients for the model (A-IMP)

Results for all impurity models using the second derivative approach and NIR spectroscopy are presented in Table 5-4. The model for methanol presents reasonable performance to predict its impurity.

Independently, all impurities were present in the biodiesel sample. The model for water also presented reasonable accuracy in prediction. However, the models for triglycerides and glycerol did not perform very well. The RPD values for methanol, water, triglycerides, and glycerol models were 3.49, 3.25, 2.52,, and 2.28, respectively. According to Williams (2001), RPD value from 0 to 2.3 is very poor and the calibration use is not recommended, from 2.4 to 3.0 is poor and the calibration only can be used in very rough screening, and from 3.1 to 4.9 is fair and the calibration can be used for screening, Additional studies are necessary to improve these models, especially for triglycerides and glycerol.

Table 5-4 Validation results for impurities prediction using models with all impurities (A-IMP)

Models A-IMP, NIR	R^2	SECV mg kg^{-1}	RMSEP mg kg^{-1}	Max Absolute error of prediction ABEP mg kg^{-1}	Average relative error AVRE (%)	RPD (SD/SECV)
Methanol	0.934	330	177	410	6.86	3.49
Water	0.918	100	93	201	12.06	3.25
Triglycerides	0.869	470	932	1680	35.03	2.52
Glycerol	0.845	50	54	123	20.04	2.28

5.5 Conclusion

Models to predict impurities such as methanol, water, triglycerides, and glycerol, in biodiesel using MIR and NIR spectroscopy methods and the PLSR were developed. When only

one impurity was present in the samples, the models developed using NIR methods for methanol, water, and glycerol performed better than MIR models. They showed an average relative error (AVRE) of 5.84, 8.63, and 7.73% respectively. The model developed for triglycerides using MIR shows better performance than NIR model showing an AVRE of 2.91 %. For all impurities models developed using NIR range and second derivative as pre-treatment, only the models developed for methanol and water presented reasonable accuracy in prediction with an AVRE of 5.11 and 8.75 % mg kg^{-1} respectively. Validation results indicated that both NIR and MIR can be used to predict the concentration of impurities in biodiesel, when only one impurity is present. NIR demonstrated to be able to predict impurities such as methanol, water, triglycerides, and glycerol when more than one impurity is present. However, further work is necessary to improve the results of the models for triglycerides and glycerol.

5.6 References

ASTM Standard D6751-11b. Standard Specification for Biodiesel Fuel Blend Stock (B100) for Middle Distillate Fuels, ASTM International, West Conshohocken, PA, 2011, DOI: 10.1520/D6751-11B, www.astm.org.

Banga, S., & Varshney, P. K. (2010). Effect of impurities on performance of biodiesel: A review. Journal of Scientific & Industrial Research, 69(8), 575-579.

Brubach, J., Mermet, A., Filabozzi, A., Gerschel, A., & Roy, P. (2005). Signatures of the hydrogen bonding in the infrared bands of water. The Journal of Chemical Physics, 122(18), 184509.

European Standard EN 14214:2008: E. Automotive fuels – fatty acid methyl esters (FAME) for diesel engines – requirements and test methods. European Committee for Standardization 2008.

Felizardo, P., Baptista, P., Uva, M. S., Menezes, J. C., & Correia, M. J. N. (2007). Monitoring biodiesel fuel quality by near infrared spectroscopy. Journal of Near Infrared Spectroscopy, 15(2), 97-105.

Gaydou, V., Kister, J., & Dupuy, N. (2011). Evaluation of multiblock NIR/MIR PLS predictive models to detect adulteration of diesel/biodiesel blends by vegetal oil. Chemometrics and Intelligent Laboratory Systems, 106(2), 190-197.

Hugenin, R., & Jones, J. (1986). Intelligent information extraction from reflectance spectra - absorption-band positions. Journal of Geophysical Research-Solid Earth and Planets, 91(B9), 9585-9598.

Mittelbach, M., Remschmidt, C., Biodiesel: the comprehensive handbook, Austria, Martin Mittelbach, 2004.

Oliveira, F., Brandao, C., Ramalho, H., da Costa, L., Suarez, P., & Rubim, 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(2), 194.

Pilar Dorado, M., Pinzi, S., de Haro, A., Font, R., & Garcia Olmo, J. (2011). Visible and NIR spectroscopy to assess biodiesel quality: Determination of alcohol and glycerol traces. Fuel, 90(6), 2321-2325.

Sablinskas, V., Steiner, G., Hof, M. (2003) Chapter 6-Application. In G. Gauglitz, and T.Vo-Dinh (Eds.), Handbook of Spectroscopy (pp.89-168). Germany: WILEY-VCH.

Sagar, Naik, Meher, (2006). Technical aspects of biodiesel production by transesterification - a review. Renewable Sustainable Energy Reviews, 10(3), 248.

SDBSWeb : http://riodb01.ibase.aist.go.jp/sdbs/ (National Institute of Advanced Industrial Science and Technology, date of access).

Silverstein, R. M., Webster, F. X., Kiemle, D. J., Spectrometric Identification of Organic Compounds, New York , U.S.A., John Wiley & Son, Inc., 2005.

Shenk, J. S., Workman, J. J., Jr., Westerhaus, M. O. (2008) Application of NIR Spectroscopy to Agricultural Products. In D. Burns and E. Ciurczak (Eds.), Handbook of Near-Infrared Analysis (pp.347-386). Boca Raton, FL: Taylor &Francis Group.

Soares, I. P., Rezende, T. F., Silva, R. C., Castro, E. V. R., & Fortes, I. C. P. (2008). Multivariate calibration by variable selection for blends of raw soybean oil/biodiesel from

different sources using fourier transform infrared spectroscopy (FTIR) spectra data. Energy & Fuels, 22(3), 2079-2083.

Van Gerpen, J. Business Management for Biodiesel Producer, National Renewable Energy Laboratory, SR-510-36242, July 2004.

Williams, P. C. (2001) Chapter 8 Implementation of Near-Infrared Technology. In P. Williams, and K. Norris (Eds.), *Near- Infrared Technology* (pp.145-169). USA: American Association of Cereal Chemists, Inc.

Yaakob, B. C. M., Syahariza, Z. A., Rohman, A. (2010) Chapter 1. Fourier Transform (FTIR) Spectroscopy: Development, Techniques, and Application in the Analyses of Fats and Oil . In O. J. Rees(Eds.), Fourier Transform Infrared Spectroscopy: Development, Techniques, and Application (pp. 1-26).New York, NY: Nova Science.

Chapter 6 - Determining the Fatty Acid Composition of Biodiesel Using FTIR-NIR Spectroscopy with Derivative Technique

6.1 Abstract

This work reports the use of near-infrared (NIR) spectroscopy to predict the fatty acid composition of biodiesel. The spectra were pre-treated with Savitzky-Golay second derivative. A partial least square regression (PLSR) method was employed to develop a calibration model based on information from five pure fatty acids methyl palmitate C16:0, methyl stearate C18:0, methyl olate C18:1, methyl linolate C18:2, and methyl linolenate C18:3 and six different types of biodiesel. The ranges from 1625 to 1785 nm and from 2100 to 2200 nm were chosen to develop the models. The root mean square error of prediction (RMSEP) for C16:0, C18:0, C18:1, C18:2, and C18:3 were 1.62, 1.37, 1.03, 0.85 and 0.69 %(w/w), respectively, based on the validation set of 20 samples, eleven from combinations of biodiesel used in the calibration and nine from combinations of biodiesel that were not used in the calibration. The effect of impurities in biodiesel on fatty acid prediction was also evaluated. When biodiesel samples contaminated with trace amounts of methanol, triglycerides, water, and glycerol were tested, the RMSEP for C16:0, C18:0, C18:1, C18:2, and C18:3 were 2.43, 1.40, 1.73, 1.58, and 0.63 %(w/w), respectively, indicating impurities did not have a significant effect on the accuracy of the models. The results showed NIR and derivative techniques can be used to accurately predict the concentration of the five fatty acid methyl esters in biodiesel, although the models for C16:0 and C18:0 were less accurate.

6.2 Introduction

Biodiesel is chemically known as a mix of mono-alkaly ester of long chains of fatty acids. Biodiesel can be derived from vegetable oil or animal fat (triglycerides). The most common fatty esters present in biodiesel are palmitic acid, stearic acid, oleic acid, linoleic acid, and linolenic acid. When triglycerides have been transesterified with methanol, the biodiesel is usually referred to as fatty acid methyl ester (FAME). The fatty acid profile of the fuel is mentioned as an important factor in the determination of biodiesel properties and characteristics.

Because of its significance, the determination of fatty acids concentration in biodiesel is an important activity, when the quality of the fuel is monitored. Extensive work related to the effects of the fatty acids profile on the performance of the fuel, has been found, G. Knothe (2008), Falk, (2004). ASTM standard D 6751 presents the characteristics that the fuel has to meet in order to be used, with no direct regulations about fatty acid profile. The European standard (EN14214) limits the concentration of linolenic acid methyl ester for biodiesel to 12 % (m/m).

Successful applications of infrared spectroscopy have been reported to determine the fatty acid profile of substances and biodiesel using near infrared (NIR), as can be seen in the following studies. I. Gonzales et al. (2003), H. Azizian, et al (2010), N. Prieto, et al. (2012), P. Baptista (2008). However, this work was focused on the use of derivative spectra techniques to obtain the needed information from combination bands. Few works have been found related to the derivative technique, most of them are based on a theoretical approach, Huguenin, R. & Jones, J. (1986), Morrey, J. R. (1968), Tsai, F. & Philpot, W. (1998), Sato, T. (2002).

Using derivative analysis, Huguenin and Jones (1986) presented an algorithm that performs band separation from a combination band in reflectance spectra. This analysis used a spectrum of six overlapping constituent absorption bands, each being of Gaussian shape but of different widths, strengths, and degrees of overlap. This analysis was based on finding the combination where a derivative is zero for all wavelengths of the six basic constituents of the spectra. This condition was reached at the fifth derivative. The authors mentioned that the highest derivative order was necessary due to the effect of few factors that usually affect the reflectance spectra. Morrey (1968) previously developed a similar analysis using transmittance spectra. The convolution to zero was obtained in the third order derivative. In both cases, the authors match conclusions related to the error caused by an adjacent peak. They concluded that additional work must be done to develop a separation peak when they are too close. Tsai & Philpot (1998,) made a replication of the study of Huguenin and Jones (1986) and found similar conclusions related to the adjacent peak, and then developed a hypothesis that when the spectra feature of interest is larger than the band separation, it should be detected by the derivative methods. To prove his hypothesis, Tsai presented a new experiment using another five different synthetic spectra to apply the Huguenin and Jones methods. This experiment confirmed the hypothesis previously presented and concluded that the performance of this tool has a strong

influence on the selected parameter of the derivative method. The author also indicated that these must be selected for each particular set of spectra and purpose of the analysis. An estimation method for fatty acid composition in oil using NIR was also developed by Sato, T. (2002). The study presented a second derivative analysis to identify the corresponding peak for C16:0, C18:0, C18:1, C18:2, C18:3, and C22:1 fatty acids. Using IDAS software, an iterative study was developed examining the moving average (MA), the size of the derivative segments (SEG,) and the gap between derivative segments (GAP) on the second derivative. Sato tested the proposed method using the spectra of rapeseed oil as an example of mixtures of ester. The band for each fatty acid in the second derivative, for C18:3, C18;2, C18:1, C18:0, C16:0, and C22:1 were listed. The absorption bands were identified at 1708, 1712, 1724, 1730, 1728, and 1726 nm, respectively, when the MA= 4 nm, the SEG=12nm, and Gap=12 nm. Sato concludes that with this method it is possible to estimate the fatty acid profile roughly, simply, and rapidly.

The listed work shows the possibility of derivative technique methods to extract information from combination bands on the NIR spectra. Since no report was found on predicting fatty acid composition of biodiesel using NIR spectroscopy and developing band separation, the objectives of this study were to develop predicting models for fatty acid composition of biodiesel using FTIR – NIR spectroscopy and derivative techniques.

6.3 Materials and Methods

6.3.1 Samples preparation

For this study, pure fatty acid and biodiesel samples were used. Pure methyl palmitate (C16:0), methyl stearate (C18:0), methyl oleate (C18:1), methyl linoleate (C18:2), and methyl linolenate (C18:3) were purchased from NU-Chek-Prep Inc. Elysian, MN. The used biodiesels were prepared from food grade soybean oil, corn oil, canola oil, flaxseed oil, animal fat, olive oil, peanut oil, coconut oil, and a mix of peanut, olive, and soybean oil; purchased from local grocery store, and palm oil purchased from Country Soap Shack (Missouri, USA). All biodiesel samples were freshly produced in the author's laboratory using a standard base-catalyzed transterification process followed by recurrent water washing and a final drying process. The fatty acid profiles of the ten biodiesel fuels are shown in Table 6-1, which shows that the selected biodiesel samples include a broad range of fatty acids. The animal fat methyl ester (AFME) presents

abundant C16:0 and C18:1; palm methyl ester (PAME) presents high content of C16:0 and C18:1, canola methyl ester (CAME), peanut methyl ester (PEME), and olive methyl ester (OLME) are rich in C18:1, corn methyl ester (COME) and soybean methyl ester (SOME) were selected for their high content of C18:2, and flaxseed methyl ester (FXME) was included for its high content of C18:3. To increase the variability of fatty acid content a mix of olive, peanut, and soybean oil was used to prepare a mixed methyl ester (MXME.) This methyl ester presents high content of C18:1 and medium amounts of C16:0, C18:2, and C18:3. Coconut methyl ester (CCME) was used because it presents high content of C12:0 and C14:0 and low amounts of studied fatty acid.

Table 6-1 Fatty acid profile (mass %) of biodiesel samples prepared for this work.

FAME(mass%)	C8:0	C10:0	C12:0	C14:0	C16:0	C18:0	C18:1	C18:2	C18:3
COME	0.02	0.00	0.00	0.07	10.88	2.27	27.63	53.95	2.36
FXME	0.00	0.00	0.08	0.07	5.61	3.09	14.79	15.48	57.67
PAME	0.02	0.03	0.34	1.23	44.57	4.39	40.96	8.56	0.17
CAME	0.00	0.01	0.01	0.08	4.14	1.84	66.99	17.59	6.56
PEME	0.01	0.01	0.12	0.09	9.82	2.75	55.97	21.77	0.21
OLME	0.00	0.00	0.10	0.05	14.07	2.82	65.07	12.31	0.61
MXME	0.01	0.00	0.08	0.20	14.90	3.23	49.04	15.65	13.26
CCME	7.91	6.34	46.11	17.58	8.85	2.60	6.61	1.75	0.03
AFME	0.01	0.10	0.10	1.52	24.50	17.42	36.34	15.41	0.67
SOME	0.00	0.00	0.08	0.10	10.40	4.01	20.92	50.24	7.31

The calibration set was prepared with combinations of six biodiesels (COME, CAME, FXME, CCME, AFME and PAME). The concentration (% w/w) of methyl palmitate (C16:0), methyl stearate (C18:0), methyl oleate (C18:1), methyl linoleate (C18:2), and methyl linolenate (18:3) in the samples were ranged from: 5.35 to 44.57 % w/w, 1.92 to 4.39 % w/w , 14.80 to 66.21 % w/w, 8.56 to 52.21% w/w, and 0.18 to 57.67% w/w, respectively. The fatty acid profile

of each sample was calculated using the fatty acid profile of six original biodiesel samples with its corresponding mass percentage present in the sample. A total of one hundred and thirty-seven samples were prepared. From this set one hundred and twenty-six were used in the calibration and eleven samples were randomly selected to be used in the validation. To verify the robustness of the models under unknown biodiesel samples, the model was validated with a set of nine additional samples, consisting of pure biodiesel and samples prepared from biodiesel not used in the calibration (PEME, OLME, MXME, SOME). The range of fatty acids for each sample was kept within the calibration limits for each model. Additional experiments were developed to determine the effect of impurities in biodiesel over the fatty acid models. A set of samples of biodiesel from canola and corn oil, contaminated with methanol, triglycerides, water, and glycerol from other experiments were used to test the fatty acid models

6.3.2 Fatty acid profile measurement

The fatty acid profiles of nine biodiesels were obtained using gas chromatography. According to standard procedures detailed below, approximately 25 mg of biodiesel were dissolved in 4ml of benzene containing methyl-C13 internal standard. Samples were analyzed for fatty acid methyl esters using a HP 5890 GC with a FID detector and a SP-2560 capillary column (100m x .25mm x .2μ film, Supelco, Inc., Bellefonte, PA). Injection port and detector temperatures were 250^{o}C with a flow rate of 1 ml/min helium and a split ratio of 100:1. Injection volume was 1μl. Oven temperature began at 140^{o}C and increased at 2^{o}C/min to 200^{o}C then at 4^{o}C/min to 245^{o}C and held for 17 minutes.

6.3.3 FTIR-NIR spectroscopy scan

All biodiesel samples and pure C16:0, C18:1, C18:2, and C18:3 were scanned at a room temperature of 22-24^{o}C on a FT-IR/FT-NIR spectrometer (Perkin Elmer spectrum 400, Shelton, CT) for NIR range. The calibration and validation samples were placed in a quartz cuvette cell (Labomed Inc. Culver City, CA) of 5 mm pathlength and a transmittance scan was obtained. NIR spectra data were recorded as the absorbance in the wavelength range from 900 to 2500 nm at 1 cm^{-1} interval. All spectrums were recorded once for each sample and were obtained as an average of thirty-two scans. When the pure C18:0 was scanned, a near infrared reflectance accessory (NIRA) was used (Perkin Elmer, Liantrisant, UK). The spectrum was obtained in the range from1000 to 2500 nm at 1 cm^{-1} interval.

6.3.4 Calibration models

The regions between 1625 and 1785 nm and between 2100 and 2200 nm, in the NIR range were used to estimate the concentrations of C16:0, C18:0, C18:1, C18:2, and C18:3 in biodiesel samples. These regions were chosen considering the results of the derivative analysis and based on the criterion for detecting band center positions for each fatty acid. Grams/AI software version 9.1 (Thermo Fisher Scientific Inc.) and the PLSR method were used to develop the models. The performance of the models was assessed by the following statistical parameter: the correlation coefficient (R^2), standard error of cross validation (SECV), the root mean square error of prediction (RMSEP), the absolute error of prediction (AEP), the average relative error of prediction (AREP), and RPD value.

6.3.5 Second derivative technique

The bands in the near infrared region are difficult to assign to specific compounds because a single band in this region is the result of combinations of fundamental bands and overtones. Derivative techniques have been mentioned over the years as a possible solution to extract the information contained in these bands. This technique is used to correct the baseline, reduce the broad band, reduce the scattering effect, and perform the band separation, Huguenin et al. This can be first or higher order, the second order derivative spectra is commonly used because the band intensity and peak location are maintained as in the original spectra, Shenk et al. This characteristic could be very useful for the identification of the FAME present in biodiesel samples. To determine which derivative is appropriate to resolve the overlapping bands, raw transmittance spectra of a biodiesel sample (CAME) and the raw transmittance spectra of C16:0, C18:1, C18:2, and C18:3 in NIR range were used, which are shown in Figure 6-1. The spectrum of C18:0 is not shown because, it was obtained with a reflectance accessory using different range and absorbance scale. Considering the most common biodiesel contains the above listed fatty acids in more than 95%, for this analysis, the spectra of biodiesel was taken as the result of the five overlapping fatty acids absorption band. Figure 6-2 shows the second through fourth derivative of the spectra of CAME. The derivatives were calculated using Savitzky-Golay second derivative with third order polynomial and fifty-one points. According to Morrey (1968) and considering y (ν) as the raw spectra, detecting band center positions in transmittance spectra is reached when the following conditions are met.

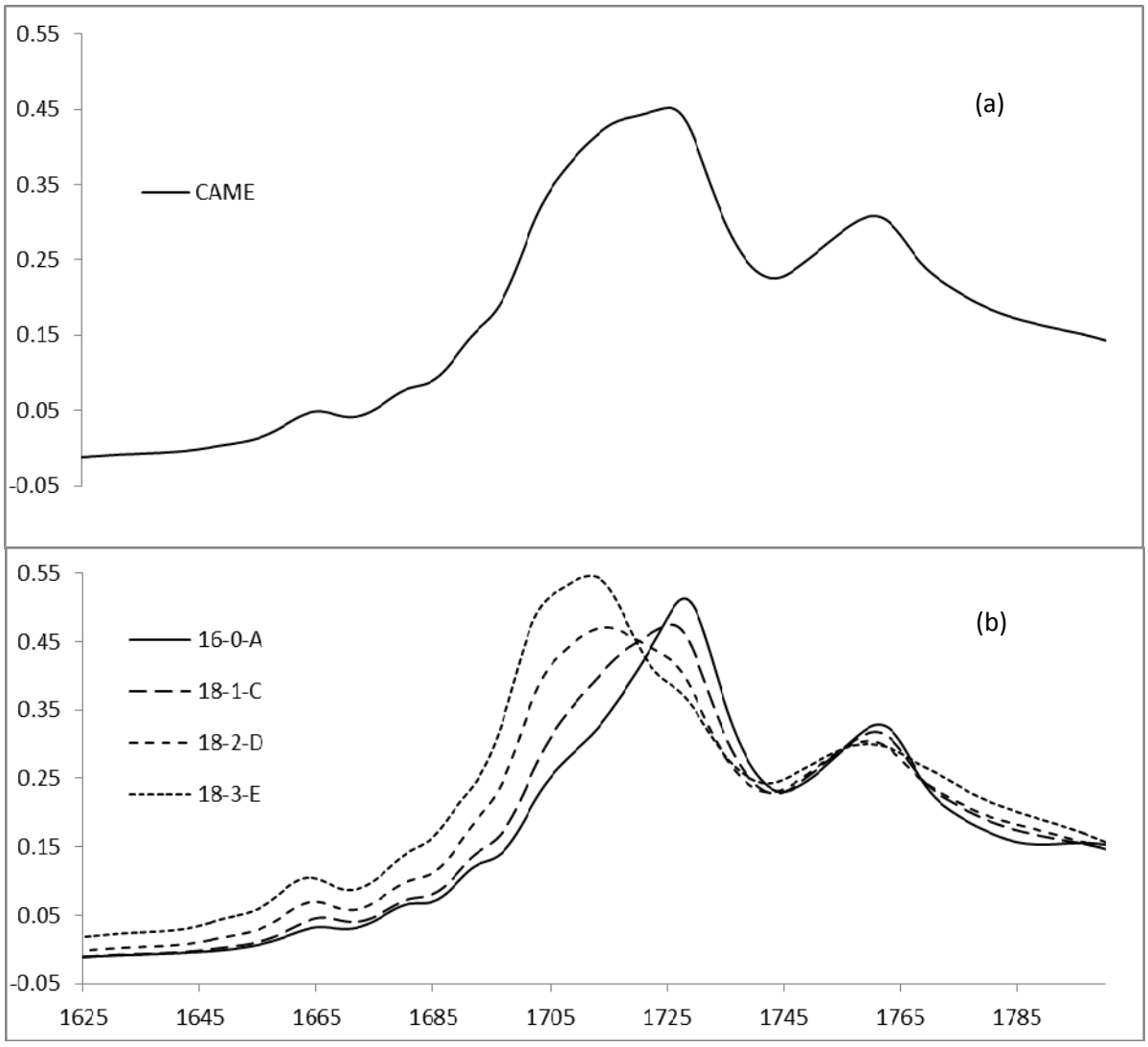

Figure 6-1 Raw spectra in specific range of NIR of (a) biodiesel from canola (CAME), and (b) pure C16:0, pure C18:1, pure C18:2, and pure C18:3.

$$y''(\nu) < 0 \qquad y'''(\nu) = 0 \qquad y''''(\nu) > 0 \qquad (1)$$

The third derivative zero crossing was used to verify the band center positions, where the band identification criterion was satisfied. The presented analysis was developed over the range between 1625 to 1785 nm, where different bands for each pure fatty acid were observed. Additionally, this is the range recommended by Sato (2008) in his analysis about identification of fatty acids in oil.

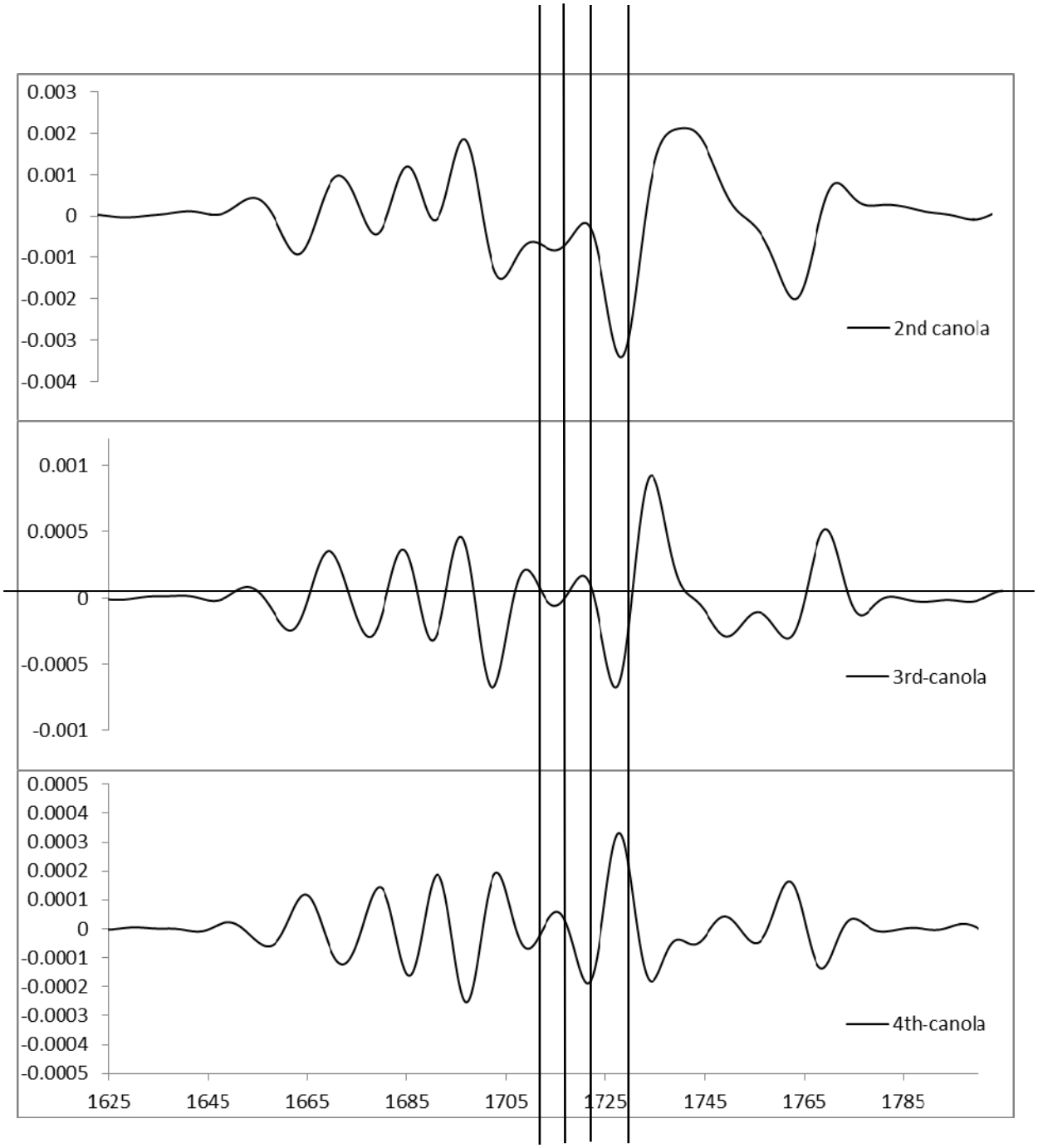

Figure 6-2 The second (top) through fourth (bottom) derivative spectra of the CAME. The four vertical lines represent the fatty acid absorption bands central position in this range.

6.4 Results and Discussion

6.4.1 Second derivative analysis

Considering the results of the derivative analysis, the absorption band for each fatty acid in the biodiesel samples was defined. Models to identify the concentration of methyl palmitate (C16:0), methyl stearate (C18:0), methyl oleate (C18:1), methyl linoleate (C18:2), and methyl linolenate (C18:3) in biodiesel samples were developed over the range from 1625 to 1785 nm and from 2100 to 2200 nm. Specific peak in the second derivative spectra was identified for C18:3 at 1710nm. For C18:2, the peak was found at 1714 nm. And for the case of C18:1, its peak was observed at 1720nm. The distinctive peaks for C18:0 were located at 1730 nm using the raw spectra. For methyl palmitate (C16:0) the peak was located at 1728 nm, because the peaks for C16:0 and C18:0 are close to the second derivative spectra they present only one peak around 1729 nm. The region from 2100 to 2200 also presents different peaks for C16:0 and C18:0 in the raw spectra. But again, when the derivative technique was applied, only one peak was observed in the second derivative spectra. However, some information was extracted when both ranges were used in the calibration. Near infrared spectroscopy data, pre-treated by Savitzky-Golay's second derivative, and partial least squares regression (PLSR) methods were used for the development of a calibration model for each fatty acid methyl ester (FAME).

6.4.2 Model to predict methyl palmitate (C16:0)

The calibration result for the model to predict methyl palmitate, represented by R^2, SECV, and number of factor were 0.92, 2.77 %(w/w) and 6, respectively. The RPD value was 3.69. When the validation set was used in the model to predict the fatty acid concentration in biodiesel, the results showed for RMSEP, Max Absolute Error, and Average Relative Error of Prediction the following values: 2.64 % (w/w), 7.01 % (w/w), and 12.45%, which are considered acceptable. Figure 6-3 shows the comparison of predicted and measured methyl palmitate from the validation set.

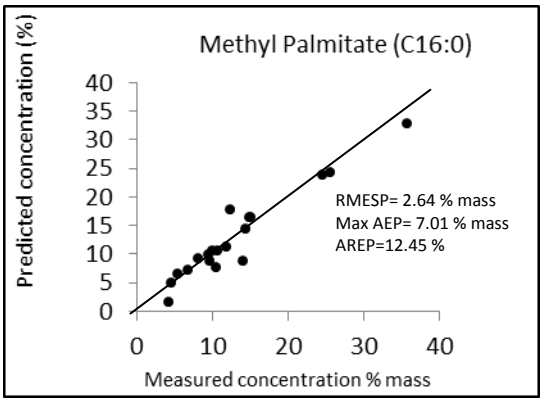

Figure 6-3 Predicted vs. measured of methyl palmitate for validation set using the NIR spectra and Savitzky-Golay second derivative.

6.4.3 Model to predict methyl stearate (C18:0)

The model to predict methyl stearate presented the following results R^2= 0.86, SECV= 1.90 % (w/w), and the number of factors was 7. The RPD value was 1.84, which 3 is very poor and the use of this calibration is not recommended. When the validation set was applied to the model for predicting the fatty acid concentration in biodiesel, the results shown for RMSEP=1.37 % (w/w), Max Absolute Error was 3.79 %(w/w), and the Average Relative error of Prediction was 25.31 %, which are considered acceptable. Figure 6-4 shows the comparison of predicted and measured methyl stearate from the validation set.

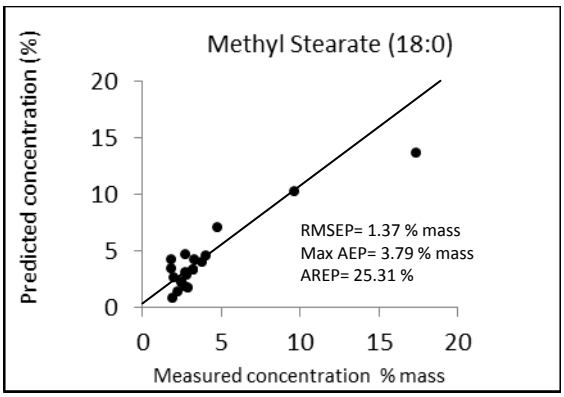

Figure 6-4 Predicted vs. measured of methyl stearate for validation set using the NIR spectra and Savitzky-Golay second derivative.

6.4.4 Model to predict methyl oleate (C18:1)

The R^2, SECV, and number of factors were 0.98, 1.74 %(w/w), and 6, respectively for the calibration model to predict methyl oleate. The RPD value was 10.90. The results of the validation process showed the following values RMSEP=1.03 % (w/w), Max AEP=2.65 % (w/w), and AREP = 1.76 %, which are considered very accurate. Figure 6-5 shows the comparison of predicted and measured methyl oleate from the validation set.

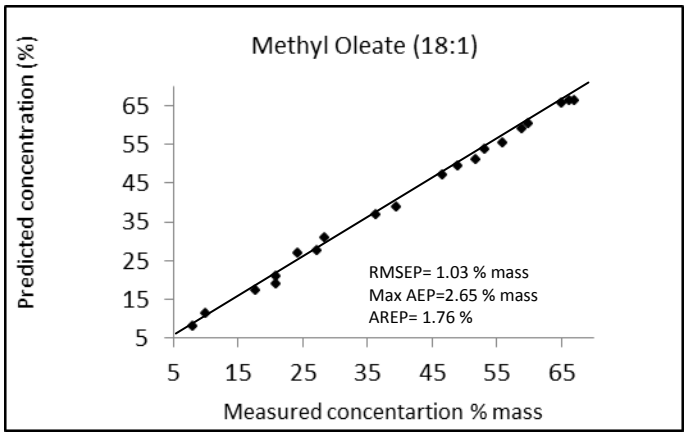

Figure 6-5 Predicted vs. measured of methyl oleate for validation set using the NIR spectra and Savitzky-Golay second derivative.

6.4.5 Model to predict methyl linoleate (C18:2)

The calibration result for the model to predict methyl linoleate, represented by R^2, SECV, and number of factors were 0.97, 2.15 %(w/w) and 6, respectively. The RPD value was 7.09, good enough for process control (Williams, 2001). When the validation set was used in the model to predict the fatty acid concentration in biodiesel, the results for RMSEP, Max Absolute Error, and Average Relative Error of Prediction showed the following values: 0.85 % (w/w), 2.34 % (w/w), and 2.84 %, which are considered very accurate. Figure 6-6 shows the comparison of predicted and measured methyl linoleate from the validation set.

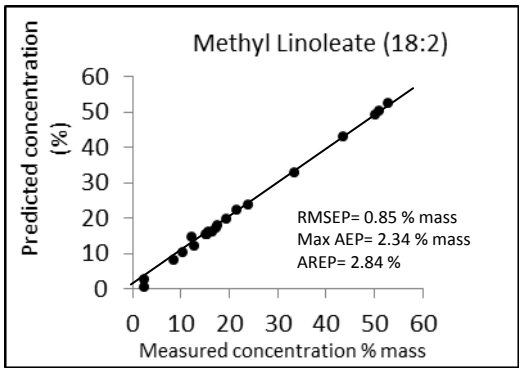

Figure 6-6 Predicted vs. measured of methyl linoleate for validation set using the NIR spectra and Savitzky-Golay second derivative.

6.4.6 Model to predict methyl linolenate (C18:3)

The calibration result for the model to predict methyl linolenate, represented by R^2, SECV, and number of factor were 0.99, 1.18 %(w/w) and 6, respectively. The RPD value was 9.66. When the validation set was used in the model to predict the fatty acid concentration in biodiesel, the results shown for RMSEP, Max Absolute Error, and Average Relative Error of Prediction were the following values: 0.69 % (w/w), 2.04 % (w/w) and 6.24 %, which are considered very accurate. Figure 6-7 shows the comparison of predicted and measured methyl linolenate from the validation set.

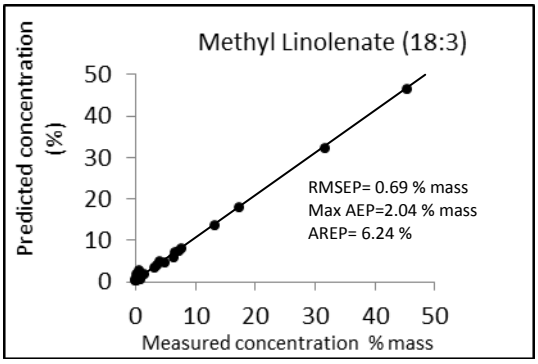

Figure 6-7 Predicted vs. measured of methyl linolenate for validation set using the NIR spectra and Savitzky-Golay second derivative.

Regression coefficients for the calibration models of fatty acid using the second derivative are shown in Figure 6-8. The models for C16:0, C18:0, C18:1, C18:2, and C18:3 show coefficients near 1730 nm which are related to pure methyl palmitate absorption band, near

1728 nm , related to methyl stearate, near 1714 nm which is related to pure methyl oleate absorption band, near 1722 nm , related to methyl linoleate, and near 1710 nm, related to methyl linolenate.

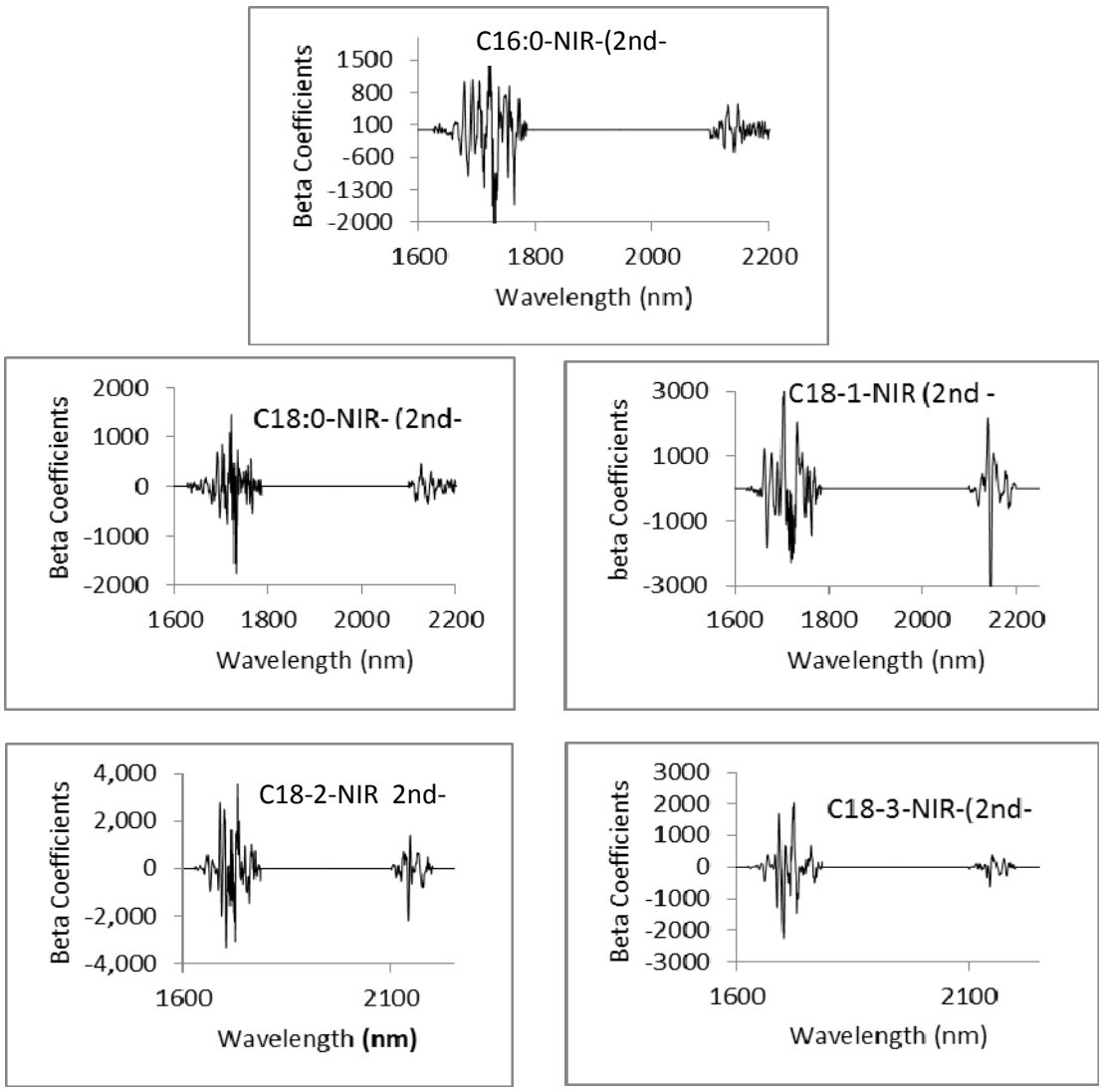

Figure 6-8 Regression coefficients for the fatty acid model using 2nd derivative

6.4.7 Effect of impurities on the fatty acid models.

To determine the effect of impurities over the fatty acid model a set of samples contaminated with impurities, from other experiments were used to test the models. A mix of biodiesels from canola oil and from corn oil contaminated with methanol, triglycerides, water,

and glycerol were used. The concentration of impurities were ranged: from 1470 to 4280 mg kg^{-1} for methanol, from 1300 to 4290 mg kg^{-1} for triglycerides, from 450 to 1290 mg kg^{-1} for water, and from 170 to 430 mg kg^{-1} for glycerol. When the described validation set was used in the model to predict the fatty acid concentration in biodiesel, the RMSEP for C16:0, C18:0, C18:1, C18:2, and C18:3 models were: 2.30, 1.40, 1.73, 1.58, and 0.63 % (w/w), respectively. No major affectations from the impurities were observed in the results, considering that RMSEP were very similar to the previous validated set (biodiesel without impurities). Figure 6-9 shows the comparison of predicted and measured methyls from the validation set.

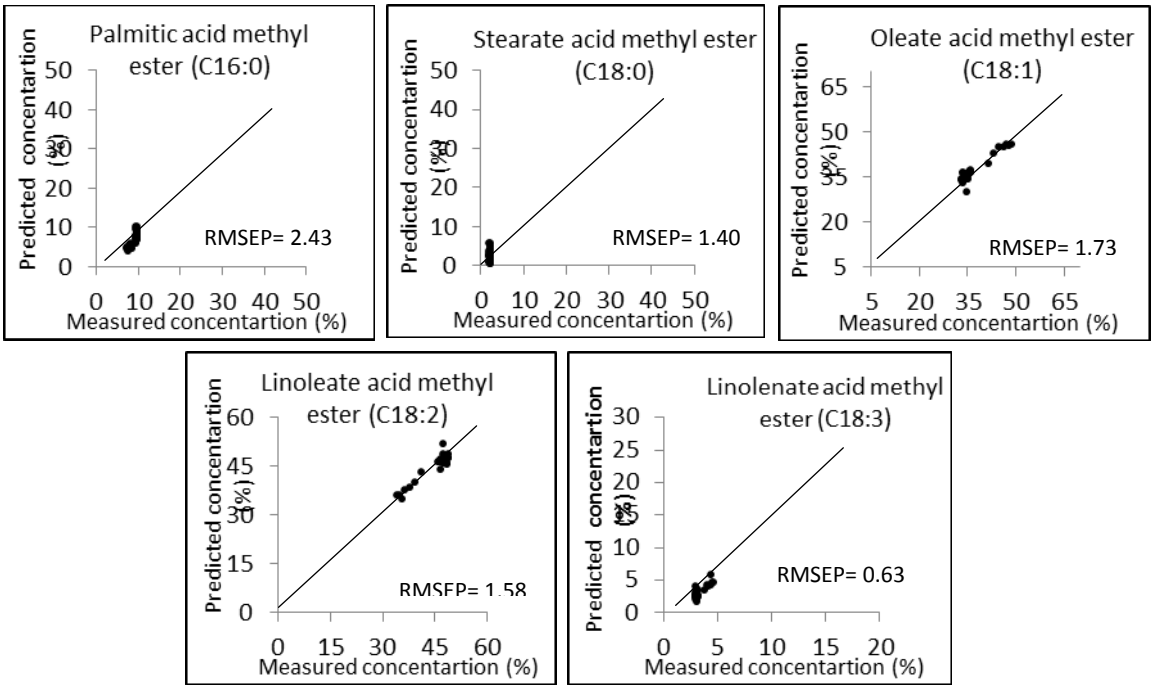

Figure 6-9 Predicted vs. measured of **C16:0, C18:0, C18:1, C18:2, and C18:3** for validation set containing impurities, using the NIR spectra and Savitzky-Golay second derivative.

The characteristics of the performance of the models that were developed using NIR and derivative techniques for concentration of C16:0, C18:0, C18:1, C18:2, and C18:3 in biodiesel are summarized in Table 6-2. The lower performance was observed in the model for C18:0, it can be attributable to the fact that biodiesel presents a short range in the concentration levels for this fatty acid.

Table 6-2 Performance factor of the models to predict fatty acid composition of biodiesel

	Model For:	Correlation coefficient (R^2)	Number of factors	(SECV) %(w/w)	(RMSEP) %(w/w)	Max. (AEP) %(w/w)	(AREP) %
Validation samples (without impurities)	C16:0	0.925	6	2.77	2.64	7.01	12.45
	C18:0	0.860	7	1.90	1.37	3.79	25.35
	C18:1	0.988	6	1.74	1.03	2.65	1.76
	C18:2	0.970	6	2.15	0.85	2.34	2.84
	C18:3	0.994	6	1.18	0.69	2.04	6.24
Validation samples (with impurities)	C16:0	0.925	6	2.77	2.43	4.01	23.67
	C18:0	0.860	7	1.90	1.40	3.37	47.68
	C18:1	0.988	6	1.74	1.73	3.22	3.43
	C18:2	0.970	6	2.15	1.58	2.92	2.65
	C18:3	0.994	6	1.18	0.63	1.46	15.00

6.5 Conclusion

Models to predict the concentration of fatty acids (C16:0, C18:0, C18:1, C18:2 and C18:3) present in biodiesel samples were developed using NIR spectra, PLSR method, and derivative technique. The regions between 1625 and 1785 nm and between 2100 and 2200 nm were chosen to develop the models. The RMSEP of the models for C16:0, C18:0, C18:1, C18:2, and C18:3 were 2.64, 1.37, 1.03, 0.85, and 0.69 % (w/w), respectively, they were slightly better that the RMSEP obtained from the previous experiments when the raw spectra were used to predict the fatty acid concentration in biodiesel. However, when the RPD values were compared, models developed with raw spectra showed higher RPD values. Derivative techniques were useful for extracting the information from combination bands in the spectra. The models developed with this technique presented accurate results in predicting C18:1, C18:2, and C18:3. Less accurate results were found in the models for C16:0 and C18:0, but the model can be considered as a promising option to predict this fatty acid.

6.6 References

Allen, . (1999). Predicting the viscosity of biodiesel fuels from their fatty acid ester composition. Fuel, 78(11), 1319.

Azizian, H., Kramer, J. K. G., & Winsborough, S. (2007). Factors influencing the fatty acid determination in fats and oils using fourier transform near-infrared spectroscopy. European Journal of Lipid Science and Technology, 109(9), 960-968.

Azizian, H., Kramer, J. K. G., Ehler, S. & Curtis J. M. (2010). Rapid quantitation of fish oil fatty acids and their ethyl esters by FT-NIR models. European Journal of Lipid Science and Technology, 112, 452-462.

Baptista, P., Felizardo, P., Menezes, J. C., & Correia, M. J. N. (2008). Multivariate near infrared spectroscopy models for predicting the methyl esters content in biodiesel. Analytica Chimica Acta, 607(2), 153-159.

Dardenne, . (2007). Prediction of fatty acid contents by mid-infrared spectrometry. Journal of Dairy Science, 90, 274.

De Marchi, M., Penasa, M., Cecchinato, A., Mele, M., & Secchiari, P. (2011). Effectiveness of mid-infrared spectroscopy to predict fatty acid composition of brown swiss bovine milk. Animal, 5(10), 1653-1658.

De Voort, , Simpson, , & Aryee, . (2009). FTIR determination of free fatty acids in fish oils intended for biodiesel production. Process Biochemistry, 44(4), 401.

Falk, . (2004). The effect of fatty acid composition on biodiesel oxidative stability. European Journal of Lipid Science and Technology, 106(12), 837.

Felizardo, P., Baptista, P., Uva, M. S., Menezes, J. C., & Correia, M. J. N. (2007). Monitoring biodiesel fuel quality by near infrared spectroscopy. Journal of Near Infrared Spectroscopy, 15(2), 97-105.

Gjerlaug Enger, E., Aass, L., Odegard, J., Kongsro, J., & Vangen, O. (2011). Genetic parameters of fat quality in pigs measured by near-infrared spectroscopy. Animal, 5(10), 1495-1505.

Gonzalez-Martin, I., Gonzalez-Perez, C., Hernandez-Mendez, J., & Alvarez-Garcia, N. (2003). Determination of fatty acids in the subcutaneous fat of iberian breed swine by near infrared spectroscopy (NIRS) with a fibre-optic probe. Meat Science, 65(2), 713-719.

Guy, F., Prache, S., Thomas, A., Bauchart, D., & Andueza, D. (2011). Prediction of lamb meat fatty acid composition using near-infrared reflectance spectroscopy (NIRS). Food Chemistry, 127(3), 1280-1286.

Hugenin, R., & Jones, J. (1986). Intelligent information extraction from reflectance spectra - absorption-band positions. Journal of Geophysical Research-Solid Earth and Planets, 91(B9), 9585-9598.

Knothe, . (2008). "Designer" biodiesel: Optimizing fatty ester (composition to improve fuel properties. Energy Fuels, 22(2), 1358.

Knothe, G. (2005). Dependence of biodiesel fuel properties on the structure of fatty acid alkyl esters. Fuel Processing Technology, 86(10), 1059-1070.

Koprna, R., Nerusil, P., Kolovrat, O., Kucera, V., & Kohoutek, A. (2006). Estimation of fatty acid content in intact seeds of oilseed rape (brassica napus L.) lines using near-infrared spectroscopy. Czech Journal of Genetics and Plant Breeding, 42(4), 132-136.

Morrey, J. R. (1968). On determining spectral peak positions from composite spectra with a digital computer. Analytical Chemistry, 40(6), 905.

Patil, A. G., Oak, M. D., Taware, S. P., Tamhankar, S. A., & Rao, V. S. (2010). Nondestructive estimation of fatty acid composition in soybean [glycine max (L.) merrill] seeds using near-infrared transmittance spectroscopy. Food Chemistry, 120(4), 1210-1217.

Prieto, N., Lopez Campos, O., McAllister, T. A., & Aalhus, J. L. (2012). Near infrared reflectance spectroscopy predicts the content of polyunsaturated fatty acids and biohydrogenation products in the subcutaneous fat of beef cows fed flaxseed. Meat Science, 90(1), 43-51.

Rudolphi, S., Witzke-Ehbrecht, S. v., & Becker, H. C. (2005). Estimation of oil content in safflower by near infrared reflectance spectroscopy. Proceedings of the VIth International.

Sablinskas, V., Steiner, G., Hof, M. (2003) Chapter 6-Application. In G. Gauglitz, and T.Vo-Dinh (Eds.), Handbook of Spectroscopy (pp.89-168). Germany: WILEY-VCH.

Sato, T. (2002). New estimation method for fatty acid composition in oil using near infrared spectroscopy. Bioscience Biotechnology and Biochemistry, 66(12), 2543-2548.

Shenk, J. S., Workman, J. J., Jr., Westerhaus, M. O. (2008) Application of NIR Spectroscopy to Agricultural Products. In D. Burns and E. Ciurczak (Eds.), Handbook of Near-Infrared Analysis (pp.347-386). Boca Raton, FL: Taylor &Francis Group.

Soyeurt, H., Dardenne, P., Dehareng, F., Lognay, G., Veselko, D., Marlier, M., et al. (2006). Estimating fatty acid content in cow milk using mid-infrared spectrometry. Journal of Dairy Science, 89(9), 3690-3695.

Tsai, F. (1998). Derivative analysis of hyperspectral data. Remote Sensing of Environment, 66(1), 41.

Velasco, L., Mollers, C., & Becker, H. C. (1999). Estimation of seed weight, oil content and fatty acid composition in intact single seeds of rapeseed (brassica napus L.) by near-infrared reflectance spectroscopy. Euphytica, 106(1), 79-85.

Williams, P. C. (2001) Chapter 8 Implementation of Near-Infrared Technology. In P. Williams, and K. Norris (Eds.), *Near- Infrared Technology* (pp.145-169). USA: American Association of Cereal Chemists, Inc.

Yaakob, B. C. M., Syahariza, Z. A., Rohman, A. (2010) Chapter 1. Fourier Transform (FTIR) Spectroscopy: Development, Techniques, and Application in the Analyses of Fats and Oil . In O. J. Rees(Eds.), Fourier Transform Infrared Spectroscopy: Development, Techniques, and Application (pp. 1-26).New York, NY: Nova Science.

Chapter 7 - Conclusions and Future Work

7.1 Conclusion

Blending level of biodiesel-diesel mixtures can be determined by NIR spectroscopy and multiple linear regression methods (MLR). Developed models showed accurate prediction in the selected ranges, even when biodiesel source was unknown in the validation process (RMSEP =3.2 %, v/v). Type of diesel fuel did not show significant effect in the performance of the models. Specific gravity of biodiesel-diesel blend was also determined by NIR and MLR method. Similarly to the model developed for blending level, no effects were observed related to the types of biodiesel or diesel used in this study.

Models to determine the fatty acid concentration in biodiesel using NIR, MIR, and PLSR method were developed. Accurate results were obtained from the models developed and validated with specific types of biodiesels. Less accurate results were observed in the models developed with heterogeneous types of biodiesel and validated with unknown biodiesel samples, indicating the adverse effects of biodiesel source in predicting its fatty acid profile. Such effects were minimized by developing models in narrower concentration ranges. For instance, RMSEP was reduced from 1.29 to 0.66%mass for the model of C16:0 and from 0.76 to 0.27%mass for the model of C18:0. MIR spectroscopy showed slightly better prediction accuracy than NIR. Nevertheless, both MIR and NIR spectroscopy methods can be used to predict the concentration of fatty acid methyl esters in biodiesel, even if the source of biodiesel is unknown.

Two different types of models to predict impurities including methanol, water, triglycerides, and glycerol in biodiesel were successfully developed using both MIR and NIR spectroscopy ranges and the PLSR method. NIR method showed better performance than MIR when methanol, water and glycerol were predicted in models with only one impurity. Model for triglycerides performed better when MIR method was used. To develop models with all impurities, only NIR spectroscopy with derivative techniques was feasible. The obtained RMSEP of the model for methanol, water, triglycerides, and glycerol were 177 mg kg-1, 93 mg kg-1, 932 mg kg-1, and 54 mg kg-1, respectively. Although the models for methanol and water showed better accuracy in prediction than the models for triglycerides and glycerol, all models can be used to predict impurities in biodiesel with reasonable accuracy when all impurities listed above are present.

Derivative technique was used to improve prediction of the concentration of fatty acids (C16:0, C18:0, C18:1, C18:2 and C18:3) present in biodiesel samples using NIR spectra. Accurate results were obtained with the models to predict methyl oleate, methyl linoleate, and methyl linolenate. The RMSEP for the model to predict C18:1, C18:2, and C18:3 were 1.03%(w/w), 0.85%(w/w), and 0.69%(w/w), respectively. Even though, less accuracy was obtained in the models to predict methyl palmitate and methyl stearate, the technique showed great potential to be used in biodiesel spectra analysis.

7.2 Future Work

The recommended next step in detecting biodiesel blending level is to develop an inexpensive portable device based on the finding from this research. This device could use infrared light emitted diode (L.E.D) of identified wavelength from this research with a sensor to detect the received energy, the difference between emitted and received energy will be considered as the absorbed energy in the sample. The absorbed energy is related to the biodiesel level present in the sample.

The models to determine of methyl palmitate (C16:0) and methyl stearate (C18:0) require additional work to improve their accuracy. The improvement achieved developing sub-model to predict the concentration of C16:0 and C18:0 suggests the way forward. A narrow range for calibration models is recommended. Additionally, several options can be analyzed such as different statistic methods (MLR and PCA), artificial neural network, and derivative technique using superior order in the polynomial to improve the results.

To develop future models to detect impurities in biodiesel when all impurities are present, the following are recommended:

- To measure the impurity concentration in each sample, using the reference methods.
- Derivative technique method is also recommended with variation on the principal characteristic (order of polynomial and number of point for the derivative).
- Considering that biodiesel with four impurities is a complex chemical structure, powerful tools to develop models such as artificial neural network (ANN) could help to improve the model.