DEVELOPMENT OF NEW CHEMOMETRICS APPROACHES TO DETERMINE PHYSICAL AND CHEMICAL PROPERTIES OF CRUDE DISTILLATION UNIT PRODUCTS BASED ON MOLECULAR SPECTROSCOPY

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

DOCTOR OF PHILOSOPHY

in Chemistry

by
Ayten Ekin MEŞE SEZEN

December 2022
İZMİR

ACKNOWLEDGEMENTS

There are several people without whom it would not have been possible to complete this master thesis and to whom I am grateful.

To my supervisor, Prof. Dr. Durmuş ÖZDEMİR, for his full support, expert guidance, understanding and encouragement for eight years. His in-depth knowledge and patience made it possible to complete this thesis and I cannot image this study could not be finished without his endless help. I am extremely grateful being part of a work team of his research group.

Besides my advisor, I would like to thank to rest of my thesis committee: Prof. Dr. Figen KOREL, Prof. Dr. Hasan ERTAŞ, Prof. Dr. Şerife Hanım YALÇIN and Assoc. Prof. Dr. Levent PELİT.

To Turkish Petroleum Refineries Corporation (TÜPRAŞ) for all supports and to everyone working in Quality Control Laboratories and R&D Laboratories for collaborative research support.

To my lovely brother, Mehmet Batuhan MEŞE for always being my full supporter in all exams and important stages during my doctoral education. He was not only my brother, but also my roommate for 5 years. I am very grateful to him for enduring me in all my mood changes and unbearable times, for not missing my coffee, for preventing me from starving. Also, I'm sorry to alienate him from graduate school.

To my beloved husband, Berke SEZEN, for being by my side when I feel desperate and helping me for everything I need. I'm happy that his last name is on the cover of the thesis as well.

Finally, my sincere thanks to my parents, Dilek and Metin MEŞE for supporting me spiritually throughout my entire study and encouraged my activities.

ABSTRACT

DEVELOPMENT OF NEW CHEMOMETRICS APPROACHES TO DETERMINE PHYSICAL AND CHEMICAL PROPERTIES OF CRUDE DISTILLATION UNIT PRODUCTS BASED ON MOLECULAR SPECTROSCOPY

Crude distillation units are the first processing units of crude oils based on fractional distillation. The properties of the petroleum products obtained from refinery units are frequently analyzed to ensure that the off-spec product cannot be obtained and that the process is working under the desired conditions. This study aims to develop a method based on multivariate data analysis to determine physical and chemical properties of petroleum samples as an alternative to time-consuming and conventional analytical methods.

Four different petroleum products obtained from CDU for years were selected and used in this study, which are heavy and light diesel, heavy and light straight run naphtha. Four different spectroscopic methods which are UV-Vis, Fluorescence, FT-NIR and FTIR-ATR spectroscopy, were performed and compared. Multivariate calibration models were developed using Partial least Squares (PLS) and Genetic Inverse Least Squares (GILS) algorithms.

For heavy and light diesel, predictive performance of three different spectroscopic methods were compared and for heavy diesel UV-Vis spectroscopy, for light diesel FT-NIR spectroscopy was selected for most of the parameters. Developed models by fluorescence analysis of light diesel samples conducted with two different measurement modes and synchronized fluorescence spectral data has resulted in better models compared to total fluorescence spectra. Studies with straight run naphtha samples were obtained from three different refineries and prediction performances were compared. All obtained model results indicates that developed methodology can be used in routine operations instead of conventional analytical methods.

ÖZET

HAM PETROL ÜNİTELERİ ÜRÜNLERİNİN FİZİKSEL VE KİMYASAL ÖZELLİKLERİNİN BELİRLENMESİNDE MOLEKÜLER SPEKTROSKOPİYE DAYALI YENİ KEMOMETRİK YAKLAŞIMLARIN GELİŞTİRİLMESİ

Karmaşık proseslere ait petrol rafinerilerinde ham petrol öncelikli olarak ham petrol damıtma ünitelerinde işlenerek fraksiyonel damıtma yöntemiyle ayrıştırılır. İşlenen ham petrol farklı kaynaklara ve özelliklere sahip olsa bile, elde edilen ürünün özelliklerinin belirli standartlara uyması beklenmektedir. Bu amaçla ürün özellikleri standart analiz metotlar yardımıyla laboratuvarlarda düzenli olarak takip edilmektedir. Bu çalışmada zaman alıcı klasik analiz yöntemlerine alternatif olarak petrol ürünlerinin fiziksel ve kimyasal özelliklerini belirlemek için çok değişkenli veri analizine dayalı bir yöntem geliştirilmesi amaçlanmıştır.

Çalışmalar ağır dizel, hafif dizel, ağır ve hafif nafta olmak üzere dört ana ürün grubuyla gerçekleştirilmiştir. Ürünler direkt olarak rafineri üretimden elde edilmiş olup, modelleme çalışmaları dört farklı spektroskopik veri ile gerçekleştirilmiştir. UV-Vis, Floresans, FT-NIR ve FTIR-ATR spektroskopik ölçümleri alınan numunelerde modeller Kısmi En Küçük Kareler (PLS) ve Genetik En Küçük Ters Kareler (GILS) algoritmaları kullanılarak geliştirilmiştir.

Dizel numunelerinde üç farklı spektroskopik veriyle gerçekleştirilen model sonuçları karşılaştırılmış ve ağır dizel numunesi için UV-Vis spektroskopisi, hafif dizel için ise FT-NIR spektroskopisi birçok parametre için en iyi model seçilmiştir. Hafif dizellerin floresans analizleri iki farklı modda gerçekleştirilmiş ve senkronize modda toplanan verilerle gerçekleştirilen modeller toplam floresans moduna göre tahmin başarısı daha yüksek modellerle sonuçlanmıştır. Nafta örnekleriyle üç farklı rafineride çalışma gerçekleştirilmiş olup, tahmin performansları karşılaştırılmıştır.

Seçilen tüm model sonuçları, çalışma kapsamında geliştirilen metodolojinin klasik yöntemler yerine kullanılabileceğini göstermektedir.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

LIST OF TABLES

Table **Page**

Table **Page**

LIST OF ABBREVIATION

IBP	Initial Boiling Point
FBP	Final Boiling Point
T(x)	Temperature of (x) % (w/ w%) percent recovery of distillate
HVDZ	Heavy Diesel
LTDZ	Light Diesel
HSRN	Heavy Straight Run Naphtha
LSRN	Light Straight Run Naphtha
FTIR	Fourier Transform Infrared
ATR	Attenuated Total Reflectance
FT-NIR	Fourier Transform Near Infrared
UV-Vis	Ultraviolet-Visible
PLS	Partial Least Squares
GILS	Genetic Inverse Least Squares
CDU	Crude Distillation Unit
SECV	Standard Error of Cross-Validation
SEP	Standard Error of Prediction

CHAPTER 1

INTRODUCTION

Oil refineries have the most complex process in chemical industry. They consist of many different physical and chemical processes. They are mainly atmospheric and vacuum distillation units, cracking isomerization, hydrogenation, desulfurization, aromatization and blending. Many processes are connected to each other, and they work to be the other's charge. In case of any malfunction, unit shut down or out of control processes are affected quickly and serious loses can occur in refinery profitability. Therefore, production is carried out according to certain specifications. Refineries often process different types of crude oil that vary in content. Changes in crude oil prices, capability limits of crude oil tanks, political instabilities of crude oil exporting countries, regulations in product specifications, make it inevitable for crude oil exchanges or the blending of crude oil. Crude oil blending is one of the most important actions of refineries to increase profit margin. Variations in the crude oil composition affect the planned production capacities in order to meet the final product quantities. As each refinery units' scheme is unique, the structure is very complex. Most refineries are designed to process crude oil in a certain API and produce products in a certain range. API stands for the American Petroleum Institute gravity and it is a measure of how heavy or light petroleum liquid is compared to water.

Nowadays, refineries must respond quickly to changes in crude oil, changes in final product demands and leads. The required flexibility in refinery planning and in the complexity of various processes can only be achieved by strictly observing the change of each refinery unit and the flow of output product. The composition in each product stream is the shaped according to the composition in crude oil.

Optimization of crude distillation unit process conditions is the most important parameter in every refinery. Delays in the analysis cause delays in the adjustment of the process conditions. Failure to take an action immediately in the case of defective malfunction may lead to decrease in production or even to unit shut down. Profitability of refinery increases with production of required distillates at maximum efficiency with minimum cost. To achieve this, strict monitoring of the charge and output of each refinery

unit must be performed. Refineries need to respond quickly to changes in crude oil compositions or changes leading to end-product demands.

1.1. Standard Analysis

Determination of the physical properties for the intermediate products of the crude oil unit in the refinery that depends on conventional analytical methods is time-consuming, requiring relatively large sample volume and expensive operations. The extension of the time from sampling to analysis reporting can result in off-spec products at the refinery. Each one of these reference analyses is performed separately according to the appropriate American Society for Testing and Materials (ASTM) or European (EU) standard methods. As a result, laboratory analysis consists of several laborious steps in the form of sampling, sample storage, sample preparation, measurement, data verification and reporting. Among the physical properties routinely monitored are distillation points, API gravity, flash point, vapor pressure and freezing point. The standards that are used for references analysis for this study are given in Table 1.1 for each parameter.

Table 1.1. Standard methods for reference analysis

Analysis	Standard Methods
Distillation Points	EN-ISO-3405; ASTM D86 ASTM D2887 (for heavy diesel only)
API/Density	EN-ISO-3675; ASTM 1298
Flash Point	EN-ISO-2719; ASTM D96 IP 170 (for kerosene)
Vapor Pressure	IP 394; ASTM 5191

EN-ISO-3405 is the test method for atmospheric distillation of petroleum product using a laboratory batch distillation unit to quantitatively determine boiling range characteristics of such products as light and middle distillates and natural gasolines [1]. The boiling range gives information on the composition, the properties and the behavior of the fuel during storage and volatility is the major determinant of the tendency of a

hydrocarbon mixture. Distillation limits are often included in petroleum product specifications, process refinery and control applications.

ASTM D2887 is the standard test method for boiling range distribution of petroleum fractions by gas chromatography [2]. This method cannot be used for the analysis of gasoline samples or gasoline components since it is limited to samples having a boiling range greater than 55 °C. This test method is sometimes used to replace conventional distillation methods (D86). Boiling range distributions obtained by this method are equivalent to those obtained by true boiling point (TBP) distillation, not equivalent to low efficiency distillation such as EN-ISO-3405.

EN ISO 3675 is the standard test method for determination of density of crude oil and liquid petroleum products in laboratory [3]. This method is also called as Hydrometer method. Density is a fundamental physical property to characterize both light and heavy fractions of petroleum products. This standard specifies the density of oil and non-oil products, liquid petroleum products, liquid crude oil under normal conditions, at 15 °C by using glass hydrometer.

EN ISO 2719 and IP 170 are the test methods for determination of flash point [4]. Flash point measures the tendency of the specimen to form a flammable mixture with air under controlled laboratory conditions. This method gives information about overall flammability hazard of a material.

IP 394 the test method for determination of air saturated vapor pressure (ASVP) and calculated dry vapor pressure equivalent (DVPE) [5]. Vapor pressure is one measure of the volatility characteristics of fuels used in many differing types of engines with large variations in operating temperatures. It is used as a classification criterion for the safe handling and carriage pf petroleum products, feedstocks, and components. Specifications for volatile petroleum products generally include vapor pressure limits to ensure products of suitable volatility performance. Fuels having a high vapor pressure may vaporize too readily in the fuel handling systems, resulting in decreased flow to the engine.

IP 435 is the test method for determination of the freezing point of aviation turbine fuels by the automatic phase transition method [6]. The freezing point of an aviation fuel is the lowest temperature at which the fuel remains free of solid hydrocarbon crystals that can restrict the flow of fuel through filters if present in the fuel system of the aircraft. The temperature of the fuel in the aircraft tank normally falls during flight depending on aircraft speed, altitude, and flight duration. The freezing point of the fuel must always be lower than the minimum operational tank temperature.

To determine physical properties of one petroleum product, at least two different standard analysis method should be performed. Each conventional analytical methods, explained above, are time-consuming analysis. Especially, determination of distillation temperature of samples can take more than one hour. Additionally, those analysis requires trained personal and high investment cost.

There are several studies in literature which propose alternative methods to determination of physical parameters of petroleum products. Following section summarizes some of these studies given in the literature.

1.2. Literature Reviews

Chung et al. [7] have studied with six different petroleum products (LSR, Naphtha Kerosene, LGO, Gasoline and Diesel), which are obtained over 4 months and suggested a method for rapid classification of different types of samples by using near-infrared spectroscopy with reflection probe. Principal component analysis (PCA) combined with Mahalanobis distance was used for classification and they suggest that with the help of NIR spectroscopy, identification of petroleum products can be achieved. Using NIR spectra, Kim and his co-workers studied real-time classification for petroleum products [8]. Proposed classification method has been applied to six different petroleum products which are diesel, gasoline, kerosene, light gas oil, light straight run and naphtha. In the study, where PCA was used for feature selection and Bayesian classifier was used for classification, error results were reported as less than 6%.

Kelly and Callis [9] were performed an analysis with finished gasoline products. Since three different hydrocarbon classes, aromatics, olefins and saturates, are several important parameters, this study aims to perform a fast analysis technique, which can be used instead of reference methods. Samples were collected over 19 months and NIR spectra of those samples were collected. Stagewise multilinear regression (MLR) and Partial Least Square (PLS) was used for statistical analysis. This study shows that NIR spectroscopy combined with multivariate analysis gives results at least as good as reference methods for volume percentages of hydrocarbon classes. In other study, analysis of gasoline samples was also investigated by Fodor and his co-workers [10]. In this work, unlike other studies, FT-IR spectroscopy with ATR accessory was used for

spectroscopic analysis. To be able to achieve compositional variety, a total of 800 gasoline samples were collected from different geographic locations from all over US. PLS regression analysis were performed for thirteen different parameters by their respective FTIR spectra. In this article, necessity of baseline correction of FTIR spectra, separate calibration models for summer- and winter-grade gasoline samples were studied. The results showed no significant improvement in calibration models built from baseline corrected FTIR spectra. However, because of different chemical composition for summer- and winter-grade gasoline, two different models were found to be needed for gasoline models. Al-Ghouti and his team have been studied to determine the adulteration of motor gasoline products using FTIR spectroscopy and multivariate calibration [11]. Study aims to quantify three different motor gasoline samples, which were also differ in prices, by means of density and distillation temperatures using FTIR spectroscopy combined with PLS regression method. At the end of the study, proposed method suggested for detecting any adulteration of super leaded motor gasoline from regular leaded motor gasoline. Another study with gasolines were published by Özdemir [12], to determine the octane number of gasoline using NIR spectroscopy and multivariate calibration methods. Three different genetic calibration techniques were used which are genetic regression (GR), genetic classical least squares (GCLS) and genetic inverse least squares (GILS). Calibration models to quantify octane number were built from the set of 60 gasoline samples. Each genetic algorithm-based method was compared with literature studies. Study shows that models obtained with genetic algorithms improve the accuracy of ILS and CLS techniques. For gasoline classification, Balabin and his team [13] compared different multivariate calibration methods. Nine different methods were used which are Linear Discriminant Analysis, Quadratic Discriminant Analysis (QDA), regularized discriminant analysis (RDA), soft independent modeling of class analogy (SIMCA), partial least squares (PLS), K-nearest neighbor (KNN), support vector machines (SVM), probabilistic neural network (PNN), and multilayer perceptron (ANN-MLP). From the calibration models, PNN was found to be the most effective method, however, it is also mentioned that KNN technique is much easier and gives an adequate result, so KNN was recommended.

In another study, Lysaght et al. [14] aimed to determine the percent aromatic, percent saturates and freezing point of military aviation fuel. Composition of a set of 33 JP-4 fuel samples was determined at six different laboratories and each analysis was performed

triplicate. Stagewise multiple linear regression and partial least squares regression were applied to collected NIR spectra and performance of models were compared with ASTM standards. Although errors of prediction sets found to be less than the reproducibility of ASTM methods, limitations due to dependency on reference analysis emphasized as the biggest limitation. Another study with jet fuels was, performed by Westbrook [15], aimed to use NIR spectroscopy for determination of several physical and chemical properties which are cloud point, cetane number, selected percent evaporated distillation points, density, aromatics, heat of combustion and viscosity of Army compression ignition fuels by the help of PLS regression. Performance of each model was presented for their prediction ability and best model was found for density with R^2 0.89 and for the model for cetane number has the lowers R^2, 0.25. Comparison of near-infrared and mid-infrared spectroscopy for properties of kerosene samples were performed by Chung and his co-workers [16]. In the study, performed by fifty samples collected over 3 months, comparison of two different spectroscopic methods were based on the predictive ability of PLS regression models for distillation temperatures of kerosene at different percent recoveries. Results showed that, NIR spectroscopy have better calibration performance over Mid-IR while, Mid-IR provides richer qualitative spectral information with higher resolution power.

The importance of naphtha in the production cycle has been the subject of many studies found in the literature. Spectroscopic analysis combined with chemometrics techniques allows rapid and accurate analysis for determination of naphtha composition. Although there are several studies conducted with naphtha samples obtained from actual refining process [17,18,56–60], sample collection times are only several months. Several of studies have been published to determine detailed hydrocarbon composition of naphtha samples, which are paraffin, naphthenic and aromatic structures [17,58–61]. In industrial applications, it is crucial to have samples represents whole composition variability related to process operations such as crude oil switch, conditions of upstream processes, seasonal operational strategies, and other process variations [62,63]. To be able to have perfect database, sampling stage should be extended. Ku and his co-workers [17] proposed a rapid compositional analysis by Near-Infrared spectroscopy coupled with PLS regression. In the article where paraffin, naphthenic and aromatic content were investigated, a total of 50 different naphtha samples were collected for 3 months. NIR analysis were performed by reflectance probe. It is suggested that obtained models shows excellent correlation

with respective reference analysis of parameters. Another comparison for different spectroscopic methods were also performed by Ku and Chung [18]. Six different chemical composition of naphtha samples were used to compare NIR and Raman spectroscopy. Quantitative analysis of physical and chemical properties of fifty naphtha samples was performed by PLS for each spectroscopic technique. Results showed that NIR have better calibration performance than Raman spectroscopy due to its superior signal-to-noise ratio and spectral reproducibility.

Breitkreitz and her co-workers have been studied to determine total sulfur content in diesel fuel by NIR spectroscopy and multivariate calibration [19]. By using ninety-seven different diesel samples, the performance of five different multivariate calibration approaches which are principal component regression (PCR), Partial least squares regression PLSR, multiple linear regression (MLR), variable selection based genetic algorithm (GA) and successive projection algorithm (PA). The results showed that, not only total sulfur content can be determined by NIR spectroscopy, also all the multivariate calibration models showed acceptable prediction results. However, GA and SPA was emphasized as they have the more robust models compared to other calibration approaches. To determine the main properties of diesel, which are cetane number, cetane index, density, viscosity, distillation temperatures, total aromatics, polycyclic aromatics hydrocarbons, Marinovic and his team proposed PLS calibration models with two different spectroscopic techniques (FT-IR and FT-Raman) [20]. Sample set consist of ninety commercial diesel fuels and results showed that FT-IR spectroscopy combined with PLS regression give better quantitative determination of physico-chemical properties of diesel. Genetic algorithm-based method, GILS, were also used for determination of several properties of diesel samples with the help of near infrared spectroscopy [21]. A total of 250 diesel samples were used and study shows that, GILS algorithm is able to select and extract the chemical information which is sought. For boiling point, total aromatic content, density and viscosity, GILS model results showed successful calibration models.

There are several studies about determination of Biodiesel/diesel blend properties using spectroscopic techniques combined with PLS regression [22–26]. In a study in which NIR spectroscopy used for monitoring biodiesel blends, Oliveira and his team propose a method based on multivariate control charts [22].Study shows that proposed method enhance the quality diagnostics of the model and making possible to identify out of control samples. Another study proposed FTIR spectroscopy combined with PLS

regression methods have been studied by Guerrero et. al [27]. In this study, biodiesel concentration was aimed to quantified in petrodiesel blends and results were reported as obtained models have good ability for determining the concentration of African palm biodiesel in petrodiesel-biodiesel blends. Using synchronous fluorescence, simultaneous determination of quality parameters of biodiesel/diesel blends were also studied by Insausti and his co-workers [24] by the help PLS multivariate calibration. Study was performed with 30 different diesel samples which were collected in different gas stations with biodiesel contents 5% (w/w %) and 7% (w/w %). At the results, it is reported that fatty acid methyl esters, cetane number, gross heat of combustion and color of biodiesel/diesel samples can be identified by fluorescence spectroscopy combined with PLR regression.

Aleme and his team have been proposed two papers which uses distillation curves and multivariate calibrations for determination of flash point and cetane index [28] and specific gravity and kinematic viscosity [29] in diesel using distillation curves and multivariate calibration. Unlike other studies, these studies were performed their calibration models from the results obtained from ASTM reference methods, not from spectroscopic data. From five different refineries, a total of 300 diesel samples were obtained for this study. It was suggested that ASTM D86 method together with PLS regression is effective to predict all the parameters mentioned above, regardless origin and type of diesel.

Although infrared spectroscopies (both Mid-IR and Near-IR) are extensively used for analysis of crude oil and crude oil products, fluoresce spectroscopy is also preferred because of its inhered advantages. Several studies have used fluorescence analysis of crude oil and its products and indicated that physical properties of samples, especially API values, have great impact on fluorescence emissions [51,52] . Different modes of fluorescence spectral analyses such as conventional fluorescence, total fluorescence, synchronous fluorescence, time-resolved fluorescence was conducted and compared. It was indicated that synchronous fluorescence mode is resulted in good results for determination of kerosene present in diesel [52–55].

1.3. Spectroscopic Analysis Methods

As in literature reviews, several spectroscopic methods can be used for determination of quality properties of petroleum products. Four different spectroscopic analyses were used in this thesis which are UV-Visible Spectroscopy, Fluorescence spectroscopy, Fourier Transform Near Infrared Spectroscopy and Fourier Transform Infrared Spectroscopy.

1.3.1. UV-Visible Spectroscopy

UV-Visible spectroscopy is based on the absorption of radiation from ultraviolet region i.e. 180 nm to visible region i.e.780 nm by the chemical compound. Interaction of light with molecules occurs at electronic levels. Electromagnetic radiation interacts with matter and several processes occur which are absorption, transmission, scattering or reflection. Absorption of light by matter causes transition of electrons to different energy level, from ground state to an excited state. During this transition, vibrational and rotational transitions occur as well. This method is widely used in the field of analytical chemistry for quantitative analysis since the amount of light which absorbed is linearly correlated with concentration of sample.

1.3.2. Fluorescence Spectroscopy

Similar to UV-Vis spectroscopy, absorption of electromagnetic radiation caused by photons exciting a molecule and raising it to an electronic state. In fluorescence spectroscopy, absorption spectrum of molecules that has fluorescent properties are obtained. Figure 1.2 shows the Jablonski diagram which explains the fluorescence phenomenon.

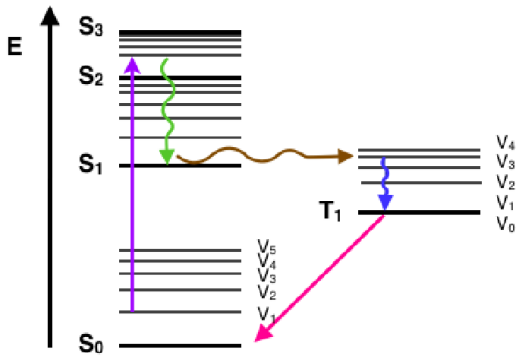

Figure 1.2. The Jablonski Diagram[30]

The main advantage of fluorescence spectroscopy is its high sensitivity of molecules which have fluorescent properties.

1.3.3. Infrared (IR) Spectroscopy

Infrared spectroscopy has been used for the identification and structural analysis of chemical compounds. Beside diatomic molecules, all functional groups absorb IR radiation and frequencies at which functional groups absorb IR radiation is different and unique. Hence, IR spectroscopy has been used for identification and structural analysis of chemical compounds. In IR regions, the vibrations of fundamental bonds on different functional groups remain reasonably independent from the rest of a molecule, while occupying a different, but repeatable, position in the spectrum. The absorption data for molecular bonds are linearly proportional to the concentration information of the sample (Eq 1.1.). It is possible to perform univariate calibration studies by using concentration information with specific absorption peaks belonging to different functional groups in the sample. However, multivariate calibration methods are needed in the analysis of overlapping peaks and combination bands that give information about more than one bond.

$$A = \varepsilon . b . c \qquad (1.1)$$

A: Absorbance, **ε** : Molar absorptivity, **b**: Length of light path, **c**: Concentration

Main advantages of IR spectroscopy are it is a non-destructive technique, it has good precision, no external calibration needed, low S/N ratio and mechanically simple

technique. Infrared region in electromagnetic spectrum is divided into three parts which are Near-IR, Middle-IR, Far-IR as shown in Table 1.2.

Table 1.2. IR Spectral Regions [31]

Region	Wavelength, μm	Wavenumber, cm^{-1}
Near	0.78 to 0.25	12800 to 4000
Middle	2.5 to 50	4000 to 200
Far	50 to 1000	4000 to 600

1.3.3.1. Fourier Transform Near Infrared Spectroscopy

Near-infrared region in electromagnetic spectrum covers the transition from visible spectral region to middle infrared spectral range which is 800-2500 nm (12800-4000 cm^{-1}). Functional groups of -CH, -OH, -SH and -NH tent to absorb NIR radiation and mainly those vibrations are observed. In FT-NIR, spectral features are seen as a result of overtones and combinations of fundamental mid-infrared bands [32,33]. Table 1.3 shows the main overtones and combination bands in the infrared region. [34]

Table 1.3. Main overtones and combination bands in the infrared region

Vibrations	Wavelength, nm
O-H First Overtone	1400 – 1450
O-H Combinations	1900 – 1975
C-H Second Overtone	1125 – 1225
C-H Combinations first overtone	1350 – 1450
C-H First overtone	1625 – 1775
C-H Combinations region	1950 – 2450

The NIR region is attractive for petroleum analysis because many of the absorption bands observed in this region arise from overtones or combinations of carbon-hydrogen stretching vibrations. Although it is hard to interpret FT-NIR spectra due to highly overlapping and broad absorption bands, chemometrics have proven its effectiveness for both quantitative and qualitative analysis in many fields in the literature.

1.3.3.2. Fourier Transform Infrared Spectroscopy

Starting with the invention of interferometers by Albert Abraham Michelson and development of Fourier Transform Infrared Spectrometers has revitalized the field of infrared spectroscopy and overcome the limitation of dispersive spectrometers.

Middle-infrared region has two main regions which are functional group region (3600 – 1200 cm^{-1}) and the fingerprint region (1200-600 cm^{-1}). Compared to fingerprint region, functional group region is easy to interpret. To be able to extract the information from fingerprint region chemometrics models have been used.

1.4. Aim of the Study

Although there are a number of studies for determination of properties of petroleum products using multivariate calibration techniques, it is observed that almost all of these studies performed with finished products or with the samples collected for only short period of time and with a little bit of changes in the process. Additionally, conventional analytical methods have important disadvantages, which are requiring long analyses times and required well-trained personnel. Since all processes are dependent to each other in refinery, long analyses and reporting times can cause lack of feedback which needed for decision making.

This study aims to develop new chemometrics approaches to cover the all the changes in this dynamic process and determine the physical and chemical properties of products obtained from crude distillation units. Since the percentage and composition of crude oil blends fed to the atmospheric distillation unit changes every day, the composition of the distillates also changes. In the refinery where samples are collected for this study, composition of crude oil varies, and more than four types of crude oil blends are processed.

CHAPTER 2

EXPERIMENTAL

2.1. Standard Analysis

All parameters of petroleum samples were determined in accredited laboratory located in Tüpraş Refineries.

Boiling range distribution of samples were analyzed by two different standard methods. For heavy diesels, this analysis was ASTM D2887 [2], while for other crude oil unit products, the EN ISO 3405 [1] method was followed. In ASTM D2887, boiling range distribution is simulated by the use of gas chromatography. A capillary (open tubular) or nonpolar column is used to elute the hydrocarbon components of the sample in order of increasing boiling point. The area under the chromatogram is recorded as the column temperature is raised at a reproducible linear rate. Under the same chromatographic conditions, a known mixture of hydrocarbons is analyzed and from the calibration curve obtained, boiling points of sample are assigned. In EN ISO 3405, the distillation is performed at ambient pressure under conditions that are designed to provide approximately one theoretical plate fractionation. A 100 mL of sample is distilled under specified condition. Each sample is assigned to one of four groups. Temperature readings and volume of condensate are carried out and observations are recorded. At the end of distillation, the observed vapor temperature can be corrected for barometric pressure. The test results are expressed as percent evaporated or percent recovered versus corresponding temperature as a plot of the distillation curve[1].

The API gravity of samples was determined according to EN ISO 3675 [3]. The sample is set so a certain temperature and transferred to the cylinder which approximately has the same temperature. Temperature equilibrium is expected to be reached and then the hydrometer scale is read. Read temperature and hydrometer value are converted to 15^{o}C using standard measurement tables.

The flash point temperatures were determined following the procedure described in EN ISO 2917[4]. In the test method, determination of flash point is performed by

Pensky-Martens closed cup method. The portion of a sample is placed into the test cup of a Pensky-Marten's apparatus and heated to give a constant temperature increase with continuous stirring. The lowest temperature in which a directed ignition source though an opening the test cup lid causes the vapor is recorded as the flash point at the absolute barometric pressure. Then the obtained temperature is corrected to atmospheric pressure.

In test method of vapor pressure, IP 394, a cooled air-saturated sample of known volume is injected into a thermostatically controlled evacuated chamber[5]. After injection into the chamber, the sample is allowed to reach thermal equilibrium at the test temperature and the resulting total pressure in the chamber is equivalent to the vapor pressure of sample. The measured total vapor pressure can be converted to a DVPE by use of a correlation equation.

2.2. Spectroscopic Analysis

Four different spectroscopic analyses were used in this thesis which are Fourier Transform Infrared Spectroscopy, Fourier Transform Near Infrared Spectroscopy UV-Visible Spectroscopy and Fluorescence spectroscopy.

For heavy and light diesel samples, FTIR analysis were carried out at room temperature with Perkin Elmer FTIR spectroscopy equipped with one-diamond ATR accessory with a spectral range from 4000 to 600 cm^{-1} The spectral resolution was 4 cm^{-1} for all spectra and 32 scans were accumulated. Near-Infrared spectra of diesel samples were collected over the 12000 to 4000 cm^{-1} spectral region with a Bruker Matrix FT-NIR Spectroscopy using 2 mm pathlength quartz cell. The spectral resolution was 2 cm^{-1} for all spectra and 8 scans were accumulated. UV-Vis spectra of diesel samples were recorded in a Perkin Elmer Lambda 25 Spectrometer with a spectral range from 200 to 700 nm using 1-cm pathlength quartz cell. A background spectrum of air was recorded in with a clean dry cell before analysis.

Both total fluorescence spectra (TFS) and synchronous fluorescence spectra (SFS) of light diesel samples were recorded on an Agilent Cary Eclipse Fluorescence Spectroscopy using 1-cm pathlength quartz cell for. For TFS, the excitation range was 320 – 410 nm (10 nm intervals) and the emission range 340-650 nm (10 nm intervals). For SFS, spectra were recorded with emission from 300-600 nm with initial $\Delta\lambda$=10 nm and an increment of 10 nm on each scan for a total of 10 scans at a scan rate of 600

nm/min. Excitation slit was set to 5 nm while emission slit was 2.5 nm with a scan rate of 600 nm/min. Samples were passes through 0.45-micron syringe filter before measurement to get rid of unwanted contamination. A background spectrum of air was recorded in with a clean dry cell before analysis.

Near-Infrared spectra of all HSRN and LSRN samples obtained from three different refinery were collected at each refinery quality control laboratory. In İzmit Refinery, spectral measurements were carried out with a Bruker Matrix FT-NIR Spectroscopy using 2 mm pathlength quartz cell over the 12000 to 4000 cm^{-1} spectral region with an air background. In İzmir and Kırıkkale refinery, Near-Infrared spectra all naphtha samples were collected over the 10000 to 4000 cm^{-1} spectral region with a Perkin Elmer Spektrum Two N FT-NIR Spectroscopy using 2 mm pathlength quartz cell with an air background. For each sample, daily spectroscopic measurements were recorded.

CHAPTER 3

DATA ANALYSIS

3.1. Data Pre-processing Techniques

In industrial applications, controlling sample composition and unwanted variations due to spectral measurements are almost impossible which resulted in poor predictive performance in developed calibration models. Depending on the spectral analysis, random measurement noise, non-linear instrument responses, systematic errors and unwanted physical or chemical variations are hard to be avoid.

This situation can be overcome in two main steps. One of them is the extending sampling time to cover changes in crude oil, refinery operations etc. Secondly, choosing proper preprocessing technique to enhance signal properties and suppress unwanted variations. Preprocessing methods can be divided into two main categories, which are spectral derivatives and scatter correction or model-based methods.

Derivative transformations are the oldest technique used to remove the effects of increased noise and scattering in spectra. The first-order derivative estimates the difference between two consecutive spectral measurements, while the second-order derivative is estimated by calculating the difference between two consecutive first-derived spectral measurements. Model-based methods, unlike derivative transformations, are able to quantify and separate the different types of chemical and physical variations in spectra.

3.1.1. Savitzky-Golay Filter

Savitzky-Golay (SG) algorithm is one of the most chosen preprocessing method to which approximates spectrum by polynomial least-square fitting inside a moving window [35–37] . Selection of window size is crucial while optimizing SG technique since every parameter interested in the spectra can place in different spectral region. Although

taking derivative of spectral data increases noise, least square fitting of the SG procedure can reduce the level of spectral noise. Using first derivative of data eliminates the baseline shift.

Savitzky-Golay Filters are widely used for smoothing and differentiation mainly in absorption spectroscopy. This preprocessing method is one of the popular choices in an averaging algorithm that fits least squares polynomial to the data points and then the value to be averaged is predicted from the polynomial. There are two important parameters in this transformation. One of them is the window opening and the other is the degree of derivative and these values are determined by the user [35,36,38–41]. The selection of window size is critical. Selected window size has strong influence on derivated curve, hence multivariate analysis. Even though same sample is used, signal bandshapes can change in different spectral region. Optimal window size for should be found before each data set obtained from each spectral region before multivariate analysis. Since derivation of data occur in between selected window size, it is important not to lose spectral information during this process and to keep meaningful spectral information within the selected window range.

3.1.2. Multiplicative Scattering Correction

In the 1980s, model-based data preprocessing methods began to take place in the literature. Multiplicative Signal Correction, MSC, method is a method used in signal processing. It was first presented by Marten et al in 1983 [42], and further studies were carried out by Geladi et al[43]. The purpose of the method is to remove the unwanted scattering effects in the spectra. MSC basically takes place in two steps:

1. Regression of each spectrum against the mean spectrum.

$$\mathbf{x}_i \approx a_i + b_i \mathbf{x}_m \tag{3.1}$$

2. The original data set is corrected with the obtained coefficients.

$$\mathbf{x}_i^{msc} = (\mathbf{x}_i - a_i)/b_i \tag{3.2}$$

where $\mathbf{x}_m$ is the mean spectrum, $\mathbf{x}_i$ is the i^{th} spectrum of the collection used for calculation. By ordinary least squares regression of $\mathbf{x}_i$ onto $\mathbf{x}_m$ estimates of *a* and *b* can be obtained.

In the original paper, reference spectrum is suggested to be calculated by using parts which do not contain spectral information in the spectra. However, it is hard to determine such region in application especially in FT-NIR spectra since signals are highly overlapping and correlated. Alternatively, most studies use entire spectrum to calculate average spectrum which is called reference spectrum.

3.1.3. Extended Multiplicative Scattering Correction

By expanding the basic logic in the MSC method, the EMSC method was proposed in 1991 and is frequently used especially in NIR spectroscopic data [44]. Unlike the MSC method, a quadratic polynomial equation is fitted to the reference spectrum and a correction vector (reference spectrum) is obtained by establishing a regression between the reference spectrum and the prior information. Correction is performed on the raw data using the reference spectrum. In the EMSC method, the trend of the data to the wavenumber axis is also often included in the 2nd order polynomial equation. [36,38].

$$\mathbf{x}_\mathrm{i} \approx a_\mathrm{i} + b_\mathrm{i} * \mathbf{X}_\mathrm{m} + d_i * \boldsymbol{\lambda} + e_i * \boldsymbol{\lambda}^2 \tag{3.3}$$

$$\mathbf{x}_\mathrm{i}^{\mathrm{emsc}} = (\mathbf{x}_\mathrm{i} - (a_\mathrm{i} + d_i * \boldsymbol{\lambda} + e_i * \boldsymbol{\lambda}^2)) / b_\mathrm{i} \tag{3.4}$$

where $\mathbf{x}_\mathrm{m}$ is the mean spectrum, $\mathbf{x}_\mathrm{i}$ is the i[th] spectrum of the collection used for calculation and $\boldsymbol{\lambda}$ is the wavelengths or wavenumbers of the spectral range for the wavelength dependency correction. By applying multiple regression of $\mathbf{x}_\mathrm{i}$ onto $\mathbf{x}_\mathrm{m}$, $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}^2$, estimates of a, b, d and e can be obtained.

It is emphasized in the literature that it is more selective and effective than a classical filtering technique in eliminating deviations and unwanted variations caused by various reasons [39,40,45]. This technique, which is used to separate and measure chemical and physical deviations, allows better interpretation of spectra and statistically more robust calibration models. Its use is increasing day by day not only in the NIR region, but also in Mie scattering in the mid-infrared region, and laser and particle size-based data in Raman spectroscopy.

The most important parameter in MSC and EMSC techniques is the correct reference spectrum calculation. Data which has an absorbance value 2.5 or higher can store meaningful spectral information. However, since Beer's rule deviates on those absorbance values, MSC techniques can be applied to those data and spectral information can be saved.

In order to prevent this and create a robust reference spectrum, two different alternatives are presented:

1. Adding predetermined weight coefficients to the wavenumber axis
2. Using iterative search method for optimal weight coefficients

The iteration method works on the theme of assigning low weighting coefficients to regions where there are very large differences between the spectra. Thus, it is ensured that the spectral regions containing information about the sample are not affected by the large differences in the noisy regions.

3.2. Calibration Methods

Calibration methods are used to find the relationship between two different sets which are output from an instrument, x, and properties of the sample, c. The mathematical formula that describes this relationship then can be used to predict the interested properties of the sample. Linear calibration methods can be used then the properties of sample and instrument response(s) are linearly correlated. These calibration techniques can be divided into two groups:

1. Univariate Calibration
2. Multivariate Calibration

When the aim is to develop a model for a single property of the sample by using a single variable, univariate calibration is used. In spectroscopic data, mostly single variable is chosen as the wavelength of highest absorbance. If multiple responses are used to model multiple properties of a sample, it is called multivariate calibration.

3.2.1. Univariate Calibration

Most common univariate calibration methods are classical univariate calibration and inverse univariate calibration. Classical univariate calibration is the most common and simple univariate calibration method and very similar to Beer-Lambert Law. General formula of classical univariate calibration is given in equation 3.5 while formula of inverse univariate calibration is given in equation 3.6.

$$x = c.s \tag{3.5}$$

$$c = x.s \tag{3.6}$$

where x is an instrument response, c is the property of sample and s is a scalar that relates these two variables.

The main difference between classical and inverse univariate calibration is that classical univariate calibration assumes all the errors in the instrumental responses while inverse univariate calibration assumes the source of errors are in concentrations. Considering preparation of sample more likely includes human error considering using volumetric flasks, sample containers than errors during sample scans considering the improvements in devices, inverse univariate calibration method is resulted in better calibration model.

To reflect baseline of the responses, intercept term is used in calibration models. Using intercept term may significantly increase the predictive performance of developed calibration model if a constant response is observed. There are two main ways to add intercept term. One of them is given below.

$$\text{If } x = c.s\,;$$

$$\mathbf{s} = (\mathbf{C}'.\mathbf{C})^{-1}.\mathbf{C}'.\mathbf{x} \tag{3.7}$$

where **C** is a two column matrix in which first columns consist of ones for fitting the intercept and the second column for the property of the training samples e.g. concentrations, **x** is the vector of responses and **s** is the vector of coefficients relating the **x** to **C**.

$$\text{If } c = x.s\,;$$

$$\mathbf{s} = (\mathbf{X}'.\mathbf{X})^{-1}.\mathbf{X}'.\mathbf{c} \tag{3.8}$$

where **c** is vector of the property of the training samples e.g. concentrations, **X** is the matrix includes two columns in which first columns consist of ones for fitting the intercept and second one is responses and **s** is the vector of coefficients relating the **c** to **X**.

Second way of adding intercept to model is mean centering the data. Mean centering is applied by subtracting the mean of properties from each property while mean of responses are subtracted from each response.

3.2.2. Multivariate Calibration

If the property of interest consists of several chemical compounds each having different wavelength of absorbance, univariate calibration methods may fail. Using multiple wavelengths resulted in better calibration models most of the time since averaging of useful information in absorbance is used. In addition, effect of random interferences and noises are less when compared to univariate calibration due to same averaging. Beside these advantages, there are some limitations as well. To be able to obtain reboots calibration models, the number of variables should be at least the number of compounds and the number of experiments should also exceed the number of variables.

Classical least squares calibration, CLS, is the multivariate form of classical univariate calibration. To obtain a reasonable model from CLS, all properties of the sample should be used while modelling. Inverse least squares, ILS, similar to inverse univariate calibration, defines the properties of a sample as a function of response. That allows to model a single component in a sample without a knowledge of the other components. However, this method prone to multicollinearity issues which resulted in overfitting.

3.2.2.1. Partial Least Squares (PLS)

Partial least squares (PLS) is a soft modeling technique in which the data are decomposed into new variables which are linear combinations of the original data using Principal Component Analysis (PCA). These new variables coming from PCA are named as principal components or factors. This method uses only the advantages of CLS and ILS methods. There is no need for the knowledge of all interfering species like CLS. Also, since new variables are obtained from PCA and these variables projects the dependent variables to a new smaller dimensional space, multicollinearity problem is mostly eliminated. Unlike other methods, PLS accounts noise in both responses and concentrations.

There are two different PLS algorithms which are called PLS1 and PLS2. Studies in this thesis PLS1 algorithm was used since it has better prediction performance. The decomposition of the data in PLS1 is given below. Figure 3.2 shows the graphical representation of equation 3.9 and 3.10. to show matrix size.

$$\mathbf{X} = \mathbf{T}.\mathbf{P} + \mathbf{E} \tag{3.9}$$

$$\mathbf{c} = \mathbf{T}.\mathbf{q} + \mathbf{f} \tag{3.10}$$

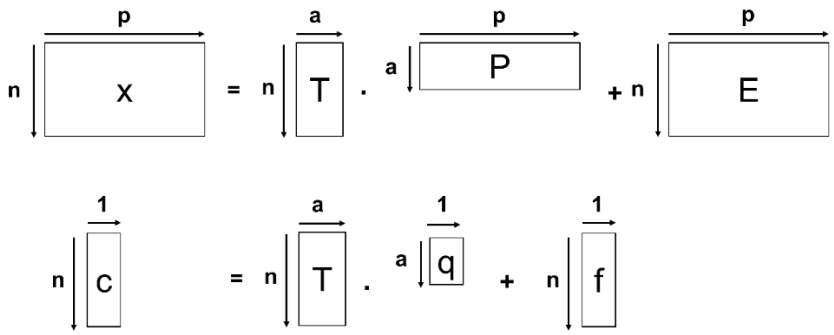

Figure 3.1. The size of matrices and vectors used in PLS.

where **X** is the matrix of responses obtained from instrument for *n* number of training samples at *p* number of variables (wavelengths or wavenumbers), **T** is the scores matrix which is used to develop a model with ILS algorithm, **P** is the loading of responses,

E is the residuals of responses. The number of latent variables is shown as a. Property of interest is shown as **c** which is usually the concentrations one of the components of the training samples, **q** is the vector of regression coefficients that relates the score matrix **T** to concentration vector **c**. Here, **f** is the vector of residuals for a number of latent variables.

Determination of latent variable is one of the most important aspect in PLS modelling. Most common way for choosing latent value is using the predicted estimation sum of squares (PRESS) values. From calculated PRESS values, latent variable is chosen when PRESS stop decreasing or before the increment in PRESS.

$$PRESS = \sum_{i=1}^{m} (\hat{c} - c_i)^2 \tag{3.11}$$

3.2.2.2. Genetic Inverse Least Squares

Genetic Inverse Least Squares (GILS) is the combination of ILS for calibration and a genetic algorithm for feature selection. GILS effectively picks and shuffles small sets of features in an iterative process to obtain better sets where the success criteria is assigned by cross-validation with ILS. These feature sets are then used for creating new models without the collinearity problem and potentially with less overfitting issues while averaging provides even further enhancement in predictive power.

There are five main steps which are initialization of gene population, fitness assignment of the population, selection of genes to be breed, cross-over and breeding, lastly the replacement of the genes with the old ones.

To find the best combination of variables and achieve desired computational time and obtain robust model, there are used defined options which are selection of number of genes in the population, coefficient threshold for the initialization of new genes, the number of iterations for breeding and replacement of old genes.

To construct a gene, random number of variables, which are a set of absorbances at specific wavenumbers, are selected. Obtained gene is subjected to the cross-validation for the determination of its fitness and coefficient. Leave-one-out cross validation is applied where a single sample is removed from the whole data and ILS model developed with rest of them and removed data is predicted. This procedure is repeated for all

samples. If R^2 of obtained gene is above defined threshold, it is added to the pool, otherwise it disposed. New genes are created until the population of gene pool reaches the user defined number. Those genes are called the parent genes. For each gene in the pool, fitness values are calculated by the formula below.

$$Fitness = \frac{1}{SECV} \tag{3.12}$$

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

where m is the number of samples, c_i is the property of sample and $\hat{c}_i$ is its prediction.

Selection of parent genes for breeding, roulette wheel method is used. In this method, each gene has an area which is proportional to their fitness value. Genes are selected by spinning the wheel and randomly chosen. The genes with highest fitness value have a more chance to be selected then genes with lower fitness values since they have higher portion on the wheel. From the selected genes from wheel, off-spring genes are created. Parent genes, or first selected genes, are cut from the middle and exchange variables so a new pair of offspring are formed.

Parent Gene Pair:

G1 = [A_{300}, A_{4300} # A_{1450}, A_{2814}]

G2 = [A_{500}, A_{1000}, A_{208} # A_{2108}, A_{1991}, A_{1993}]

Offspring:

NEWG1 = [A_{300}, A_{4300}, A_{2108}, A_{1991}, A_{1993}]

NEWG2 = [A_{1450}, A_{2814}, A_{500}, A_{1000}, A_{208}]

Parent genes are cut off from the # symbol and subjected to cross-over. Then parent genes are replaced with their offspring. The fitness of offspring is compared the fittest gene and replaced if offspring fits better. This iteration is repeated until user defined iteration number is reached. When it is done, whole procedure which starts from initialization and ends from iteration is called a run. After each run, fittest gene is found and after completing the number of runs, final model obtained from best genes.

CHAPTER 4

RESULTS AND DISCUSSION

Studies with each crude distillation unit products were performed separately and reported in order from heavy product to lightest under different sections.

Studies with diesel samples mostly not separated as light and heavy fractions of diesel in literature studies. However, in Tüpraş Refineries, diesel samples are obtained from crude distillation units are separated as light and heavy fractions due to the production parameters. Light diesel samples are obtained right above where heavy diesel samples are obtained and these samples have lower carbon number hence lower boiling points then heavy diesel samples.

4.1. Heavy Diesel Samples

Studies conducted with heavy diesel samples were performed in three different spectroscopic methods which are FTIR-ATR, FT-NIR, and UV-Vis Spectroscopy. Samples were obtained from one of the crude distillation unit in TUPRAS İzmit Refinery.

Multivariate calibration studies with heavy diesel samples collected for one year (from 2018 to 2019). During one year of sample collection time, a total of 23 different crude oil with different percentages were processed in crude distillation unit. Distillation temperatures, API gravity and flash point of the samples were analyzed in quality control laboratory located in İzmit Refinery, according to their respective ASTM methods. Distillation temperature analysis of heavy diesel samples were analyzed three times a day while API and flash point analyses were performed once a day.

Each spectroscopic analysis of samples was performed in different laboratories. To be able to compare the multivariate calibration model performance of heavy diesel samples, both calibration and validation data sets were kept same for each spectroscopic analysis.

During one year of sample collection, distillation temperature range of a total of matched 268 heavy samples were analyzed with both standard analysis method and three spectroscopic analyses. Since standard method API and flash point analyses were

performed once a day, only 88 matched sample was analyzed with three spectroscopic analyses. The distillation temperature ranges from initial boiling point (IBP) to final boiling point (FBP) along with API gravity and flash point of samples are summarized in Table 4.1. Graphical representation of Table 4.1 is shown in Figure 4.1.

Table 4.1. Data range, mean, median and standard deviation of physical properties of heavy diesel samples.

	Minimum (Min)	Maximum (Max)	Mean	Median	Standard deviation
IBP, °C	114.6	139.7	129.8	130.7	6.0
T5, °C	180.3	201.0	191.7	192.2	4.0
T10, °C	211.6	230.5	221.8	222.8	4.0
T20, °C	249.5	267.3	257.8	258.4	3.7
T30, °C	272.7	291.9	282.6	283.2	3.5
T40, °C	289.8	311.9	302.2	302.1	3.7
T50, °C	302.1	328.9	319.1	319.8	4.7
T60, °C	315.2	346.5	335.6	337.1	5.9
T70, °C	327.5	363.8	351.4	354.0	6.8
T80, °C	341.0	382.3	368.8	371.9	8.5
T90, °C	357.8	410.5	392.4	396.8	11.5
T95, °C	372.7	436.0	410.8	415.7	13.8
FBP, °C	412.5	503.0	451.9	455.0	17.0
API	30.7	35.3	33.4	33.4	1.0
Flash Point, °C	75.0	94.0	83.2	83.5	4.5

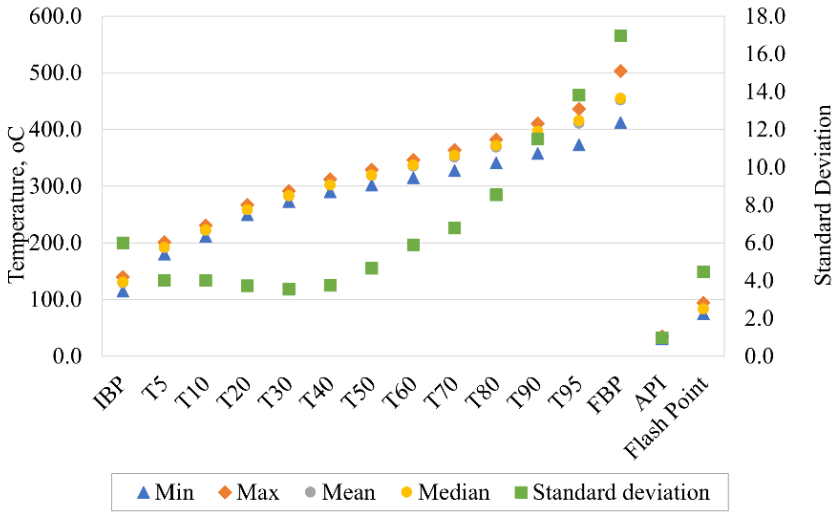

Figure 4.1. Graphical representation of Table 4.1.

According to the standard deviations, green dots whose y-axis is on the right in Figure 4.1, the parameters with relatively less variability are T30, T40 and T50 while FBP has the highest variation. Mean and median values of all parameters are close to each other which indicates the distribution of dataset can be assumed as normal.

For distillation models, from 268 samples, a total of 178 samples were assigned as calibration set to train model, a total of 60 samples were used as independent validation and remaining 30 samples were used as unknowns to observe prediction performance of chosen model. For API and flash point models, a total of 66 samples were assigned as a calibration set while the rest 22 samples were used as independent validation set. Since the number of samples in API and flash point dataset is less than distillation point dataset, prediction performance of models was decided to be observed using SEP values and no unknown set was assigned.

Each data set obtained from spectroscopic analysis were treated with different data pre-processing method and two different multivariate calibration approach were performed to preprocessed data, which will explain under spectroscopic analysis results.

4.1.1. FTIR-ATR Spectroscopic Results

Raw FTIR-ATR spectra of a total of 268 heavy diesel samples are shown in Figure 4.2.

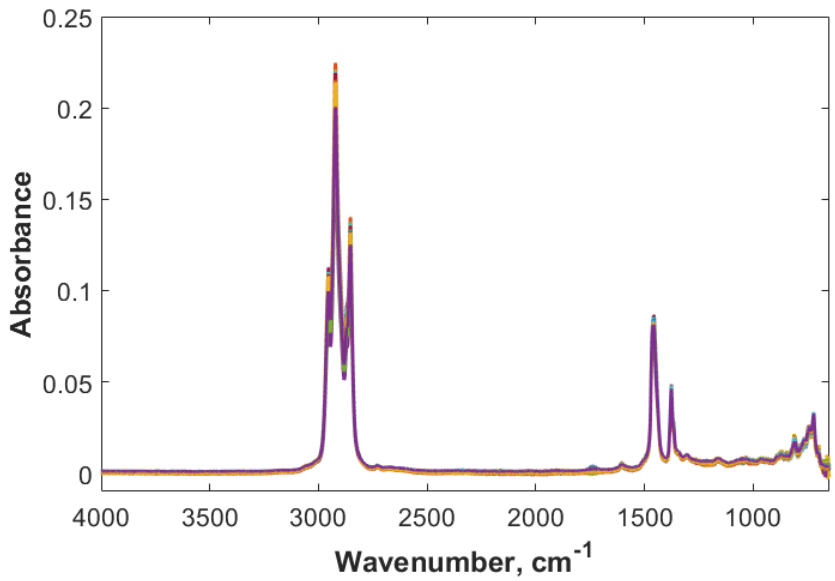

Figure 4.2. Raw FTIR-ATR Spectra of a total of 268 CDU heavy diesel samples.

In Figure 4.2., spectral regions 3100-2500 cm^{-1} correspond to C-H stretching while 1800-1000 cm^{-1} correspond to C-H bending vibrations. The absorbances around 3000-2800 cm^{-1} raised from C-H stretching vibrations of CH_2 groups and stretching band of CH_3 groups. The presence of stretching vibrations of C-N groups were indicated in region located at 1300-1250 cm^{-1} while bending vibrations of methylene groups were observed at around 1450 cm^{-1}.

Spectral data shown in Figure 4.2 was optimized for three different ASTM methods which are distillation, API gravity and flash point analyses to determine the best preprocess technique. From 13 different distillation point parameters, T50 was selected for optimization of preprocess since it has the lowest reproducibility values compared to rest of distillation points. Reproducibility value is defined as "*The difference between two single and independent results obtained by different operators working in different laboratories on identical test material would, in the long run, exceed the following values only one case in twenty"* in ASTM standards. Table 4.2 shows the reproducibility value calculation of distillation parameters from ASTM methods. The letter i next to the physical parameters states the i^{th} sample used for calculation.

Table 4.2. Reproducibility value calculation for heavy diesel according to ASTM D2887.

Parameters	**Heavy Diesel**
	Reproducibility, °C
IBP	0.066 $\mathbf{IBP_i}$
T5	0.015 ($\mathbf{T5_i}$ +100)
T10	0.015 ($\mathbf{T10_i}$ +100)
T20	0.015 ($\mathbf{T20_i}$ +100)
T30	0.013 ($\mathbf{T30_i}$ +100)
T40 – T90	4.3
T95	5.0
FBP	11.8

To the spectral data shown in Figure 4.2., three different preprocess techniques were applied which are SG filter, MSC and EMSC. By systematic varying preprocess parameters like smoothing, chancing window size, polynomial order, derivative options, different scattering correction methods, best preprocessing method was determined for each model development. Best technique was chosen for model that has lowest SEP value.

Best preprocess technique for distillation point models were obtained with a first order derivate with Savitzky-Golay algorithm using second order polynomial and 11 points window size along with Multiplicative Scattering Correction. For API and Flash Point, first order derivative with SG algorithm using first order polynomial and 13 points window size along with Extended Multiplicative Scattering Correction. Additionally, reduction of wavenumber range from 4000 cm^{-1} to 3300 cm^{-1} was resulted in models with higher predictive ability. Figure 4.3 shows FTIR-ATR spectra of heavy diesel samples in which selected preprocessed technique for distillation point parameters were applied.

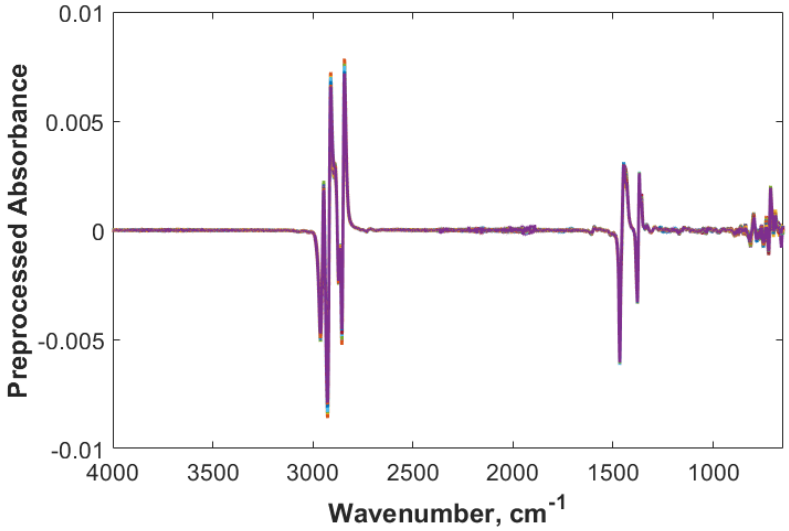

Figure 4.3. FTIR-ATR Spectra of heavy diesel samples treated by first derivative using first order polynomial and a window size of eleven followed by MSC.

As shown in Figure 4.3, taking first derivative of spectra emphasize the maximum absorbances on each spectrum. After determination of preprocessing technique, two different calibration approach were applied to the data set which and PLS and GILS algorithm. Table 4.3 shows the calculated standard error of cross-validation (SECV), standard error of prediction (SEP) and the correlation coefficient of calibration curve, R^2, of each parameter obtained from developed PLS and GILS models. Developed model that has lowest SEP values for each parameter was highlighted as bold font.

Table 4.3. Two different multivariate calibration results of heavy diesel samples obtained from FTIR-ATR along with data range and average reproducibility value (R, avg)

FTIR-ATR	Data Range			PLS				GILS			R avg
	Min	Max	Range	SECV	SEP	R^2	LVs	SECV	SEP	R^2	
IBP,°C	114.6	139.7	25.1	1.178	3.101	0.963	7	**1.619**	**2.866**	**0.938**	8.4
T5, °C	180.3	201.0	20.7	1.521	2.474	0.860	6	**1.506**	**2.342**	**0.886**	4.4
T10, °C	211.6	230.5	18.9	**0.933**	**2.215**	**0.943**	**7**	1.277	2.263	0.910	4.8
T20, °C	249.5	267.3	16.9	0.518	1.977	0.981	9	**0.892**	**1.907**	**0.951**	5.4
T30, °C	272.7	291.9	19.2	0.593	1.599	0.974	8	**0.754**	**1.536**	**0.963**	5.0
T40, °C	289.8	311.9	22.1	0.637	1.720	0.973	8	**0.791**	**1.636**	**0.963**	4.3
T50, °C	302.1	328.9	26.8	**1.192**	**1.810**	**0.940**	**6**	0.850	1.811	0.970	4.3
T60, °C	315.2	346.5	31.3	1.390	2.164	0.949	6	**1.106**	**1.974**	**0.968**	4.3
T70, °C	327.5	363.8	36.3	1.746	2.488	0.939	6	**1.546**	**2.302**	**0.953**	4.3
T80, °C	341.0	382.3	41.3	1.552	3.464	0.970	7	**1.711**	**3.108**	**0.964**	4.3
T90, °C	357.8	410.5	52.7	2.977	4.706	0.939	6	**2.675**	**4.442**	**0.952**	4.3
T95, °C	372.7	436.0	63.3	3.777	5.899	0.931	6	**3.601**	**5.755**	**0.939**	5.0
FBP, °C	412.5	503.0	84.2	7.580	8.702	0.801	5	**6.262**	**8.478**	**0.872**	11.8
API	30.7	35.3	4.8	0.045	0.194	0.998	7	**0.119**	**0.168**	**0.987**	0.5
Flash P., °C	75.0	94.0	19.0	**0.665**	**1.751**	**0.980**	**7**	2.220	1.784	0.812	6.0

As Table 4.3 indicates, beside of T10 and T50 distillation points and flash point, GILS models resulted in lowest SEP values which indicates better predictive ability. Results indicates that genetic algorithm can be selected for quantitative analysis of petroleum products conducted with FTIR-ATR spectroscopy. Another trend was observed in SEP values of distillation parameter models. Lower reproducibility value of standard analysis method has resulted in lower the standard error of prediction value. Since multivariate calibration models are trained with standard analysis results and reproducibility value indicated the nature error found in the analysis, it is accepted to have better models for parameters that have lowest reproducibility values.

4.1.2. Near-Infrared Spectroscopic Analysis Results

Raw FT-NIR spectra of a total of 268 heavy diesel samples are shown in Figure 4.4.

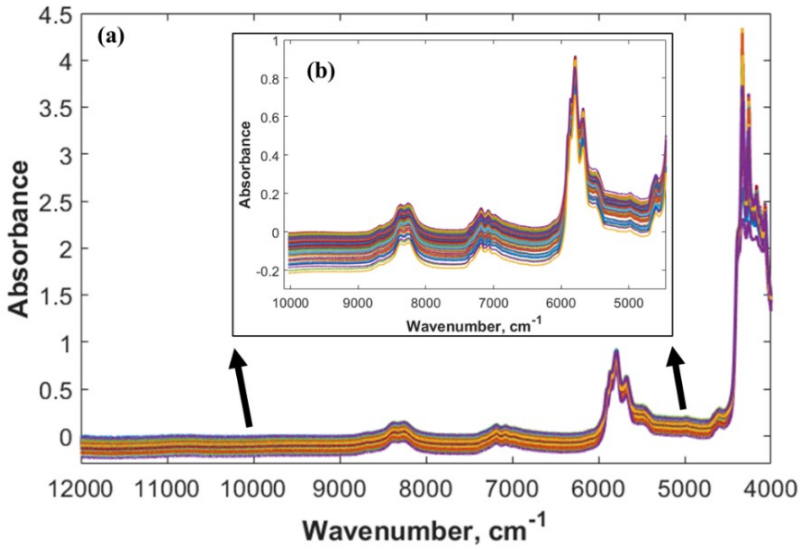

Figure 4.4. Raw FT-NIR Spectra of a total of 268 CDU heavy diesel samples (a) full range (b) narrowed range

As shown in Figure 4.4 (a), although the 4500-4000 cm^{-1} spectral range shows significant spectral variations, it contains no useful spectral information due to the strong saturation of NIR radiation. Before any further analysis, in addition to spectral range of 4500-4000 cm^{-1}, 12000-9000 cm^{-1} ranges were also removed from data set since no infrared absorption was observed between these spectral ranges. Figure 4.4 (b) shows the narrowed range NIR spectra.

In Figure 4.4 (b), spectral bands between 9000-8000 cm^{-1} corresponds to the second overtone region, 7500-7000 cm^{-1} corresponds to second combination region and 6200-5300 cm^{-1} corresponds to first overtone region of C-H bands [50] The spectral region found in 4700-4300 cm^{-1} spectral range, corresponds a part of first combination region. Although noticeable spectral variations in terms of position and intensity of peaks are observed around major bands, significant baseline shifts were observed in obtained NIR spectra. To be able to enhance signal properties and eliminate baseline and background effects, preprocessing techniques were applied to raw FT-NIR spectra.

By systematic varying preprocess parameters were also applied to FT-NIR spectra and best preprocess method was chosen for each parameter according to lowest SEP values. For distillation points, smoothing with third order polynomial fitting and 5-point window size following Extended Multiplicative Scatter Correction was chosen. For API and Flash Point, only EMSC was applied to the narrowed data set. Figure 4.5. shows the preprocessed FT-NIR spectra of heavy diesel samples.

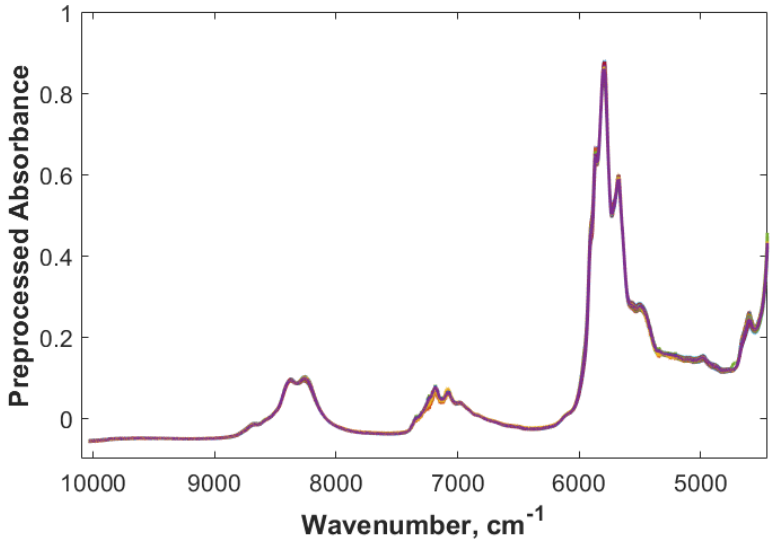

Figure 4.5. FT-NIR Spectra of heavy diesel samples treated by smoothing using first order polynomial and a window size of 5 followed by EMSC.

As shown in Figure 4.5, baseline variation has been minimized with selected preprocessed technique. To spectra shown in figure above, two different multivariate calibration approach were performed, and multivariate calibration results are given in Table 4.4 with standard error of cross-validation (SECV), standard error of prediction SEP), the correlation coefficient of calibration curve, R^2, along with latent variables for PLS models. For each parameter, multivariate calibration model with lowest SEP values was highlighted as bold font.

Table 4.4. Two different multivariate calibration results of heavy diesel samples obtained from FT-NIR along with data range and average reproducibility value (R, avg)

FT-NIR	Data Range			PLS				GILS			R avg
	Min	Max	Range	SECV	SEP	R^2	LVs	SECV	SEP	R^2	
IBP, °C	114.6	139.7	25.1	**1.499**	**2.622**	**0.940**	**13**	1.709	2.670	0.926	8.4
T5, °C	180.3	201.0	20.7	**1.331**	**1.956**	**0.891**	**13**	1.390	2.224	0.893	4.4
T10, °C	211.6	230.5	18.9	**0.884**	**2.010**	**0.951**	**14**	1.240	2.189	0.912	4.8
T20, °C	249.5	267.3	16.9	**0.829**	**1.797**	**0.952**	**13**	1.129	1.861	0.918	5.4
T30, °C	272.7	291.9	19.2	0.881	1.428	0.936	11	**0.788**	**1.415**	**0.951**	5.0
T40, °C	289.8	311.9	22.1	1.477	1.610	0.855	11	**1.233**	**1.541**	**0.903**	4.3
T50, °C	302.1	328.9	26.8	**0.859**	**1.550**	**0.969**	**12**	0.841	1.609	0.971	4.3
T60, °C	315.2	346.5	31.3	**1.440**	**1.500**	**0.946**	**12**	1.290	1.570	0.958	4.3
T70, °C	327.5	363.8	36.3	**1.009**	**1.646**	**0.980**	**15**	1.332	1.778	0.966	4.3
T80, °C	341.0	382.3	41.3	**1.283**	**2.023**	**0.979**	**14**	1.655	2.372	0.967	4.3
T90, °C	357.8	410.5	52.7	**1.668**	**2.948**	**0.981**	**17**	2.702	4.119	0.954	4.3
T95, °C	372.7	436.0	63.3	**2.385**	**3.704**	**0.973**	**14**	2.981	4.432	0.960	5.0
FBP, °C	412.5	503.0	84.2	**3.833**	**5.818**	**0.946**	**13**	4.625	6.955	0.927	11.8
API	30.7	35.3	4.8	**0.068**	**0.108**	**0.996**	**12**	0.132	0.171	0.984	0.5
Flash P., °C	75.0	94.0	19.0	**1.964**	**1.889**	**0.802**	**8**	1.573	2.063	0.887	6.0

To the contrary of multivariate calibration results of FTIR dataset, Table 4.4 indicates that PLS method resulted in lower SEP values compared to GILS models, beside for T30 and T40 distillation points. Relation between SEP and reproducibility values are observed in FT-NIR results as well. It is also observed that chosen latent variable values for FT-NIR PLS models are higher than the FTIR PLS values.

4.1.3. UV-Vis Spectroscopic Analysis Results

Raw UV-Vis spectra of a total of 268 heavy diesel samples is shown in Figure 4.6.

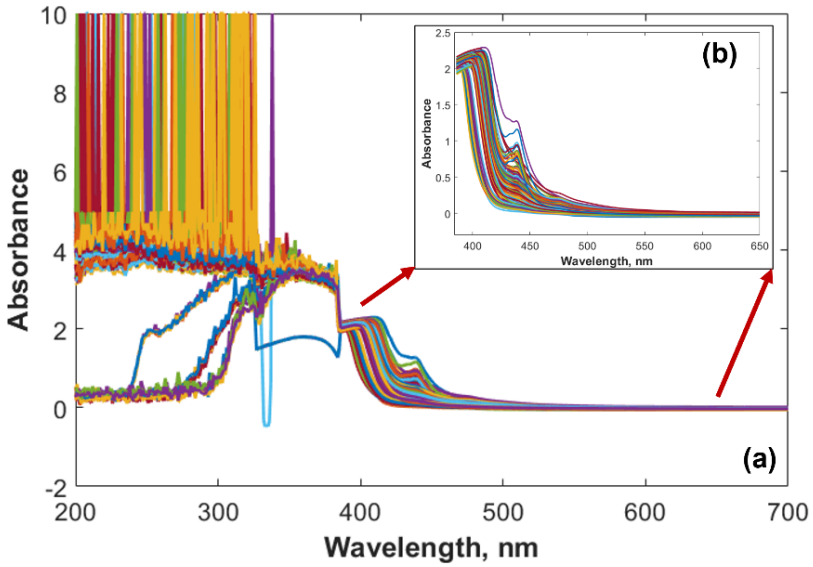

Figure 4.6. Raw UV-Vis Spectra of a total of 268 CDU heavy diesel samples (a) full range (b) narrowed range

As shown in Figure 4.6 (a), high absorption values between wavenumber from 200 to 385 nm were observed in the spectra due to the saturation. Those wavenumber regions were removed from the spectra and narrowed range UV-Vis spectra is shown in Figure 4.6 (b).

Best preprocessing technique was also searched for UV-Vis study. For distillation Savitzky-Golay filtering technique was applied to UV-Vis spectra using a window size of 9, 3rd order polynomial fitting and calculating the 1st derivative following by MSC which is shown in Figure 4.7. For API and Flash Point, smoothing was applied to Savitzky-Golay using 13-point window size and first order polynomial fitting followed by MSC.

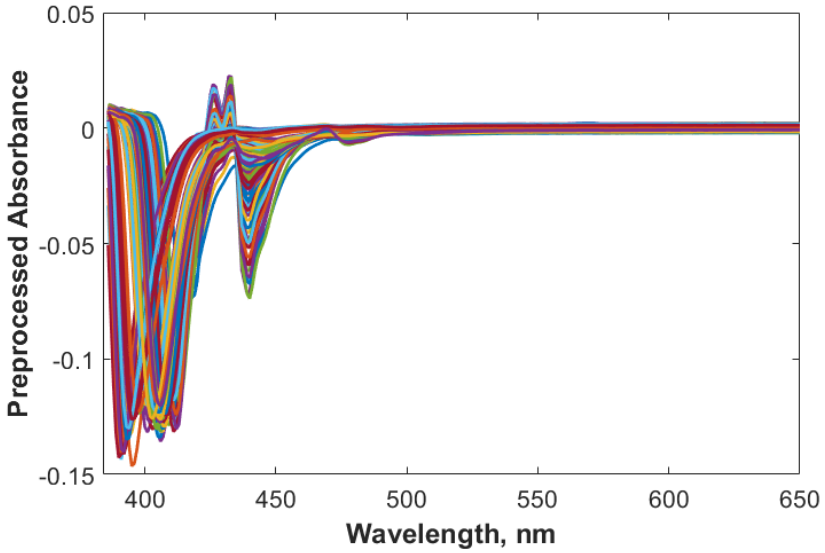

Figure 4.7. Preprocessed UV-Vis Spectra heavy diesel samples.

To the spectra shown in Figure 4.7, two different calibration approaches which are PLS and GILS are applied and SECV, SEP values along with the correlation coefficient of calibration curve, R^2 and selected latent variables of PLS models are given in Table 4.5. Multivariate calibration model with lowest SEP values for each parameter is highlighted as bold in the Table.

Table 4.5. Two different multivariate calibration results of heavy diesel samples obtained from UV-Vis along with data range and average reproducibility value (R, avg)

UV-Vis	Data Range			PLS				GILS			R avg
	Min	Max	Range	SECV	SEP	R^2	LVs	SECV	SEP	R^2	
IBP,°C	139.7	114.6	25.1	2.040	2.142	0.886	17	**1.716**	**2.087**	**0.921**	8.4
T5, °C	201.0	180.3	20.2	1.827	2.075	0.785	17	**1.559**	**1.761**	**0.849**	4.4
T10, °C	230.5	211.6	18.9	1.783	2.300	0.802	17	**1.332**	**2.022**	**0.892**	4.8
T20, °C	267.3	249.5	17.8	1.659	2.163	0.802	20	**1.353**	**1.990**	**0.871**	5.4
T30, °C	291.9	272.7	18.8	1.396	1.673	0.849	18	**1.098**	**1.657**	**0.908**	5.0
T40, °C	311.9	289.8	22.1	**1.645**	**1.807**	**0.819**	**17**	1.254	1.820	0.897	4.3
T50, °C	328.9	302.1	26.8	1.544	1.810	0.899	19	**1.262**	**1.743**	**0.934**	4.3
T60, °C	346.5	315.2	31.3	1.597	1.623	0.933	19	**1.290**	**1.570**	**0.958**	4.3
T70, °C	363.8	327.5	36.3	1.617	1.475	0.948	16	**1.227**	**1.471**	**0.970**	4.3
T80, °C	382.3	341.0	41.3	1.750	1.600	0.961	16	**1.341**	**1.510**	**0.977**	4.3
T90, °C	410.5	357.8	52.7	1.918	1.678	0.974	17	**1.438**	**1.667**	**0.986**	4.3
T95, °C	436.0	372.7	63.3	**1.810**	**1.751**	**0.984**	**17**	1.313	1.794	0.992	5.0
FBP, °C	503.0	412.5	90.5	2.823	2.312	0.975	18	**2.145**	**2.045**	**0.986**	11.8
API	35.3	30.7	4.6	0.219	0.349	0.942	18	**0.194**	**0.246**	**0.955**	0.5
Flash P., °C	94.0	75.0	19.0	2.612	2.214	0.684	11	**2.279**	**1.981**	**0.761**	6.0

Similar to developed models with FTIR spectra, GILS algorithm has resulted in lower SEP value for almost all parameters compared to PLS regression. Multivariate calibration results developed with three different spectroscopic method were compared and best model for each parameter was selected according to lowest SEP values. Selected multivariate calibration algorithms and spectroscopic methods are shown in Table 4.6.

Table 4.6. Multivariate calibration results of heavy diesel samples along with selected algorithm and spectroscopic analysis.

	Data Range			Multivariate Calibration Results				Method	Spectroscopy
	Max	**Min**	**Range**	SECV	SEP	R^2	LVs		
IBP	139.7	114.6	25.1	1.716	2.087	0.921	-	GILS	UV-Vis
T5	201.0	180.3	20.2	1.559	1.761	0.849	-	GILS	UV-Vis
T10	230.5	211.6	18.9	0.884	2.010	0.951	14	PLS	FT-NIR
T20	267.3	249.5	17.8	0.829	1.797	0.952	13	PLS	FT-NIR
T30	291.9	272.7	18.8	0.788	1.415	0.951	-	GILS	FT-NIR
T40	311.9	289.8	22.1	1.233	1.541	0.903	-	GILS	FT-NIR
T50	328.9	302.1	26.8	0.859	1.550	0.969	12	PLS	FT-NIR
T60	346.5	315.2	31.3	1.440	1.500	0.946	12	PLS	FT-NIR
T70	363.8	327.5	36.3	1.227	1.471	0.970	-	GILS	UV-Vis
T80	382.3	341.0	41.3	1.341	1.510	0.977	-	GILS	UV-Vis
T90	410.5	357.8	52.7	1.438	1.667	0.986	-	GILS	UV-Vis
T95	436.0	372.7	63.3	1.810	1.751	0.984	17	PLS	UV-Vis
FBP	503.0	412.5	90.5	2.145	2.045	0.986	-	GILS	UV-Vis
API	35.3	30.7	4.6	0.068	0.108	0.996	12	PLS	FT-NIR
Flash P.	94.0	75.0	19.0	0.665	1.751	0.980	7	PLS	FTIR

For lower boiling points, IBP and T5 and higher boiling points starting with T70, GILS multivariate calibration developed with UV-Vis spectroscopic data have resulted in better predictive ability models. Rest of distillation points, models developed with FT-NIR spectral data were selected. FTIR spectral data was only found to have best model for Flash Point.

A total of 30 samples which were separated from distillation point data sets and labelled as *unknowns* were used to observe predictive ability of developed models. For each spectroscopic data, selected and highlighted bold models shown in Table 4.3, 4.4 and 4.5 were used. Model performance of each parameter has been evaluated with reproducibility value of standard methods. Each difference between standard analysis

value and model prediction value, which will be called residuals from now on, was compared to calculated reproducibility value. Table 4.7 represents the number of samples which exceeds reproducibility value.

Table 4.7. Total number of samples which exceeds reproducibility value.

Parameters	Spectroscopy		
	FTIR	**FT-NIR**	**UV-Vis**
IBP	0	0	0
T5	2	4	1
T10	1	1	1
T20	0	0	1
T30	0	0	0
T40	0	0	1
T50	0	0	0
T60	0	0	1
T70	1	0	0
T80	3	0	0
T90	9	4	1
T95	12	5	1
FBP	5	6	0

The results in Table 4.7 show good alignment with results in Table 4.6. Models resulted in lowest SEP values also found to be the ones with lowest residuals and mostly in between reproducibility limits. Obtained results show that shown models in Table 4.6 can be selected as best models. Residual values of 30 unknown samples are shown in Figure 4.8. In each graph black points shows the difference between actual and model prediction values. Red lines represent the reproducibility value of standard method.

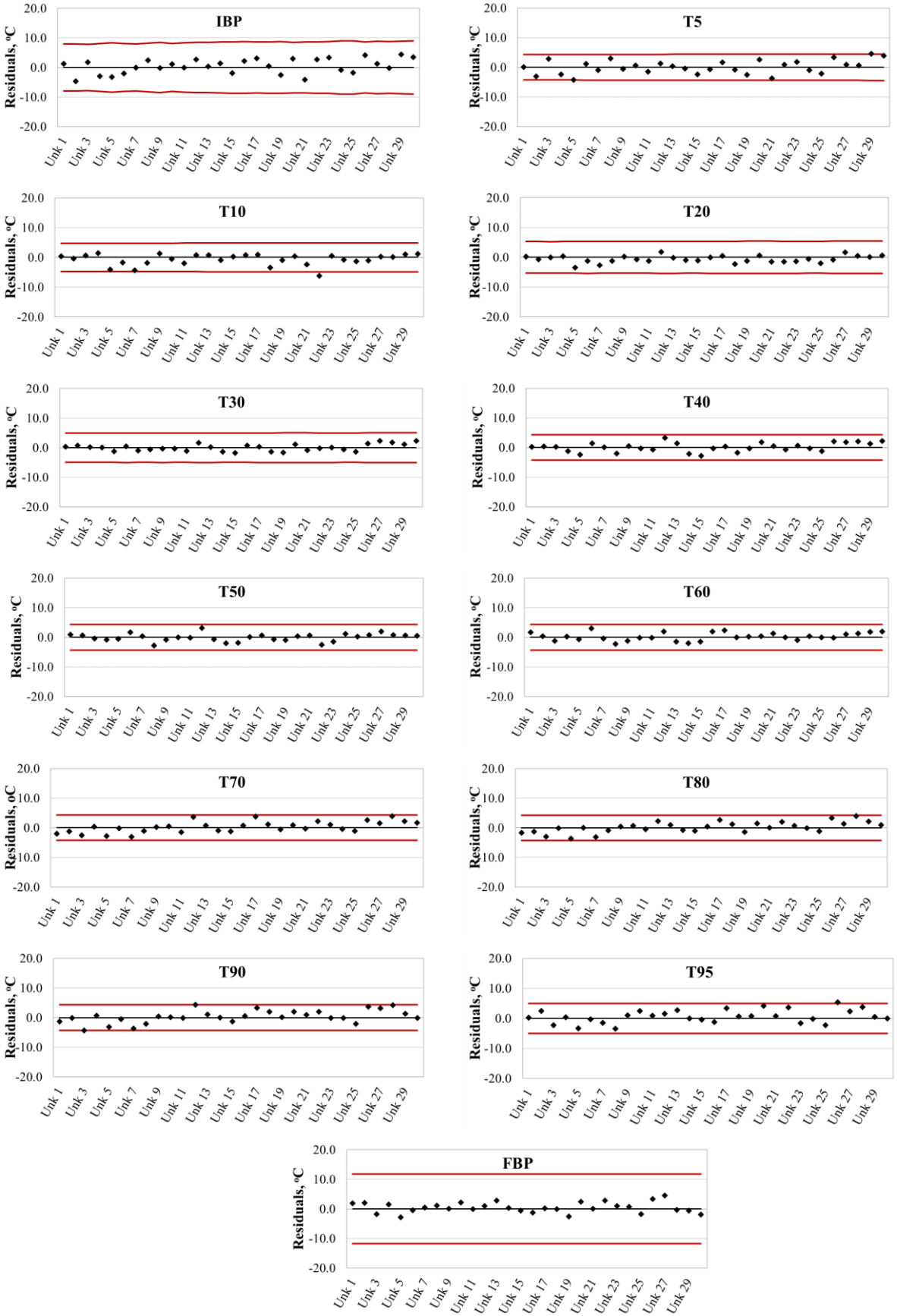

Figure 4.8. Residual graphs of 30 unknown samples from selected models for distillation points.

Residual graphs shown in Figure 4.8 indicated that combination of selected multivariate calibration algorithm and spectral analysis technique, in Table 4.5, have resulted in robust models which can be used instead of standard analysis method. Almost all residual values are in between reproducibility limits and successful predictions have been obtained. Standard method results vs selected model prediction values of developed models are given in Figure 4.9.

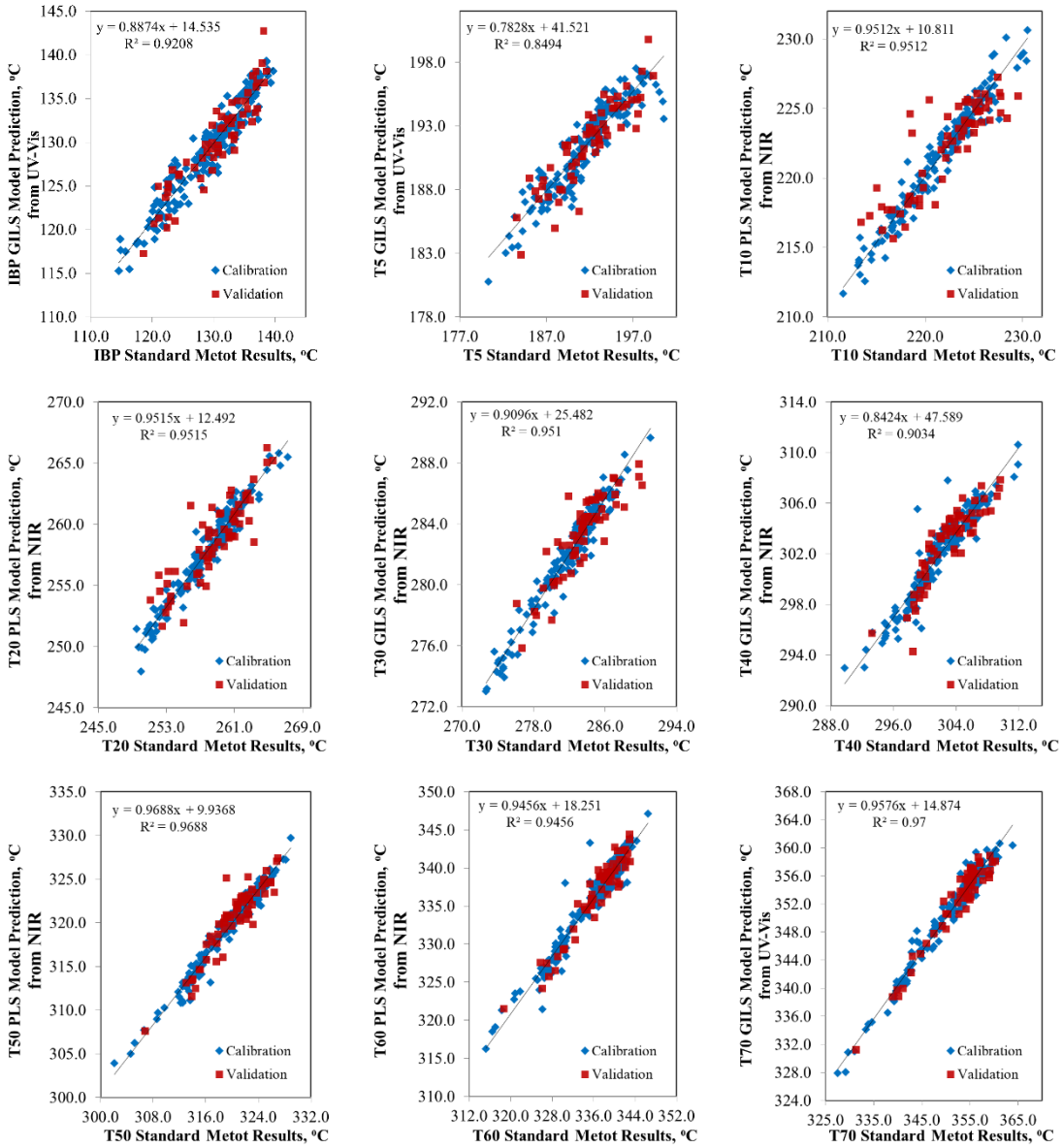

Figure 4.9. Standard analysis vs model prediction results obtained from three different spectroscopic analysis.

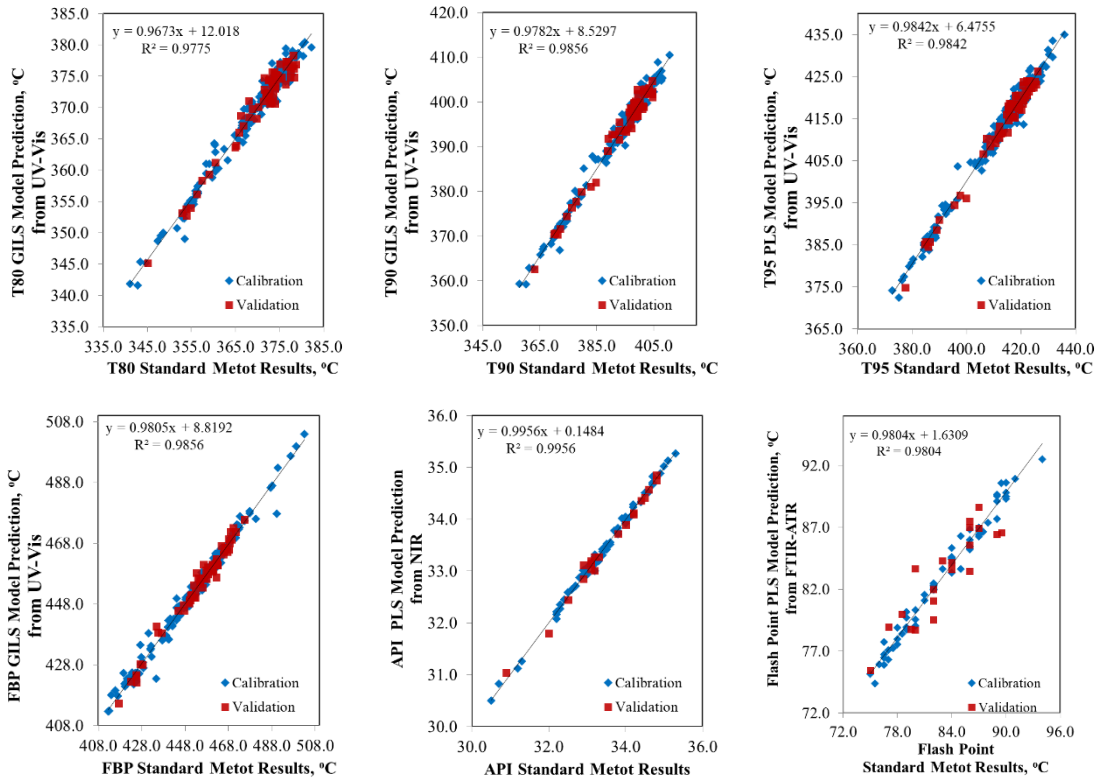

Figure 4.9. Standard analysis vs model prediction results obtained from three different spectroscopic analysis (cont'd).

According to Figure 4.9., validation set predictions of lower boiling points, which are IBP, T5, T10 and T20 are more scattered compared to rest of parameters. Success of models increases with increasing boiling point temperature when R^2 values are considered. Although FBP has the highest reproducibility value, selected model for this parameter has resulted in having one of the highest R^2 value, lowest SEP value and non-residuals of unknown sample is out of reproducibility value shown in Figure 4.8. However, as shown in Table 4.2 and 4.3, with FTIR-ATR and FT-NIR spectroscopy, higher boiling points, especially for FBP, developed models are less successful and has low prediction ability compared to others. Since higher boiling points of sample indicates the boiling point temperature of heavier chemical compounds compared to lower boiling points, models developed with UV-Vis spectroscopy includes and explains more information of those chemical components.

4.2. Light Diesel Samples

Studies with light diesel samples were conducted in two parts. Firstly, three different spectroscopic methods were compared using two different multivariate calibration approaches. Secondly, analyses performed with fluorescence spectroscopy were reported.

4.2.1. Spectroscopic Analysis Comparison

In this part, analysis of light diesel samples obtained from crude distillation unit were performed using three different spectroscopic methods which are FTIR-ATR, FT-NIR and UV-Vis Spectroscopy. Obtained spectral data were treated with different data preprocessing techniques and by the help of two different chemometric approach, best model and spectroscopic method was aimed to be found for physical parameters of light diesel samples.

Light diesel samples were collected for a year in Tüpraş İzmit Refinery. The time period of heavy diesel and light diesel sample collection time was kept same. Although the sample collection time of diesel sample fractions are same, the total of analyzed heavy diesel samples are triple of light diesel samples. Reason of this difference arise from the fact that distillation temperatures of heavy diesel samples were tracked three times of a day. Thus, collected and analyzed heavy diesel sample are much more than light diesel samples. Additionally, each spectral analysis of collected samples are performed in different laboratories. To be able to compare spectroscopic methods, samples were matched and data set with a total of consisting of the same samples was obtained. Physical properties of light diesel samples, distillation temperatures, API gravity and flash point temperatures, were analyzed and reported in quality control laboratory in İzmit Refinery. Physical properties of a total of 75 matched light diesel samples are summarized in Table 4.8. Figure 4.10 shows graphical representation of Table 4.8.

Table 4.8. Data range, mean, median and standard deviation of physical properties of light diesel samples.

	Minimum (Min)	Maximum (Max)	Mean	Median	Standard deviation
IBP, °C	233.6	247.8	240.6	240.5	3.3
T5, °C	246.2	267.7	259.7	259.7	3.9
T10, °C	258.4	274.0	266.7	266.8	3.1
T20, °C	268.9	281.6	275.0	275.3	2.9
T30, °C	274.9	288.0	280.9	281.0	3.0
T40, °C	279.6	294.2	286.7	287.0	3.4
T50, °C	284.0	299.5	291.1	291.4	3.5
T60, °C	287.9	304.8	296.2	296.4	3.9
T70, °C	291.8	310.4	300.5	300.5	4.1
T80, °C	296.3	317.4	306.3	306.3	4.7
T90, °C	299.2	327.5	313.3	313.5	5.6
T95, °C	307.2	338.0	320.4	320.3	6.6
FBP, °C	312.4	346.4	328.2	328.2	7.3
API	32.0	36.8	35.1	35.2	1.0
Flash Point, °C	107.5	130.0	117.5	117.0	5.1

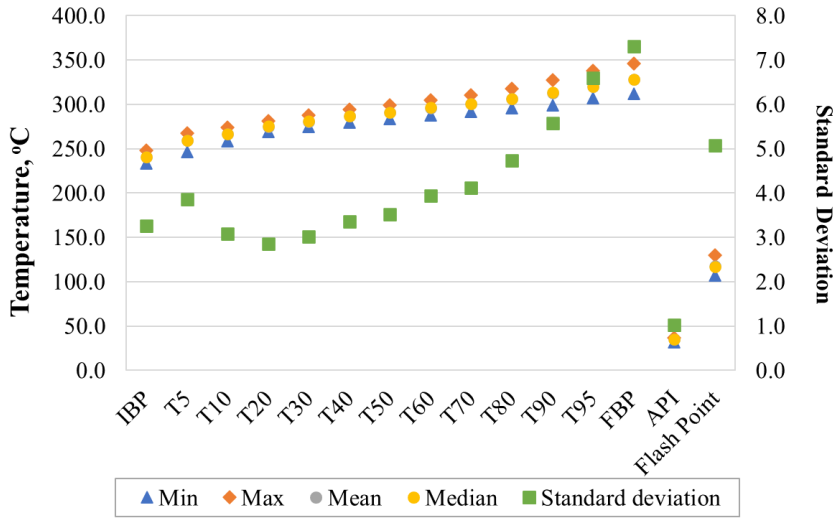

Figure 4.10. Graphical representation of Table 4.8.

Standard deviation values shown in last column in Table 4.8 and green dots in Figure 4.10 indicates that starting with T30, variability of parameter increases with increasing distillation point. API parameter has the lowest variability. Close mean and median values indicate that data has normal distribution.

From 75 light diesel samples, a total of 56 samples were assigned as calibration set to develop model, the rest of 19 samples were used as independent validation set. Unlike heavy diesel study, due to limited number of matched samples, no unknown set was separated to evaluate developed model performances. Instead of this, independent validation set prediction performance was decided to be analyzed. Each data set obtained from spectroscopic analysis were treated with different data pre-processing method and two different multivariate calibration approach were performed to preprocessed data, which will explain under spectroscopic analysis results.

4.2.1.1. FTIR-ATR Spectroscopic Analysis

Raw FTIR-ATR spectra of a total of 75 light diesel samples are shown in Figure 4.11.

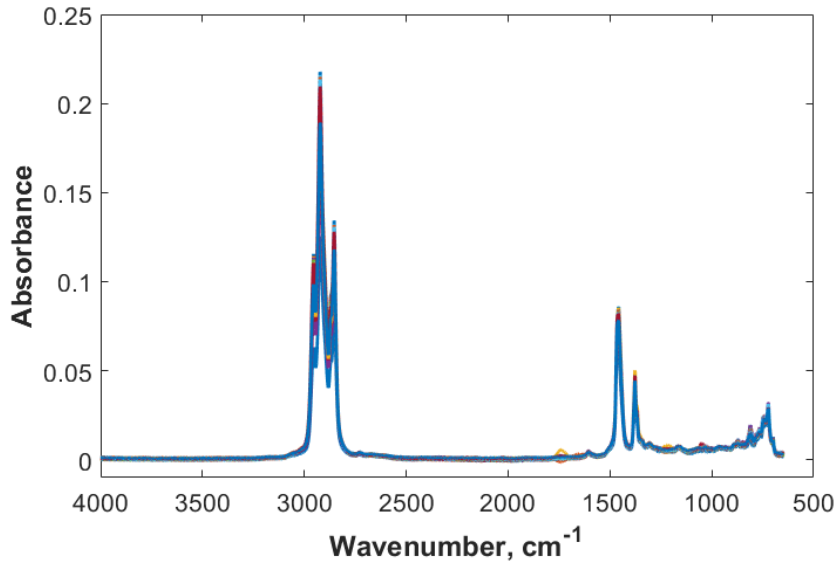

Figure 4.11. Raw FTIR-ATR Spectra of a total of 75 CDU light diesel samples.

As diesel sample fractions has very similar chemical components, raw FTIR-ATR spectra of heavy diesel, in Figure 4.2, shows similar characteristics peak positions with light diesel sample spectra in Figure 4.11. To select best preprocessing method, SG filter, MSC and EMSC were applied iteratively, and best combination of preprocessing method was chosen according to SEP values. Reduction of wavenumber range from 4000 cm^{-1} to 3300 cm^{-1} was resulted in better models, as heavy diesel study. Only MSC was applied

to FTIR-ATR spectra for multivariate calibration models. Figure 4.12 shows preprocessed FTIR-ATR spectra of light diesel samples.

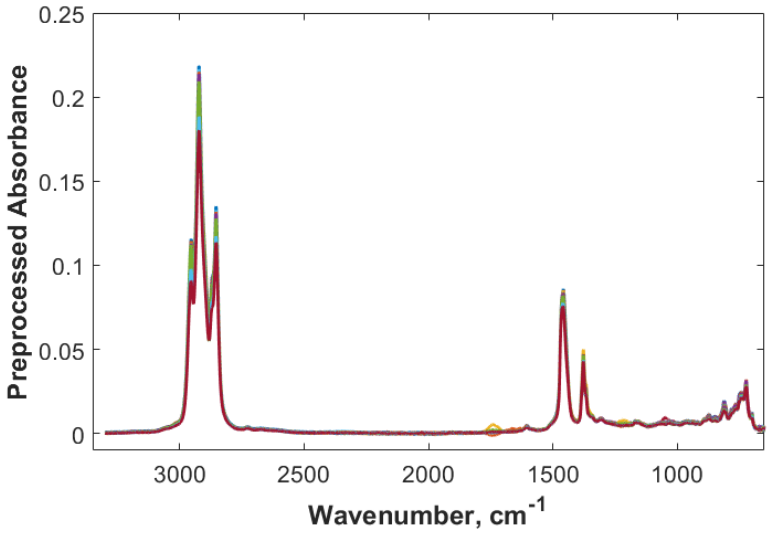

Figure 4.12. Preprocessed FTIR-ATR Spectra of light diesel samples.

Although applied preprocessing algorithm almost show no difference compared to raw spectra, differences which can not be seen by naked eyes found to be result in better models with higher predictive ability. Two different calibration approach were applied to the data set which and PLS and GILS algorithm. Table 4.9 shows the calculated standard error of cross-validation (SECV), standard error of prediction (SEP) and the correlation coefficient of calibration curve, R^2, of each parameter obtained from developed PLS and GILS models. Developed model that has lowest SEP values for each parameter was highlighted as bold font. Reproducibility values for light diesel samples are formulized based on respective ASTM analysis and formulations are shown in Table 4.10. Letter i states the i^{th} sample for the calculation.

Table 4.9. Two different multivariate calibration results of light diesel samples obtained from FTIR-ATR along with data range and average reproducibility value (R avg)

FTIR-ATR	Data Range			PLS				GILS			R avg
	Min	Max	Range	SECV	SEP	R^2	LVs	SECV	SEP	R^2	
IBP,°C	233.6	247.8	14.2	**0.120**	**2.789**	**0.999**	**12**	1.146	2.841	0.949	13.2
T5, °C	246.2	267.7	21.5	1.235	2.407	0.913	7	**1.404**	**2.270**	**0.911**	7.1
T10, °C	258.4	274.0	15.6	**0.730**	**2.019**	**0.953**	**7**	0.951	2.034	0.940	5.9
T20, °C	268.9	281.6	12.7	0.461	1.634	0.977	7	**0.698**	**1.346**	**0.960**	5.7
T30, °C	274.9	288.0	13.1	0.501	1.180	0.978	7	**0.432**	**1.161**	**0.985**	4.6
T40, °C	279.6	294.2	14.6	0.373	1.118	0.990	7	**0.505**	**0.902**	**0.984**	4.0
T50, °C	284.0	299.5	15.5	0.411	1.062	0.989	8	**0.583**	**0.765**	**0.980**	3.0
T60, °C	287.9	304.8	16.9	0.474	1.156	0.989	7	**0.537**	**0.777**	**0.988**	3.5
T70, °C	291.8	310.4	18.6	0.967	1.008	0.956	6	**0.804**	**0.866**	**0.973**	3.8
T80, °C	296.3	317.4	21.1	0.775	1.450	0.979	7	**0.950**	**1.339**	**0.971**	4.2
T90, °C	299.2	327.5	28.3	1.497	1.661	0.942	6	**1.313**	**1.629**	**0.959**	4.7
T95, °C	307.2	338.0	30.8	**1.514**	**2.391**	**0.955**	**6**	1.729	2.412	0.953	7.5
FBP, °C	312.4	346.4	34.0	**1.807**	**2.205**	**0.946**	**6**	1.579	2.336	0.961	7.1
API	32.0	36.8	4.8	0.030	0.130	0.999	9	**0.075**	**0.128**	**0.996**	0.4
Flash P., °C	107.5	130.0	22.5	**3.623**	**4.674**	**0.216**	**3**	2.032	5.268	0.800	8.4

Table 4.10. Reproducibility value calculation for light diesel calculations based on standard method.

Parameters	Light Diesel
	Reproducibility, °C
IBP	0.055 **IBPi**
T5	0.03 **T5i**
T10	0.022 **T10i**
T20	0.0208 **T20i**
T30	0.0165 **T30i**
T40	0.014 **T40i**
T50	3.0
T60	0.0117 **T60i**
T70	0.0125 **T70i**
T80	0.0136 **T80i**
T90	0.015 **T90i**
T95	0.04105 (**T95i** – 140)
FBP	7.1
API	0.6+(0.037 (**API$_i$**-60))
Flash Point	0.071 x **Flash Pi**

Bold values in Table 4.9 indicates that genetic algorithm was resulted in lower SEP values for middle point in distillation curve, T5, T20-90 and flash point analysis. For upper and lower limits, IBP and FBP and API values PLS algorithm was found to have lowest SEP values which indicates better predictive ability. Selection of multivariate calibration method for FTIR-ATR spectra shows similarity to heavy diesel study. However, reproducibility values for light diesel analysis were found higher than heavy diesel samples. Main reason is the difference in standard analysis methods. Distillation temperatures of heavy diesel samples are found by ASTM D2887 in which GC is used while distillation analyses of light diesel samples are obtained by ASTM D86.

4.2.1.2. FT-NIR Spectroscopic Analysis Results

Raw FT-NIR spectra of a total of 75 light diesel samples are shown in Figure 4.13.

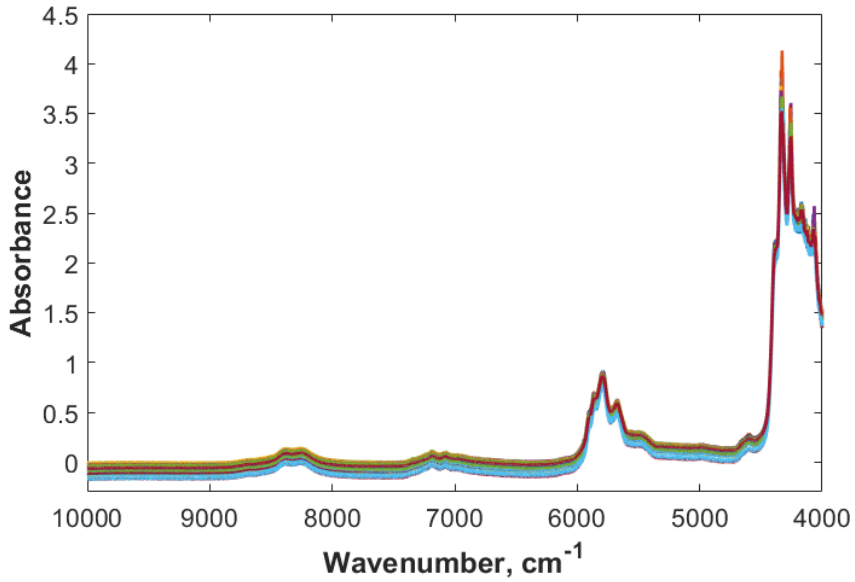

Figure 4.13. Raw FT-NIR Spectra of a total of 75 light diesel samples.

Similar to all FT-NIR spectra reported in this thesis, high absorbance values due to saturation were observed in wavenumbers between 4500 cm^{-1} and 4000 cm^{-1}. These wavenumbers were removed from dataset as well as 12000-9000 cm^{-1}.To enhance signal properties and reduce the effect of baseline shift, three different preprocessing method

were applied to this data set as well. EMSC method with smoothing with third order polynomial fitting and 9-point window size was selected to tread FT-NIR spectra for light diesel calibration study. Preprocessed spectra are shown in Figure 4.14.

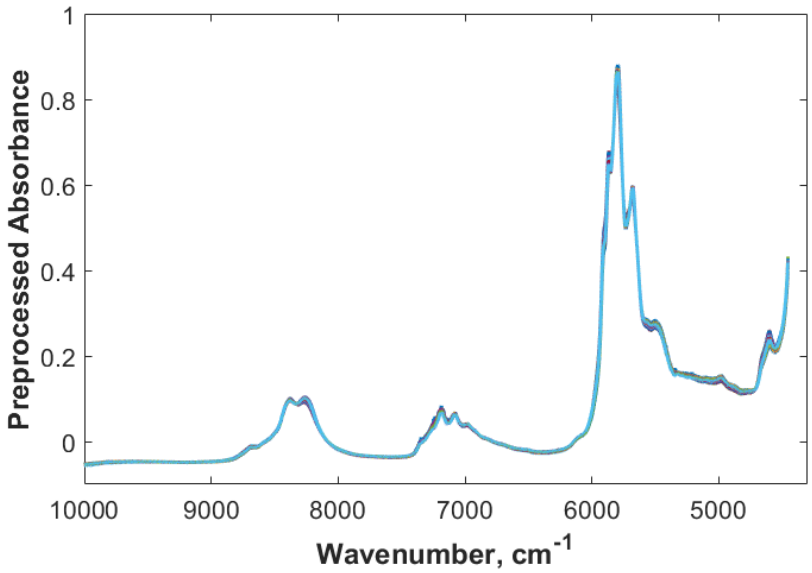

Figure 4.14. Preprocessed FT-NIR spectra of light diesel samples.

Two different calibration approaches, PLS and GILS, were also applied to spectra shown in Figure 4.14. Multivariate calibration results are given in Table 4.11 with standard error of cross-validation (SECV), standard error of prediction (SEP) and the correlation coefficient of calibration curve, R^2, along with latent variables for PLS models. For each parameter, multivariate calibration model with lowest SEP values was highlighted as bold font.

For most of the parameters, PLS algorithm was chosen similar to heavy diesel FT-NIR study. However, GILS was found to have low SEP values for several distillation points, unlike to heavy diesel study results shown in Table 4.3. As indicated in other multivariate calibration results, SEP values gets higher with high reproducibility values.

Table 4.11. Two different multivariate calibration results of light diesel samples obtained from FT-NIR along with data range and average reproducibility value (R, avg)

FT-NIR	Data Range			PLS				GILS			R avg
	Min	Max	Range	SECV	SEP	R^2	LVs	SECV	SEP	R^2	
IBP, °C	233.6	247.8	14.2	**2.529**	**2.860**	**0.418**	**5**	1.968	2.942	0.741	13.2
T5, °C	246.2	267.7	21.5	1.455	2.412	0.880	11	**3.004**	**1.894**	**0.642**	7.1
T10, °C	258.4	274.0	15.6	0.308	1.571	0.992	14	**0.884**	**1.385**	**0.946**	5.9
T20, °C	268.9	281.6	12.7	1.004	1.042	0.904	8	**0.991**	**0.834**	**0.919**	5.7
T30, °C	274.9	288.0	13.1	**0.307**	**0.939**	**0.992**	**13**	0.816	0.956	0.947	4.6
T40, °C	279.6	294.2	14.6	**0.355**	**0.719**	**0.992**	**12**	0.815	0.837	0.956	4.0
T50, °C	284.0	299.5	15.5	**0.376**	**1.061**	**0.991**	**13**	1.015	1.179	0.940	3.0
T60, °C	287.9	304.8	16.9	0.357	1.020	0.994	12	**0.799**	**1.007**	**0.971**	3.5
T70, °C	291.8	310.4	18.6	**0.365**	**1.335**	**0.994**	**13**	1.359	1.432	0.915	3.8
T80, °C	296.3	317.4	21.1	**0.450**	**1.055**	**0.993**	**12**	1.192	1.553	0.951	4.2
T90, °C	299.2	327.5	28.3	**0.659**	**1.683**	**0.988**	**12**	1.497	1.771	0.946	4.7
T95, °C	307.2	338.0	30.8	**0.746**	**2.086**	**0.989**	**12**	2.132	2.685	0.907	7.5
FBP, °C	312.4	346.4	34.0	**0.943**	**2.539**	**0.986**	**13**	2.103	2.760	0.937	7.1
API	32.0	36.8	4.8	**0.057**	**0.096**	**0.997**	**12**	0.109	0.100	0.990	0.4
Flash P., °C	107.5	130.0	22.5	3.481	4.806	0.400	5	**2.712**	**4.681**	**0.684**	8.4

4.2.1.3. UV-Vis Spectroscopic Analysis Results

Raw UV-Vis spectra of light diesel samples is shown in Figure 4.15.

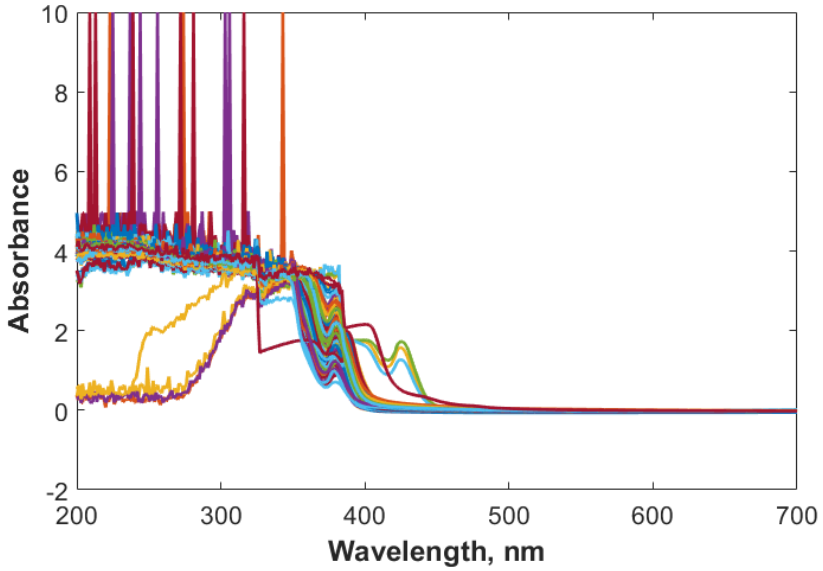

Figure 4.15. Raw UV-Vis Spectra of a total of 75 CDU light diesel samples.

Figure 4.15 shows high absorption values that can not be used for calibration studies between wavenumbers from 200 to 385 nm. Those wavenumbers were removed before treatment with any preprocessing technique. After reduction of those wavenumbers, best preprocessing technique was selected by same iterative method as other studies and first derivative with third order polynomial fitting and 9-point window size following MSC was chosen. UV-Vis spectra treated with selected preprocessing technique was shown in Figure 4.16.

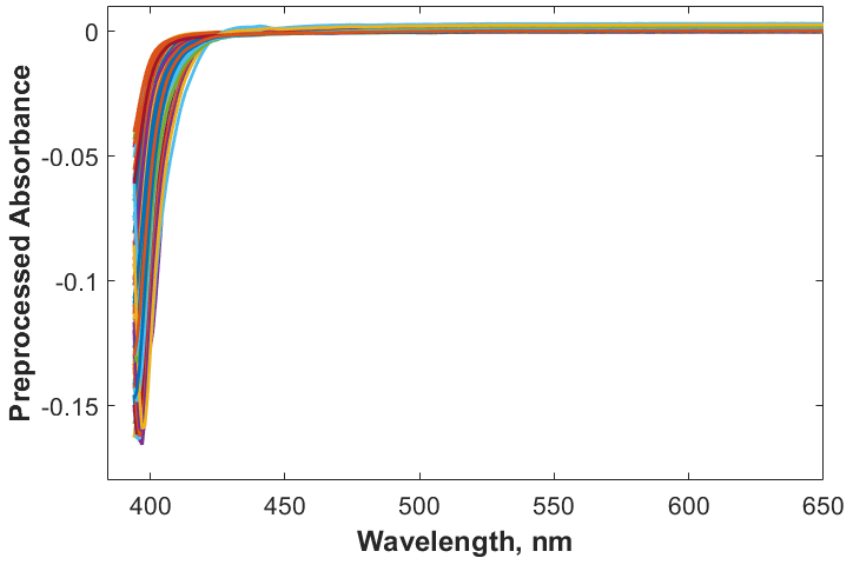

Figure 4.16. Preprocessed UV-Vis Spectra of light diesel samples.

To spectral data shown in Figure 4.16, two different calibration approaches were applied as well. Table 4.12. shows obtained multivariate calibration model results along with and the correlation coefficient of calibration curve, R^2, selected latent variables and reproducibility values.

Table 4.12. Multivariate calibration results of light diesel samples obtained from UV-Vis along with data range and average reproducibility value (R, avg)

UV-Vis	Data Range			PLS				GILS			R avg
	Min	Max	Range	SECV	SEP	R^2	LVs	SECV	SEP	R^2	
IBP,°C	233.6	247.8	14.2	1.799	4.579	0.722	12	**2.069**	**2.431**	**0.711**	13.2
T5, °C	246.2	267.7	21.5	1.153	3.301	0.909	15	**1.796**	**2.557**	**0.826**	7.1
T10, °C	258.4	274.0	15.6	1.991	2.067	0.511	7	**1.216**	**1.969**	**0.860**	5.9
T20, °C	268.9	281.6	12.7	**1.790**	**1.659**	**0.581**	**7**	1.282	2.533	0.821	5.7
T30, °C	274.9	288.0	13.1	**1.839**	**1.954**	**0.645**	**6**	1.562	2.360	0.754	4.6
T40, °C	279.6	294.2	14.6	**2.192**	**2.530**	**0.644**	**5**	1.944	2.761	0.733	4.0
T50, °C	284.0	299.5	15.5	**2.142**	**2.310**	**0.658**	**6**	1.868	2.918	0.754	3.0
T60, °C	287.9	304.8	16.9	2.453	3.073	0.673	5	**2.102**	**2.419**	**0.771**	3.5
T70, °C	291.8	310.4	18.6	1.747	2.957	0.824	6	**1.809**	**2.424**	**0.817**	3.8
T80, °C	296.3	317.4	21.1	2.368	3.473	0.752	4	**1.958**	**3.427**	**0.836**	4.2
T90, °C	299.2	327.5	28.3	2.825	3.648	0.744	6	**2.743**	**2.708**	**0.763**	4.7
T95, °C	307.2	338.0	30.8	2.610	4.124	0.816	7	**2.521**	**3.860**	**0.833**	7.5
FBP, °C	312.4	346.4	34.0	**5.013**	**4.801**	**0.530**	**4**	3.117	5.049	0.822	7.1
API	32.0	36.8	4.8	**0.624**	**0.464**	**0.678**	**9**	0.381	0.558	0.906	0.4
Flash P., °C	107.5	130.0	22.5	4.100	5.642	0.347	8	**2.418**	**4.622**	**0.810**	8.4

Selected multivariate techniques for UV-Vis data are found to be similar to FTIR-ATR study. For most of the parameters, GILS method has resulted in lower SEP values compared to PLS models. However, for API, PLS was selected while for flash point GILS results are found to have lower SEP values.

After development of different multivariate calibration models using different spectral data, all models for each parameter were compared and best model for selected parameter was chosen according to SEP values. Table 4.13 shows selected multivariate calibration method and spectroscopic measurements for each parameter.

Table 4.13. Multivariate calibration results of heavy diesel samples along with selected algorithm and spectroscopic analysis.

	Data Range			Multivariate Calibration Results				Method	Spectroscopy
	Max	Min	Range	SECV	SEP	R^2	LVs		
IBP, °C	233.6	247.8	14.2	2.069	2.431	0.711	-	GILS	UV-Vis
T5, °C	246.2	267.7	21.5	3.004	1.894	0.642	-	GILS	FT-NIR
T10, °C	258.4	274.0	15.6	0.884	1.385	0.946	-	GILS	FT-NIR
T20, °C	268.9	281.6	12.7	0.991	0.834	0.919	-	GILS	FT-NIR
T30, °C	274.9	288.0	13.1	0.307	0.939	0.992	13	PLS	FT-NIR
T40, °C	279.6	294.2	14.6	0.355	0.719	0.992	12	PLS	FT-NIR
T50, °C	284.0	299.5	15.5	0.583	0.765	0.980	-	GILS	FTIR
T60, °C	287.9	304.8	16.9	0.537	0.777	0.988	-	GILS	FTIR
T70, °C	291.8	310.4	18.6	0.804	0.866	0.973	-	GILS	FTIR
T80, °C	296.3	317.4	21.1	0.450	1.055	0.993	12	PLS	FT-NIR
T90, °C	299.2	327.5	28.3	1.313	1.629	0.959	-	GILS	FTIR
T95, °C	307.2	338.0	30.8	0.746	2.086	0.989	12	PLS	FT-NIR
FBP, °C	312.4	346.4	34.0	1.807	2.205	0.946	6	PLS	FTIR
API	32.0	36.8	4.8	0.057	0.096	0.997	12	PLS	FT-NIR
Flash P., °C	107.5	130.0	22.5	2.418	4.622	0.810	-	GILS	UV-Vis

Results shown in Table 4.13 shows some similarity to Table 4.5, in which multivariate calibration results of heavy diesel samples are shown. First of all, for both sample groups, GILS applied to UV-Vis data and PLS applied to FT-NIR data has resulted in better models for IBP and API, respectively. On the other hand, for the rest of the parameters no same combination was selected. Since light diesel and heavy diesel samples are very similar composition to each other, selection of different methods and spectroscopic analysis are not expected. Main reason of this difference can be because of difference in standard method analysis. Standard analysis of distillation point analysis of heavy diesel samples is ASTM D2887 which analyzed ana give results in weight percent while standard analyses of light diesel samples are reported in volumetric percent. This difference also can be seen in different reproducibility values as well as initial boiling point temperatures. As light diesel samples have lower number of hydrocarbon chains compared to heavy diesel, it is expected that initial boiling point temperature of light diesel should be lower than heavy diesel. However, as can be seen in Table 4.4 and Table 4.13 in data range section, IBP value of light diesel is higher than heavy diesel. This difference also because of difference in ASTM methods. Standard method results vs selected model prediction values of developed models are given in Figure 4.17.

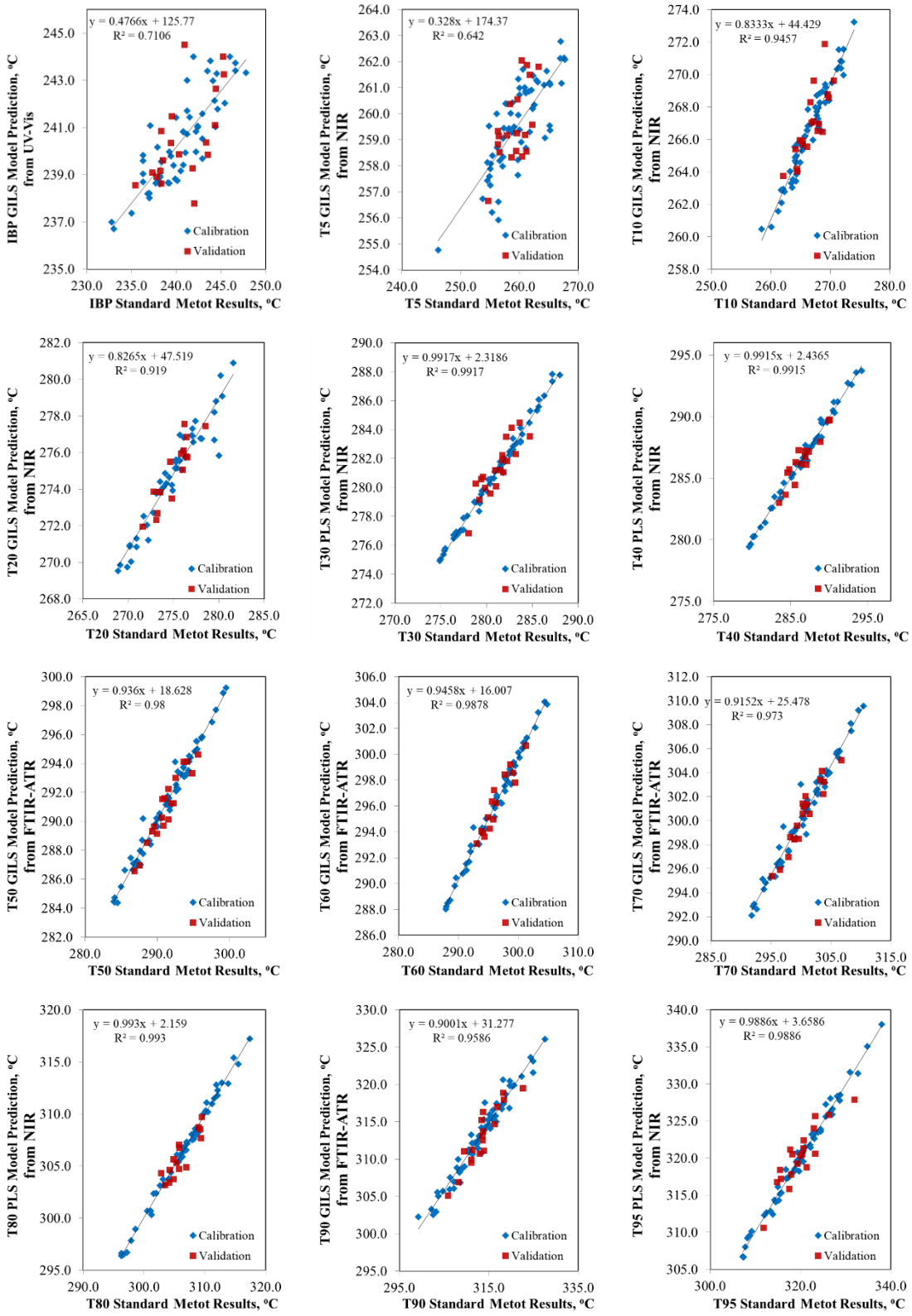

Figure 4.17. Standard analysis vs model prediction results obtained from three different spectroscopic analysis.

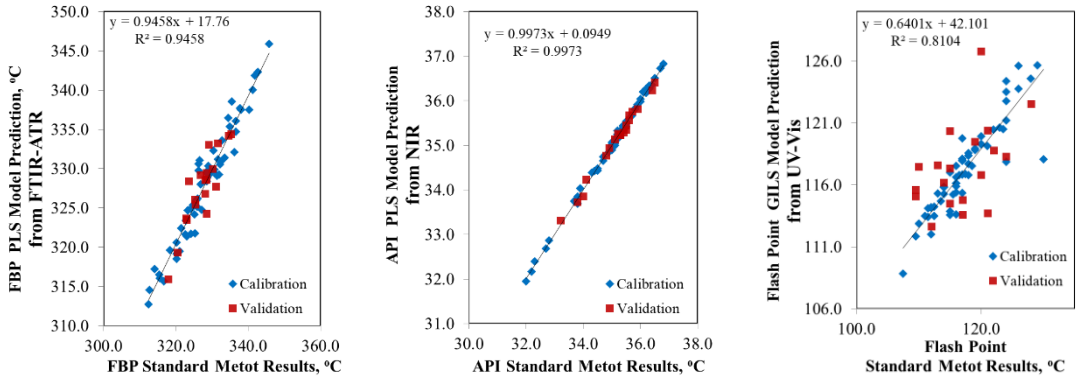

Figure 4.17. Standard analysis vs model prediction results obtained from three different spectroscopic analysis. (cont'd)

Although the lowest temperature in T5 parameter increases the dynamic range of calibration model and increases the R^2 value shown in Figure 4.17, lowest R^2 was obtained from the developed model for T5. Additionally, calibration graph of T5 shows nonlinear behavior in low temperatures. Most scattered graphs and as a result less successful models were obtained for IBP, T5 and flash point parameters.

4.2.2. Fluorescence Analysis

In this section, studies that have been conducted with light diesel samples obtained from crude distillation unit in İzmit Refinery.

A total of 116 light diesel samples obtained from crude distillation unit were collected over a year. Distillation temperatures, API gravity and flash point of samples were analyzed in quality control laboratory in İzmit Refinery according to their respective ASTM methods. Summary of primary analysis results of collected samples is shown in Table 4.14.

Table 4.14. Data range, mean, median and standard deviation of physical properties of light diesel samples.

	Min	Max	Mean	Median	Stdev
IBP, °C	191.4	223.6	205.0	204.5	6.8
T5, °C	217.3	242.2	228.1	227.5	5.4
T10, °C	227.2	250.9	238.2	237.8	5.1
T30, °C	246.8	267.2	257.7	257.9	4.3
T50, °C	257.7	280.7	271.0	271.1	4.5
T70, °C	262.1	293.4	282.5	282.7	5.5
T90, °C	279.2	313.3	299.1	299.5	6.9
T95, °C	281.5	326.3	308.0	308.2	8.5
FBP, °C	291.6	338.1	318.4	317.9	9.6
API, at 60F	34.0	39.1	37.3	37.7	1.2
Flash Point, °C	81.0	104.0	91.1	91.0	4.8

Table 4.14 indicates that properties of samples have changed during the year. Standard deviation of each parameter shows that lower (such as initial boiling point, IBP and T5) and higher cut point (T95 and FBP) of light diesel samples are the most changed parameters. Mean and median values show the data set has a symmetrical distribution.

Fluorescence analysis of diesel samples were carried out in two different mode which are total fluorescence and synchronous fluorescence modes.

In total fluorescence mode, emission spectrum at different excitation wavelengths were recorded. Since diesel samples have complex hydrocarbon molecules, emission spectra can change while changing the excitation wavelength. In synchronous fluorescence mode, both the emission and excitation wavelengths are scanned simultaneously. Emission and excitation wavelength are chosen at the beginning of the measurement along with wavelength difference which specifies increment number of wavelength after each scan.

As stated in the literature lighter petroleum oils exhibit more intense emission then the heavier oils because of quenching process.[51,52] The reason of that samples has the highest and lowest API values were selected to be shown. The TFS results of these two samples are depicted in Figure 4.18 and 4.19.

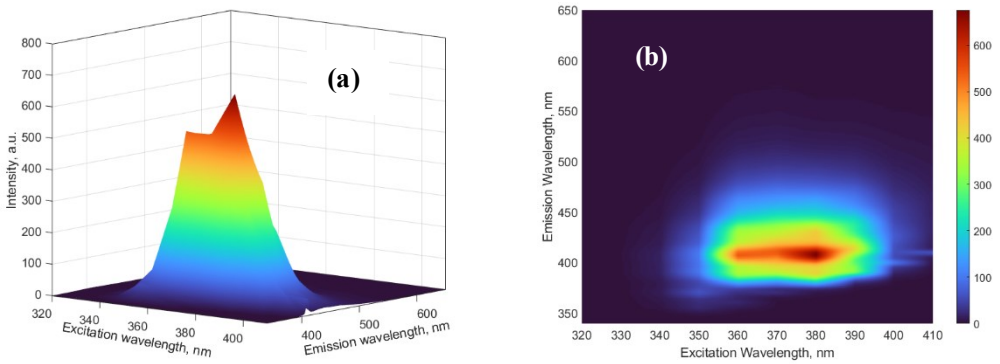

Figure 4.18. Total Fluorescence Spectra of light diesel sample with highest API value (a) 3D topographical diagram (b) contour map

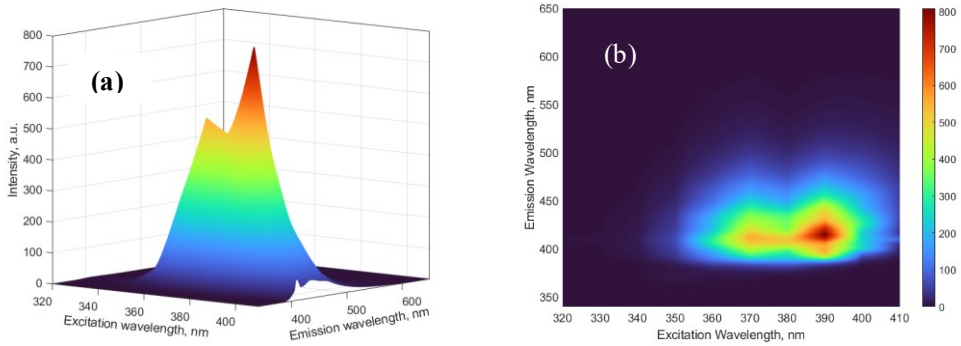

Figure 4.19. Total Fluorescence Spectra of light diesel sample with lowest API value (a) 3D topographical diagram (b) contour map.

Although studies in literature states that lighter crude oil with higher API values exhibit stronger fluorescence emissions than heavier crude oils with lower API values, Figure 4.18 and 4.19 show the opposite. It is also stated that fluorescence emission of sample is not only depend on API grade, but also depend on fluorescent aromatic compounds. As shown in also Figure 4.18 and 4.19, fluorescence intensity of light diesel sample with lowest API (Figure 4.19.a.) is higher than the sample with highest API (Figure 4.19.b.) Contour maps shown in Figure 4.18.b and 4.19b indicates the red-shift in emission maximum. The emission maximum of Figure 4.18b is located in 408 nm While emission maximum of Figure 4.19.b. is located around 420 nm because of enhanced energy transfer.

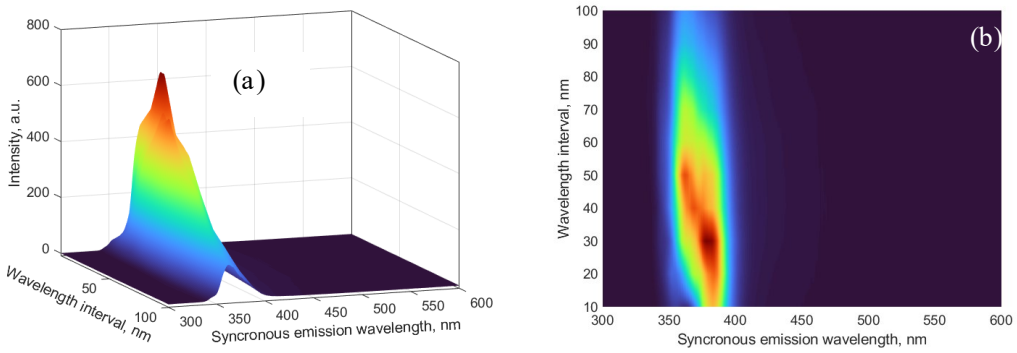

Figure 4.20. Synchronous Fluorescence Spectra of light diesel sample with highest API value (a) 3D topographical diagram (b) contour map.

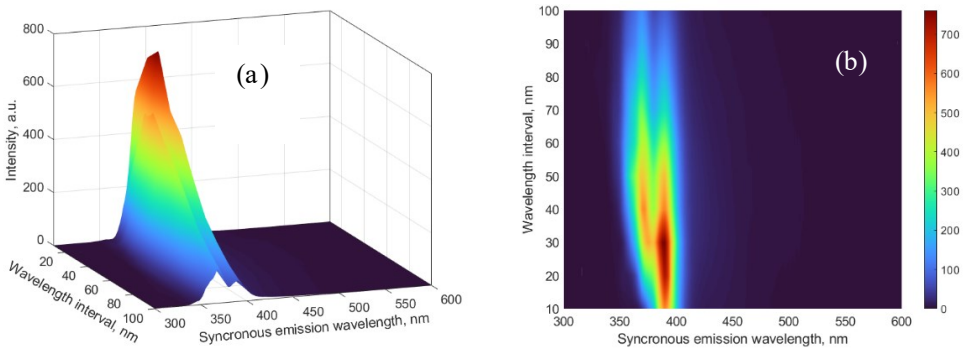

Figure 4.21. Synchronous Fluorescence Spectra of light diesel sample with lowest API value (a) 3D topographical diagram (b) contour map.

The trend of a net increase in fluorescence intensity for heavier product observed in total fluorescence spectra is also observed in synchronous fluorescence shown in Figure 4.20 and Figure 4.21. Unlike total fluorescence, synchronous fluorescence spectra show better defined shoulders which might contain valuable information that will be discussed in multivariate calibration section. This phenomenon is also stated in literature as in synchronous mode sample spectra does not present the artifacts due to Rayleigh and Raman scattering. Similar to total fluorescence, red shift in the emission is observed upon going from lighter to heaver product.

Both data sets, total fluorescence spectra and synchronous spectra, were arranged into matrix characterized by samples as columns and fluorescence intensities as rows to proceed with multivariate calibration process.

Two different calibration approaches were performed which are Genetic Inverse Least squares (GILS) and Partial Least Squares (PLS) Regression. From 116 samples, a total of 86 were assigned as calibration set to develop model and the rest of 71 samples were used as independent validation set to observed prediction ability of the model. After

developing PLS and GILS models for each data set, standard error of cross-validation (SECV), standard error of prediction (SEP) and the correlation coefficient of calibration curve, R^2, of each parameter were calculated and compared. Table 4.15 show multivariate calibration results obtained from total fluorescence spectra while Table 4.16 shows results obtained from synchronous fluorescence spectra of light diesel samples.

Table 4.15. Multivariate calibration results of total fluorescence spectra of light diesel samples.

TFS Mode	Data Range			Multivariate Calibration Results							R (Avg)
				PLS				GILS			
	Min	Max	Range	SECV	SEP	R^2	LVs	SECV	SEP	R^2	
IBP, °C	191.4	223.6	28.7	3.078	4.797	0.782	8	2.191	4.635	0.893	11.3
T5, °C	217.3	242.2	24.9	2.017	2.509	0.871	13	1.653	2.771	0.918	6.9
T10, °C	227.2	250.9	23.7	1.491	2.289	0.920	14	1.351	2.474	0.937	5.3
T30, °C	246.8	267.2	20.4	0.577	1.364	0.983	20	1.335	1.520	0.916	4.2
T50, °C	257.7	280.7	23.0	**0.716**	**1.208**	**0.976**	**17**	1.031	1.351	0.952	3.0
T70, °C	262.1	293.4	31.3	0.898	1.694	0.974	17	1.324	1.962	0.946	3.5
T90, °C	279.2	313.3	34.1	**0.748**	**1.434**	**0.988**	**17**	1.079	1.559	0.976	4.4
T95, °C	281.5	326.3	44.8	1.139	1.949	0.982	16	1.296	1.738	0.977	6.7
FBP, °C	291.6	338.1	46.5	**0.463**	**0.901**	**0.998**	**19**	0.942	1.159	0.990	7.1
API	34.0	39.1	4.9	0.161	0.332	0.982	17	0.244	0.355	0.961	3.3
Flash Point, °C	81.0	104.0	23.0	**1.174**	**1.751**	**0.941**	**17**	1.833	1.862	0.863	6.6

Table 4.16. Multivariate calibration results of synchronous fluorescence spectra of light diesel samples.

SFS Mode	Data Range			Multivariate Calibration Results							R (Avg)
				PLS				GILS			
	Min	Max	Range	SECV	SEP	R^2	LVs	SECV	SEP	R^2	
IBP, °C	191.4	223.6	28.7	**3.931**	**4.007**	**0.674**	**8**	3.142	4.168	0.796	11.3
T5, °C	217.3	242.2	24.9	1.762	2.614	0.893	13	**1.643**	**2.386**	**0.910**	6.9
T10, °C	227.2	250.9	23.7	**1.672**	**1.918**	**0.900**	**13**	1.440	2.021	0.928	5.3
T30, °C	246.8	267.2	20.4	0.950	1.332	0.953	16	**1.175**	**1.169**	**0.931**	4.2
T50, °C	257.7	280.7	23.0	0.853	1.220	0.966	17	1.183	1.208	0.936	3.0
T70, °C	262.1	293.4	31.3	**0.860**	**1.667**	**0.967**	**15**	1.019	1.719	0.956	3.5
T90, °C	279.2	313.3	34.1	0.888	1.462	0.984	15	1.210	1.495	0.970	4.4
T95, °C	281.5	326.3	44.8	1.482	1.655	0.967	12	**1.387**	**1.652**	**0.972**	6.7
FBP, °C	291.6	338.1	46.5	1.545	1.063	0.974	12	1.233	1.519	0.984	7.1
API	34.0	39.1	4.9	**0.169**	**0.298**	**0.982**	**17**	0.251	0.333	0.961	3.3
Flash Point, °C	81.0	104.0	23.0	1.627	2.221	0.893	15	1.731	2.394	0.885	6.6

For both total fluorescence and synchronous measurements, almost for all of the parameters, PLS multivariate calibration results resulted in lower standard error of prediction values than GILS results. When both calibration and prediction standard errors were analyzed separately, highest prediction errors were observed in parameters have higher reproducibility values. It is an expected outcome as developed models highly based on primary analysis results.

When calibration results of two different fluorescence mode were compared, models developed from synchronous fluorescence spectra has resulted in lower calibration and prediction errors. This might be the results of better-defined contours with shoulders shown in Figure 4.20 and Figure 4.21. Standard method results vs selected model prediction values of developed models are given in Figure 4.22.

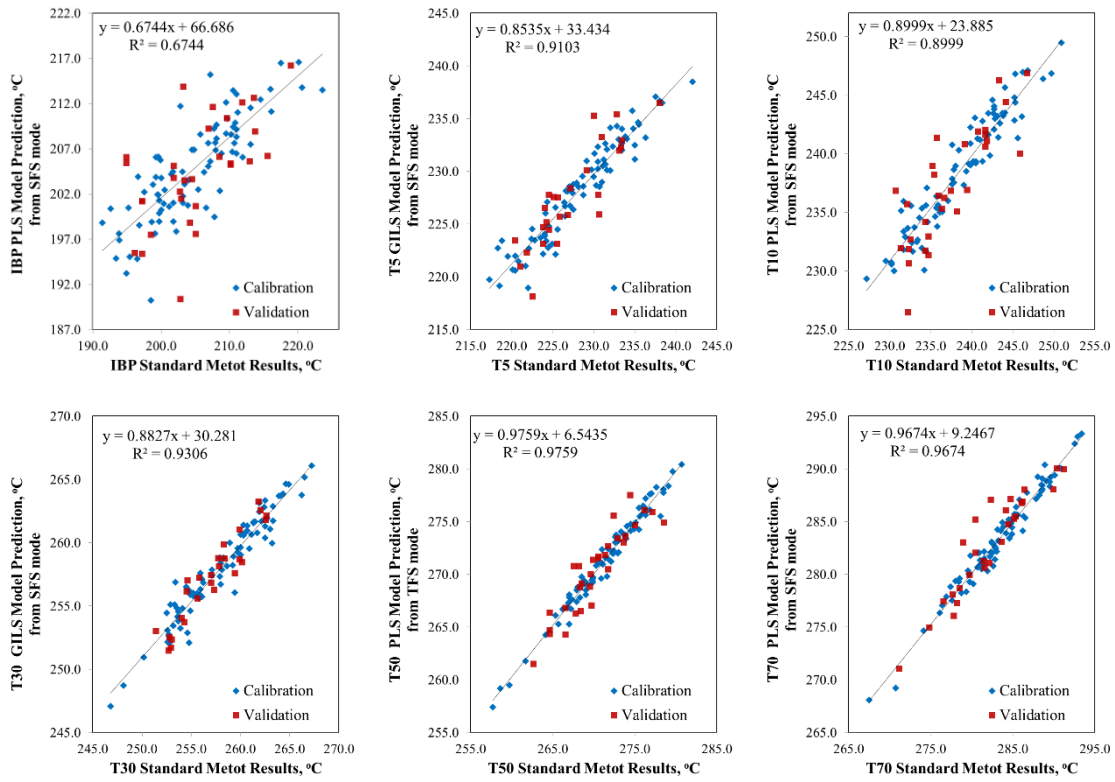

Figure 4.22. Standard analysis vs model prediction results obtained from two different measurement mode of fluorescence spectroscopy.

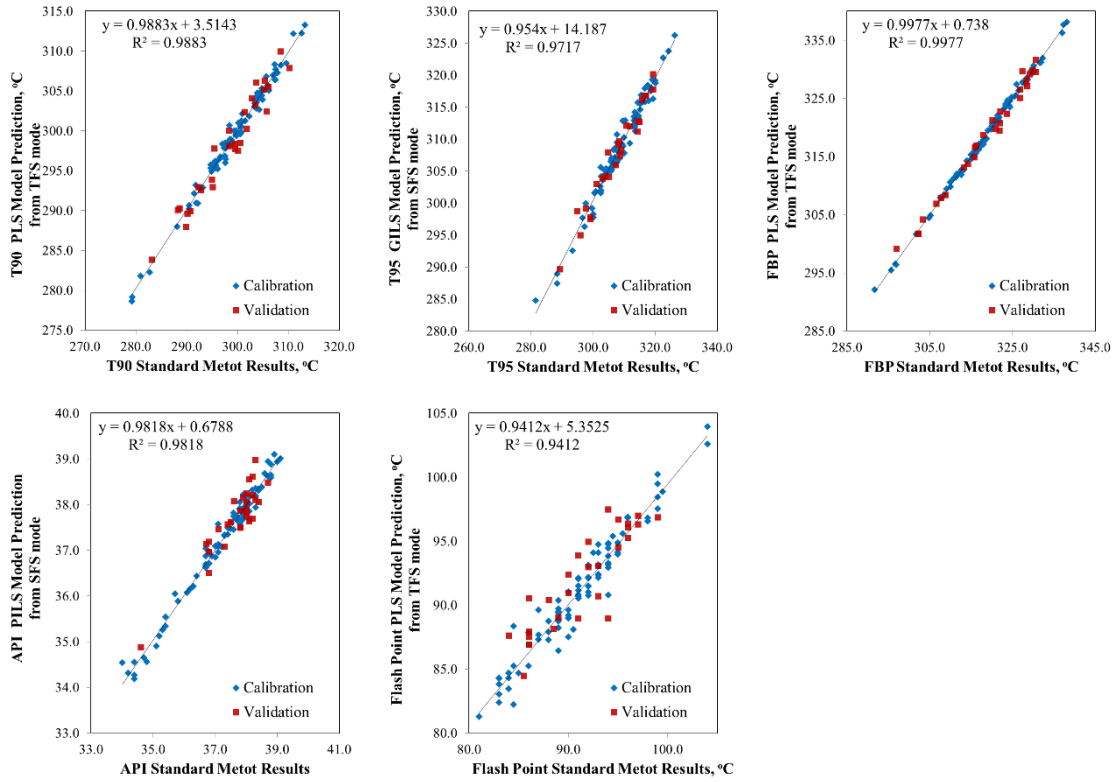

Figure 4.22. Standard analysis vs model prediction results obtained from two different measurement mode of fluorescence spectroscopy (cont'd).

Figure 4.22 shows the actual versus model prediction graphs of selected measurement mode and multivariate calibration algorithm for all parameters. Developed model of IBP are observed to have lowest R2 value which can be explained by reproducibility value of IBP. However, although the reproducibility value of FBP is high and close to reproducibility value of T5, TFS mode measurement combined with PLS model has resulted in best calibration models of all distillation temperatures considering R^2 and SEP values.

4.3. Straight Run Naphtha Samples

In this section, studies that have been conducted with heavy straight run naphtha (HSRN) and light straight run naphtha (LSRN) samples obtained from three different crude distillation units belonging to three different refineries which are İzmit Refinery, İzmir Refinery and Kırıkkale Refinery. Naphtha is colorless, volatile, and flammable liquid obtained by distillation of crude oil in atmospheric distillation column. It is an

important light distillate and used as feedstock in petroleum and petrochemical industry such as in gasoline blending and monomer synthesis for the polymers.

Two different types of naphtha are obtained in the process regarding as cut point temperatures: the first type having five to six carbon containing hydrocarbons is called Light Straight Run Naphtha (LSRN), the second type called Heavy Straight Run Naphtha (HSRN) could have six to ten carbon atoms in the molecular structure. LSRN can be used as a feedstock for isomerization unit while HSRN can be used as a feedstock for platformer unit. Both two units are designed to increase octane number of the gasoline.

HSRN and LSRN samples were collected daily from three different crude oil units in three different refineries. Distillation temperatures, API gravity and vapor pressure of the samples were analyzed in quality control laboratory located in each refinery, according to their respective ASTM methods.

In İzmit refinery, a total of 301 HSRN samples and a total of 276 LSRN samples were collected between May 2018 and July 2021. In İzmir refinery, a total of 281 HSRN samples and a total of 256 LSRN samples and were collected between August 2019 and September 2021. Lastly, in Kırıkkale refinery, and a total of 231 HSRN samples and a total of 232 LSRN samples were collected between April 2019 and September 2021. The distillation temperatures ranges at different recovery level (from initial boiling point (IBP) to final boiling point (FBP)) along with API gravity and vapor pressure are summarized in Table 4.8. Distillation temperatures was obtained using ASTM D86 method, while API values were measured EN ISO 2719 and vapor pressure values were obtained using IP 394 method.

As shown in Table 4.17, although the name of samples is the same for each refinery, the physical properties of samples differ from each other. When the distillation curves were investigated, although HSRN samples of İzmir refinery has the lowest initial boiling point, final boiling point of HSRN samples is not much different from other refineries. The HSRN samples produced by İzmir refinery contain a wider variety of chemical structures than other refineries when considering the difference between the initial boiling point and final boiling point. Considering standard deviation of each sample groups, there is less production variation in Kırıkkale refinery. Distillation temperatures of HSRN samples are expected to be higher than LSRN samples, since they obtain hydrocarbons with a higher carbon number. However, in İzmir refinery, final boiling point of LSRN samples is higher than HSRN samples. The reason of that is the design of naphtha splitter in CDU have been designed differently in İzmir refinery in accordance

with the production. It is also known that, in İzmit and İzmit refinery, number of different crude oil which are processed in crude distillation units higher than Kırıkkale Refinery. Because of this reason, physical properties of samples produced in Kırıkkale refinery do not change as much as other refineries.

Table 4.17. Data range, average and standard deviation of physical parameters of all refinery's naphtha samples.

		İzmit HSRN				İzmir HSRN				Kırıkkale HSRN			
		Min	**Max**	**Mean**	**Standard deviation**	**Min**	**Max**	**Mean**	**Standard deviation**	**Min**	**Max**	**Mean**	**Standard deviation**
IBP, °C		81	107.6	93.5	4.9	35.3	65.5	45.8	3.9	84.4	102.5	93.2	3.3
T5, °C		93.1	112	101.8	3.2	66.4	87	76.1	3.5	96.1	110.3	102.9	2.4
T10, °C		96.5	113.4	103.9	2.9	77.5	94.5	84.1	3.1	99.6	112.4	105.2	2.2
T30, °C		103.7	117.2	109.9	2.5	89	115.5	97.5	4.1	106.2	119	111.7	2.2
T50, °C		111.2	124.1	117.2	2.6	95.7	132	107.2	5.4	113.5	128.1	120.5	2.5
T70, °C		119.4	135.1	126.6	3.4	101.8	147.7	116.1	6.6	122.8	142.4	132.6	3.3
T90, °C		130.9	150.5	139.5	4.3	110	171.7	126.6	7.9	136.5	160.6	148.9	3.9
T95, °C		135.8	157.1	144.6	4.7	113.8	184.5	130.7	8.7	141.8	167.6	155.6	4.1
FBP, °C		144.4	172.3	155.3	5.2	124.6	206.3	140.5	10	151.6	183.1	167.5	4.1
API		55.4	61.8	59.3	1.1	56.3	66.7	63.2	1.4	56.2	60.2	58.4	0.8
IBP, °C		25.9	37.5	33	2.1	30.8	43.9	37.5	2.6	22	35.5	29.7	1.6
T5, °C		33.4	46.7	42	2	43.6	62.2	55.7	3.6	30	44.2	36.8	1.8
T10, °C		36	48.4	43.9	1.9	50.8	68.5	62.3	3.2	40	46.1	38.7	2.8
T30, °C		41.1	53.1	48.2	2.1	69.7	86.8	80	3	35.5	50.3	43.8	1.5
T50, °C		45.2	60.9	53.4	2.7	89.8	106.2	98.6	3.3	40.1	55.2	49.1	1.5
T70, °C		49.2	70.4	61	3.4	108.5	126.2	117.3	4.1	45.8	61.3	55.4	1.7
T90, °C		59.8	86.8	73.3	4.5	130	161.7	141.1	6.9	57.1	68.4	63.7	1.2
T95, °C		64.6	97.8	79.5	6	138	178.7	151.4	8.3	62	70.9	65.8	1.2
FBP, °C		67.9	115.7	91.9	7.4	148.6	195.3	167.7	9.8	66.1	80.7	70.8	2.4
API		79.2	86.7	82.7	1.4	63	73.1	67.8	1.2	81.4	89.7	85.8	0.9
V.P., °C		69.5	106.9	80.9	5.5	32.9	66.1	44.3	5.4	82.6	122.9	97.7	5.3

4.3.1. FT-NIR Spectroscopic Analysis Results

Figure 4.23, 4.24 and 4.25 shows raw FT-NIR spectra of İzmit, İzmir and Kırıkkale refinery, respectively, in each figure, spectra on the left side belongs to HSRN samples while on the right side belongs to LSRN samples.

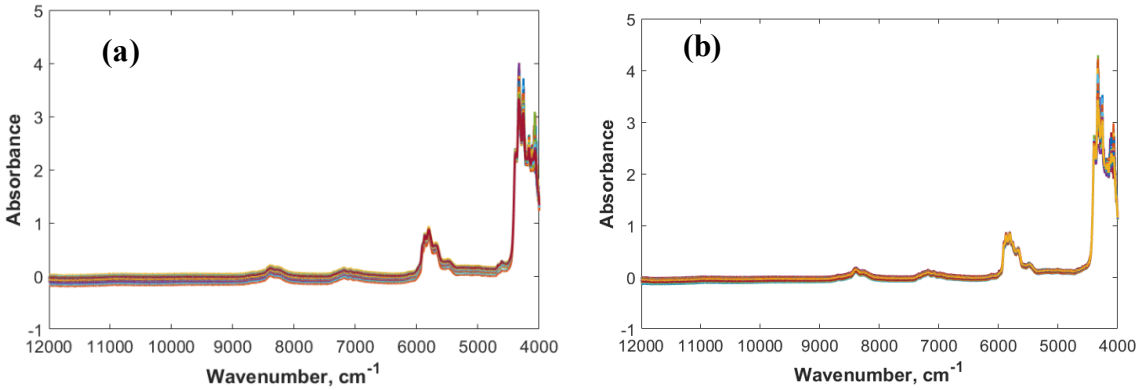

Figure 4.23. FT-NIR spectra HSRN(a) and LSRN(b) samples belongs to İzmit Refinery.

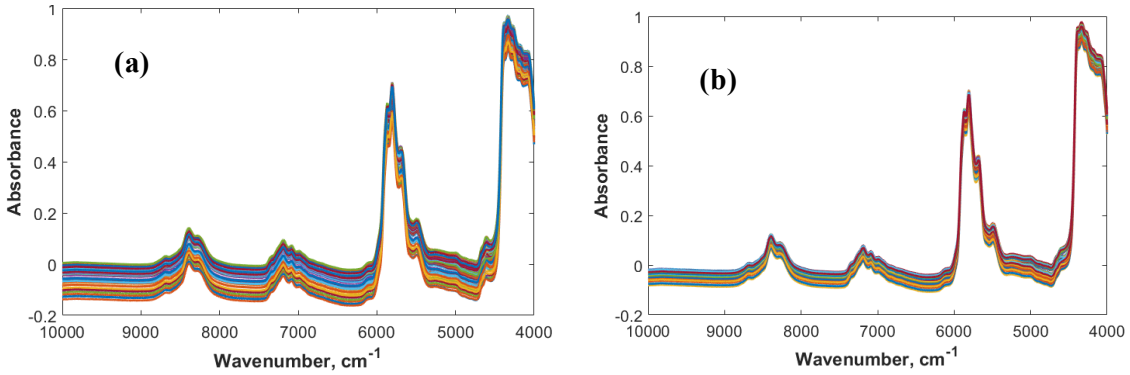

Figure 4.24. FT-NIR spectra HSRN(a) and LSRN(b) samples belongs to İzmir Refinery.

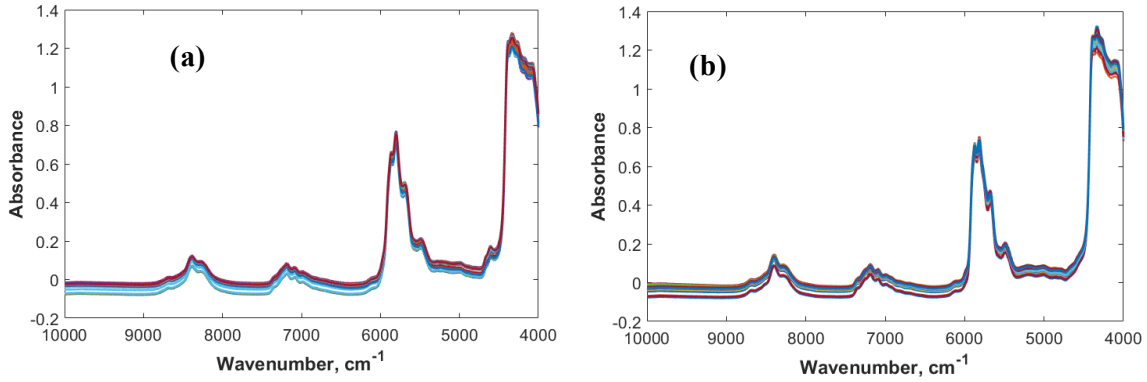

Figure 4.25. FT-NIR spectra HSRN(a) and LSRN(b) samples belongs to İzmir Refinery.

When FT-NIR samples of naphtha samples were investigated, some differences were observed as shown in Figure 4.23, 4.24 and 4.25. Firstly, in Figure 4.24, which belongs to samples in İzmit refinery, although 4500-4000 cm^{-1} spectral range shows significant spectral variations, it contains no useful spectral information due to the strong saturation of NIR radiation. However, in Figure 4.24 and 4.25, no such strong saturation was observed. The main reason of the difference is that two different brands of spectroscopy were used in this study which are Bruker and Perkin Elmer. Due to strong saturation, spectral range of 4500-4000 cm^{-1} were removed from data set belongs to İzmit refinery. Also, 12000-9000 cm^{-1} ranges were also removed from the same data set since no infrared absorption was observed between these spectral ranges.

In addition, in each obtained spectrum, significant baseline shifts were observed. However, effect of baseline shifts was also differing from refinery to refinery. Baseline shift is observed more in refineries, where product composition changes more. That is why, among the sample measurements, the least baseline shift problem is seen in the sample measurements obtained from Kırıkkale refinery. Extended Multiplicative Scatter Correction (EMSC) pre-processing method was applied to each data set to enhance signal properties and suppress unwanted variations before multivariate calibration and shown in Figure 4.26, 4.27 and 4.28.

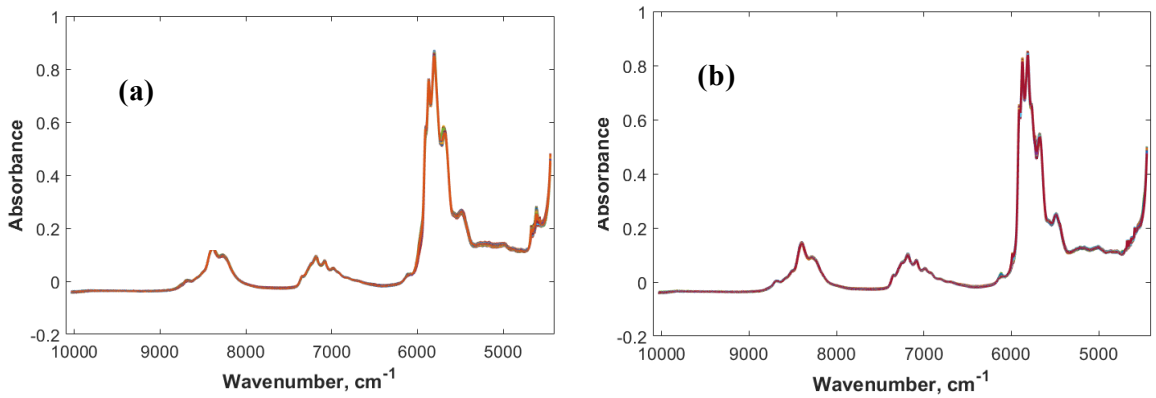

Figure 4.26. EMSC corrected FT-NIR spectra HSRN(a) and LSRN(b) samples belongs to İzmit Refinery

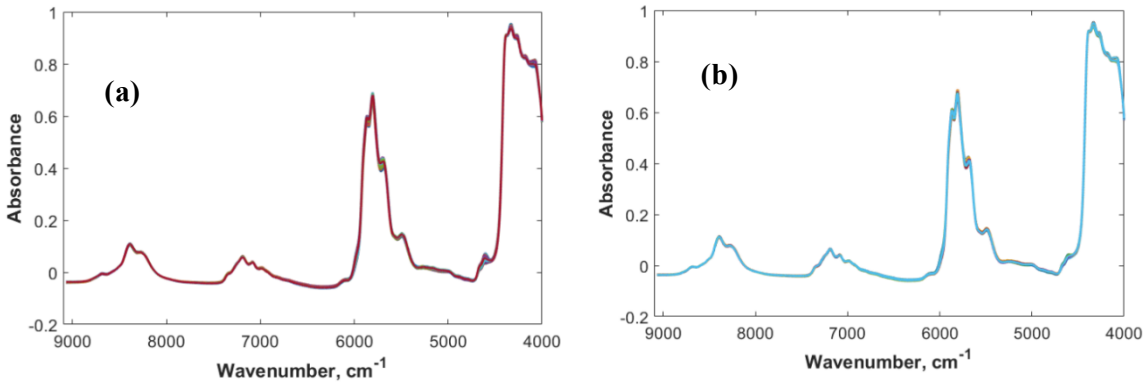

Figure 4.27. EMSC corrected FT-NIR spectra HSRN(a) and LSRN(b) samples belongs to İzmir Refinery

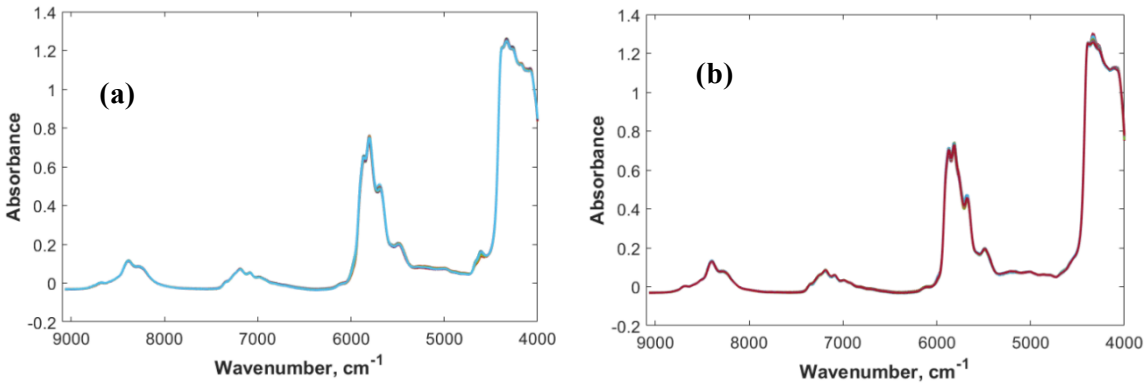

Figure 4.28. EMSC corrected FT-NIR spectra HSRN(a) and LSRN(b) samples belongs to Kırıkkale Refinery

To EMSC corrected spectral data shown in Figure 4.26, 4.27 and 4.28 two different calibration approached were performed which are Genetic Inverse Least squares (GILS) and Partial Least Squares (PLS) Regression.

From 301 HSRN samples belongs to İzmit Refinery, a total of 230 samples were assigned as calibration set to develop model and the rest of 71 samples were used as independent validation set to observed prediction ability of the model. From 276 LSRN samples belongs to İzmit Refinery, a total of 210 samples were assigned as calibration set to develop model and the rest of 66 samples were used as independent validation set to observed prediction ability of the model

From 281 HSRN samples belongs to İzmir Refinery, a total of 210 samples were assigned as calibration set to develop model and the rest of 71 samples were used as independent validation set to observed prediction ability of the model. From 256 LSRN samples, a total of 192 samples were assigned as calibration set to develop model and the rest of 64 samples were used as independent validation set to observed prediction ability of the model

From 231 HSRN samples belongs to Kırıkkale Refinery, a total of 175 samples were assigned as calibration set to develop model and the rest of 56 samples were used as independent validation set to observed prediction ability of the model. From 232 LSRN samples a total of 175 samples were assigned as calibration set to develop model and the rest of 57 samples were used as independent validation set to observed prediction ability of the model.

After developing PLS and GILS models for each data set, standard error of cross-validation (SECV) and standard error of prediction (SEP) values of each parameter were calculated and compared. For each parameter, best multivariate calibration method was chosen according to lowest SECV and SEP values. Calculated standard error of calibration and standard error of validation parameters along with name of the standard method and reproducibility values for final selected models for EMSC pre-processed FT-NIR spectra for HSRN and LSRN samples of each refinery is given in Table 4.18. Calculated reproducibility values for straight run naphtha based on standard method are given in Table 4.19. Letter i in Table 4.19 represents i^{th} sample for calculation. In Table 4.18, parameters those with LV means PLS models were selected.

Table 4.18. Multivariate Calibration results along with name of the standard method and number of latent variables.

Parameters		SECV	SEP	R^2	LV		SECV	SEP	R^2	LV		SECV	SEP	R^2	LV
IBP, °C	**İzmit HSRN**	0.966	1.482	0.962	-	**İzmir HSRN**	1.834	2.688	0.711	-	**Kırıkkale HSRN**	0.708	1.295	0.959	12
T5, °C		0.553	0.787	0.972	-		1.204	1.658	0.868	-		0.405	0.799	0.977	-
T10, °C		0.462	0.625	0.976	-		0.571	0.832	0.964	-		0.268	0.688	0.988	-
T30, °C		0.315	0.507	0.985	-		0.292	0.461	0.995	-		0.278	0.531	0.986	-
T50, °C		0.268	0.511	0.990	-		0.276	0.467	0.997	-		0.219	0.480	0.993	-
T70, °C		0.330	0.808	0.991	-		0.361	0.538	0.997	-		0.207	0.519	0.996	-
T90, °C		0.457	0.967	0.989	-		0.428	0.741	0.997	-		0.387	0.637	0.990	-
T95, °C		0.605	1.293	0.984	-		0.550	0.910	0.995	-		0.573	0.785	0.979	-
FBP, °C		1.134	1.968	0.951	-		1.764	3.031	0.949	-		1.258	1.507	0.895	-
API		0.075	0.135	0.996	-		0.146	0.312	0.988	-		0.082	0.131	0.988	-
IBP, °C	**İzmit LSRN**	0.642	0.810	0.905	11	**İzmir LSRN**	0.660	1.051	0.938	-	**Kırıkkale LSRN**	0.791	0.942	0.722	-
T5, °C		0.293	0.458	0.980	-		0.788	1.084	0.945	11		0.263	0.468	0.977	-
T10, °C		0.256	0.368	0.983	-		0.591	0.730	0.967	11		0.173	0.336	0.988	-
T30, °C		0.230	0.353	0.988	-		0.383	0.686	0.985	-		0.122	0.252	0.994	-
T50, °C		0.239	0.391	0.993	-		0.341	0.665	0.990	-		0.158	0.397	0.992	-
T70, °C		0.262	0.422	0.995	-		0.302	0.643	0.995	-		0.146	0.390	0.993	-
T90, °C		0.485	0.903	0.989	-		0.773	1.463	0.989	-		0.158	0.326	0.983	-
T95, °C		0.721	1.261	0.986	-		0.955	1.950	0.989	-		0.211	0.411	0.972	-
FBP, °C		0.821	2.176	0.987	17		2.451	3.519	0.933	-		1.117	1.661	0.761	-
API		0.124	0.185	0.992	-		0.060	0.183	0.997	-		0.104	0.212	0.989	-
V.P., °C		0.570	0.896	0.990	-		1.036	1.834	0.964	-		0.972	1.448	0.967	-

Table 4.19. Reproducibility value calculation for naphtha calculations based on standard method.

Parameters	Straight Run Naphtha
	Reproducibility, °C
IBP	4.7
T5	2.5+2.8(0.43((**T10i - IBPi**) /10) +0.24)
T10	1.9+2.8(0.43((**T20i - T5i**)/15) +0.24)
T20	1.9+2.8 (0.43 ((**T30i - T10i**)/15) +0.24)
T30	1.8+2.8(0.43((**T40i- T20i**)/20) +0.24)
T40	1.8+2.8(0.43((**T50i- T30i**)/20) +0.24)
T50	1.9+2.8(0.43((**T60i- T40i**)/20) +0.24)
T60	1.9+2.8(0.43((**T70i- T50i**)/20) +0.24)
T70	2.1+2.8(0.43((**T80i- T60i**)/20) +0.24)
T80	2.1+2.8(0.43((**T90i- T70i**)/20) +0.24)
T90	2.8+2.8(0.43((**T95i- T80i**)/15) +0.24)
T95	3.6+2.8(0.43((**FBPi - T90i**)/8) +0.24)
FBP	7.1
API	0.5
Vapor P.	0.01014(**Vapor P.i** +160)

As shown in Table 4.18, for most of the parameters, GILS models have been found as the best model. When three refineries were compared, for almost all parameters, lowest error values were found in samples which belongs to İzmit refinery. This can be explained by the sample collection times. Measurements in İzmit refinery have been performed daily for three years, which is a year longer than other refineries. Since extending sampling time results in more sample variation in data set, more robust models are obtained in İzmit refinery samples than other refineries. Highest SEP values are observed in IBP and FBP. This can be explained by reproducibility values of standard methods. Reproducibility values of IBP and FBP are higher than other parameters hence these models give the highest standard errors.

After developing multivariate calibration models for HSRN and LSRN samples for each refinery, models were uploaded to each laboratory and performance of model predictions has started to be observed. Predictions were made with new samples produced in refineries daily. Developed model performances with blind unknown samples are presented in the graphs below. For each sample group, two different graphs are presented. Graphs on the left side represent the laboratory results for given parameters. Orange

points represent reference values obtained from standard method while blue points represent model predictions. Graphs on the right side shows the difference between actual and model prediction values. Red lines represent the reproducibility value of standard method. Model performance of given parameters have been evaluated with reproducibility value of standard methods.

Considering the multitude of graphics, the results are presented under the headings in order to avoid confusion.

4.3.1.1. İzmit Refinery – Sample Predictions

In İzmit refinery, predictions of developed multivariate calibration models for HSRN and LSRN samples had been observed for 87 days and results are shown in Figure 4.29 for HSRN samples and in Figure 4.30 for LSRN samples.

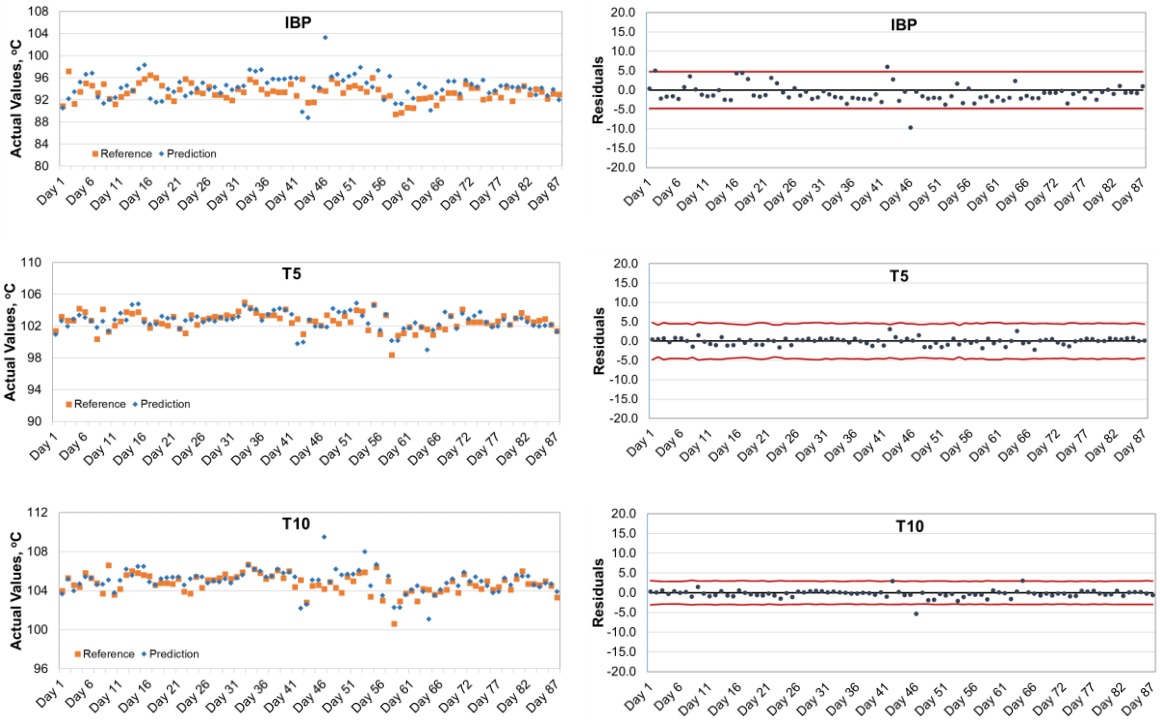

Figure 4.29. Model predictions of HSRN samples belongs to İzmit Refinery.

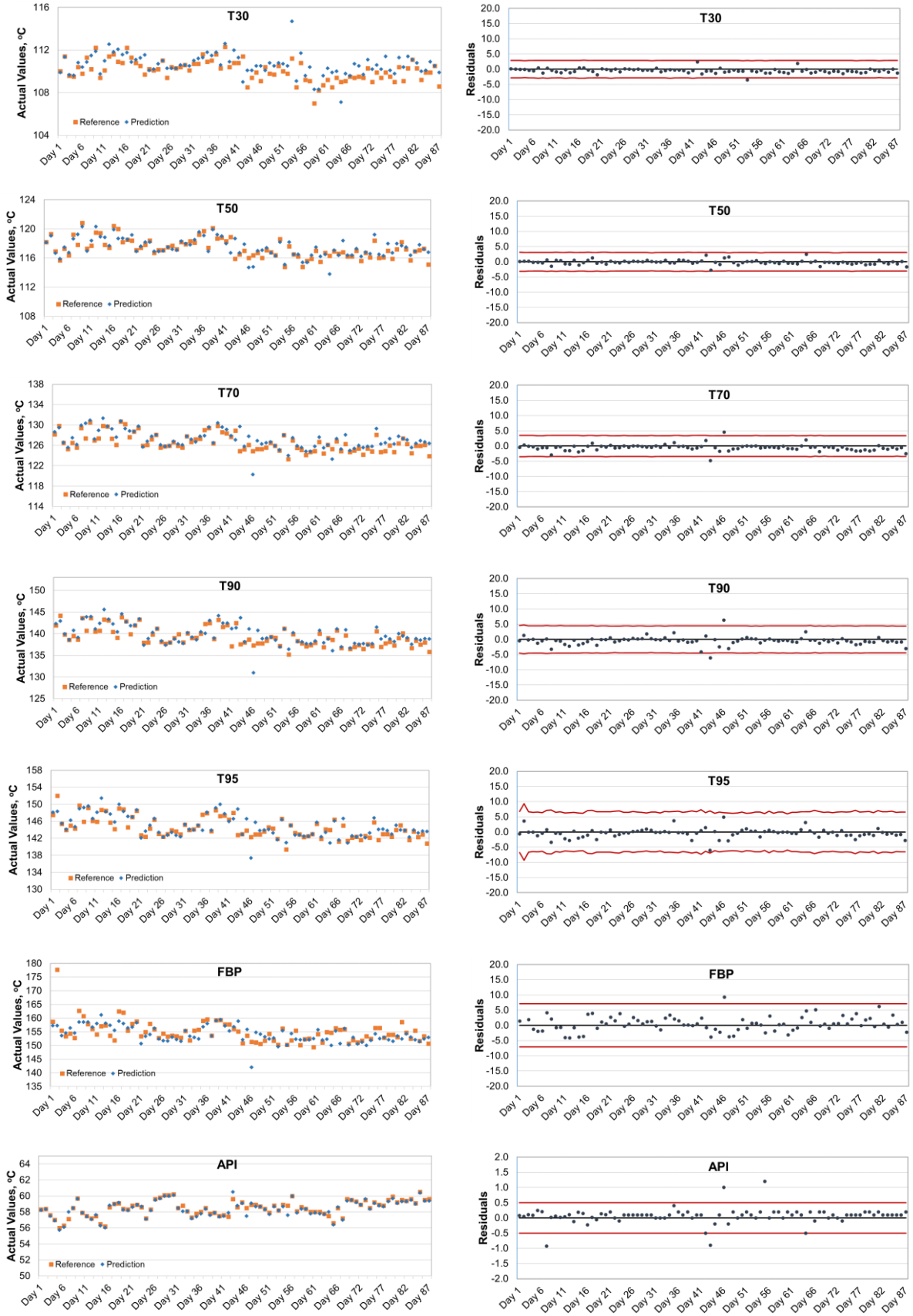

Figure 4.29. Model predictions of HSRN samples belongs to İzmit Refinery (cont'd).

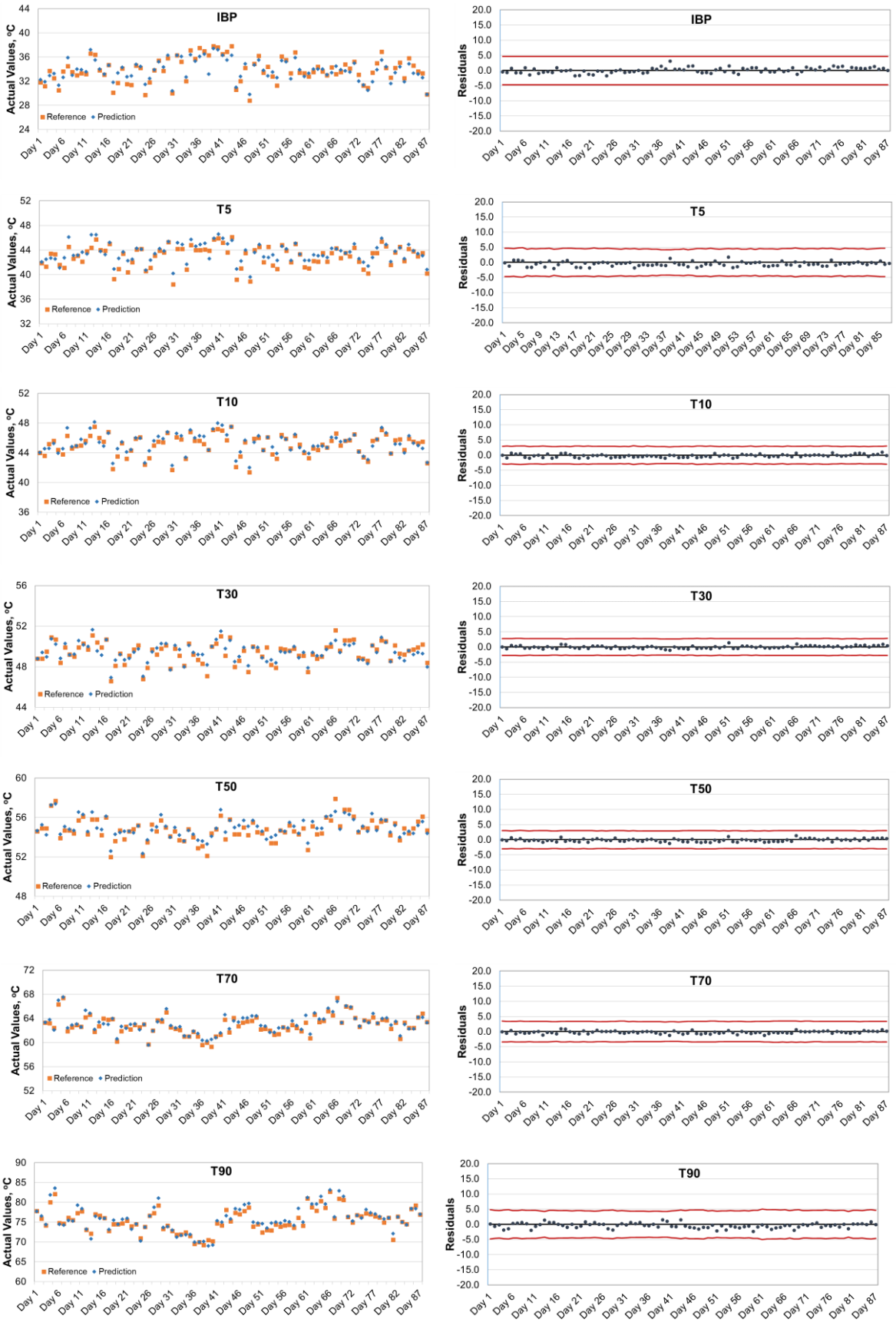

Figure 4.30. Model predictions of LSRN samples belongs to İzmit Refinery.

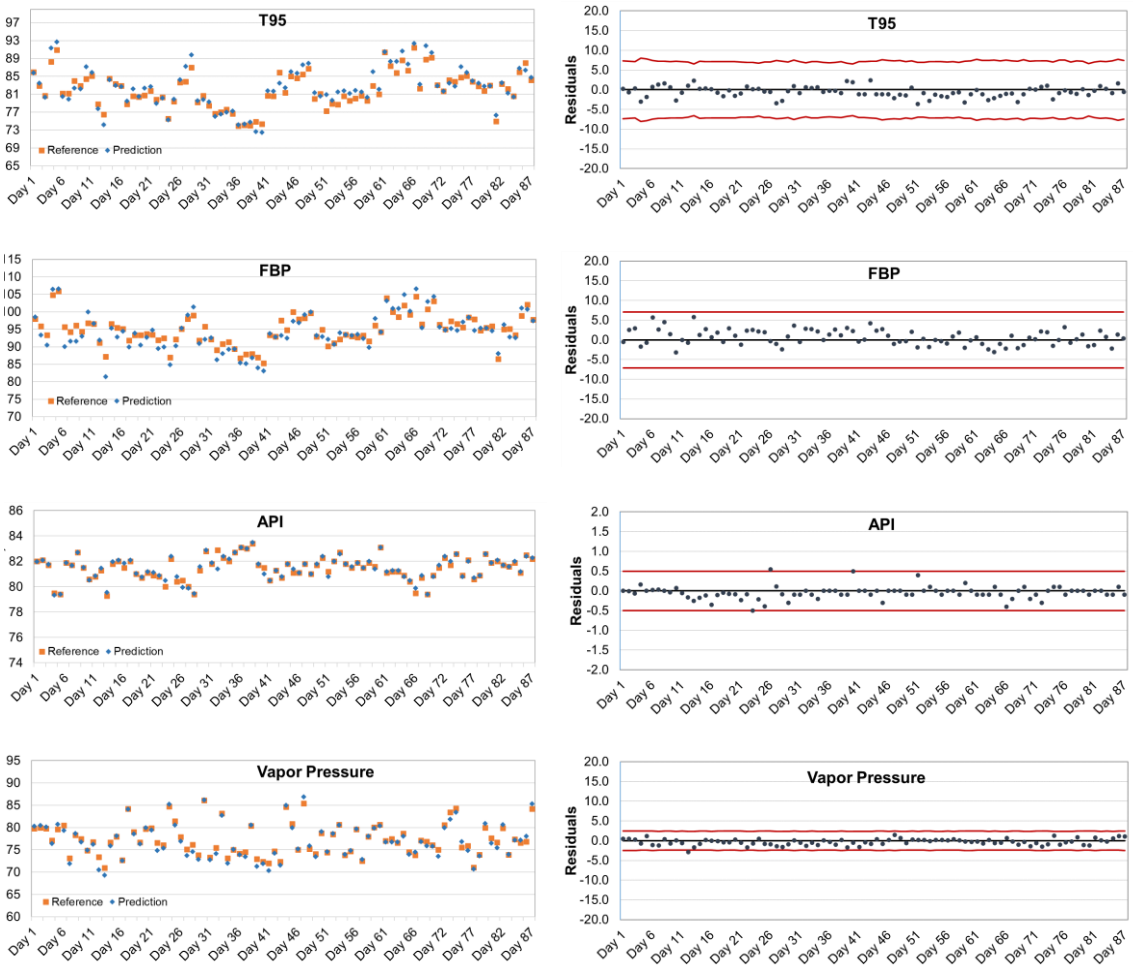

Figure 4.30. Model predictions of LSRN samples belonging to İzmit Refinery (cont'd).

In HSRN sample predictions, except for a few samples, it was observed that the difference between the laboratory result and the model estimates never exceeded the laboratory reproducibility value, which is shown as a red line on the left side. When the results are examined closely, it was observed that high erroneous results were obtained in distillation temperature and API predictions on the same days which are day 43 and day 46. and unusual spectra, compared to other samples, was observed for both days wh ich indicated that a human error was made during the spectroscopic measurements.

When all parameters were examined in LSRN sample predictions, successful predictions have been obtained for all 87 days.

4.3.1.2. İzmir Refinery –Sample Predictions

In İzmir refinery, predictions of developed multivariate calibration models for HSRN and LSRN samples had been observed for 76 days and results are shown in Figure 4.31 for HSRN samples and in Figure 4.32 for LSRN samples.

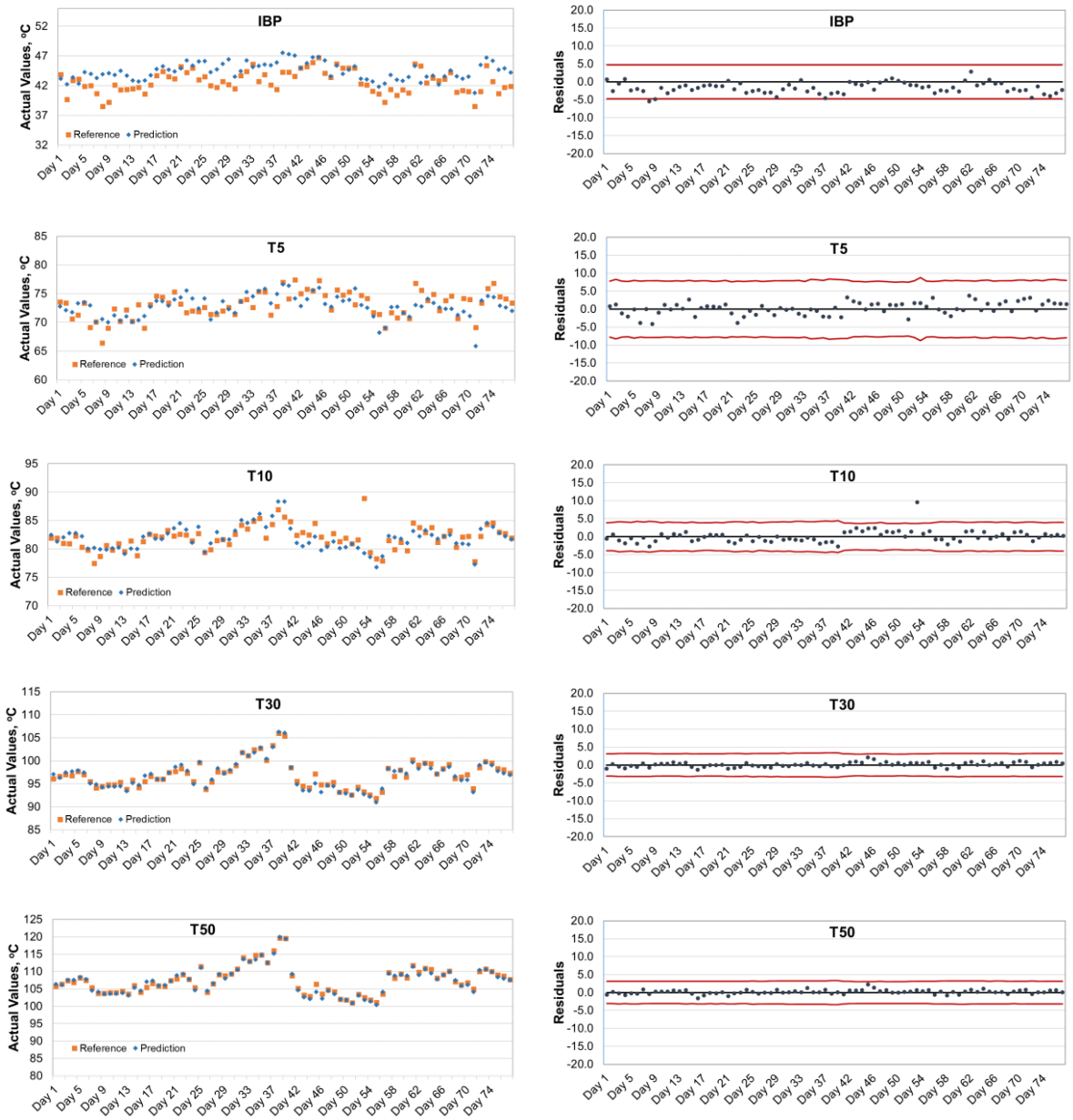

Figure 4.31. Model predictions of HSRN samples belongs to İzmir Refinery.

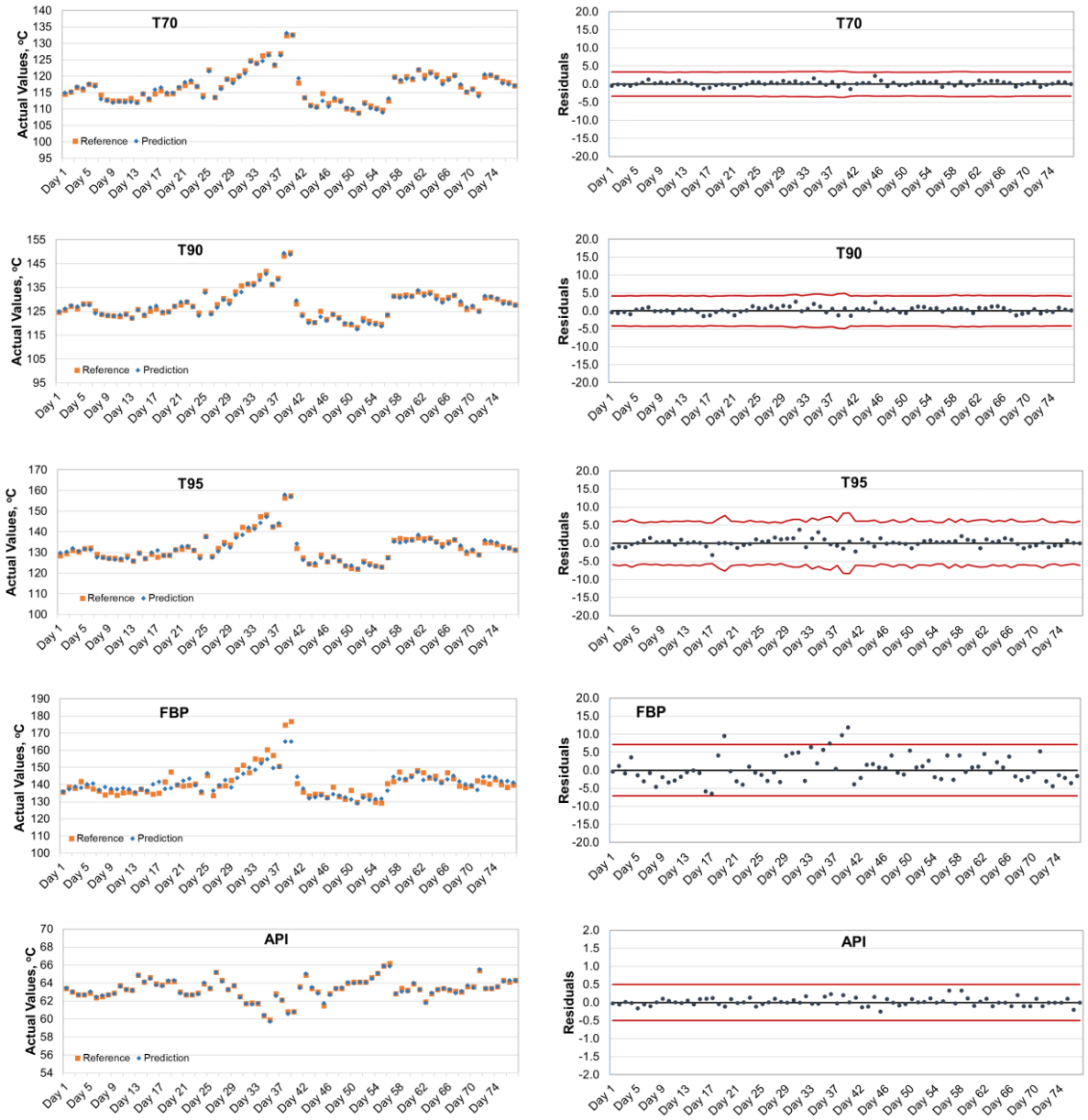

Figure 4.31. Model predictions of HSRN samples belongs to İzmir Refinery (cont'd).

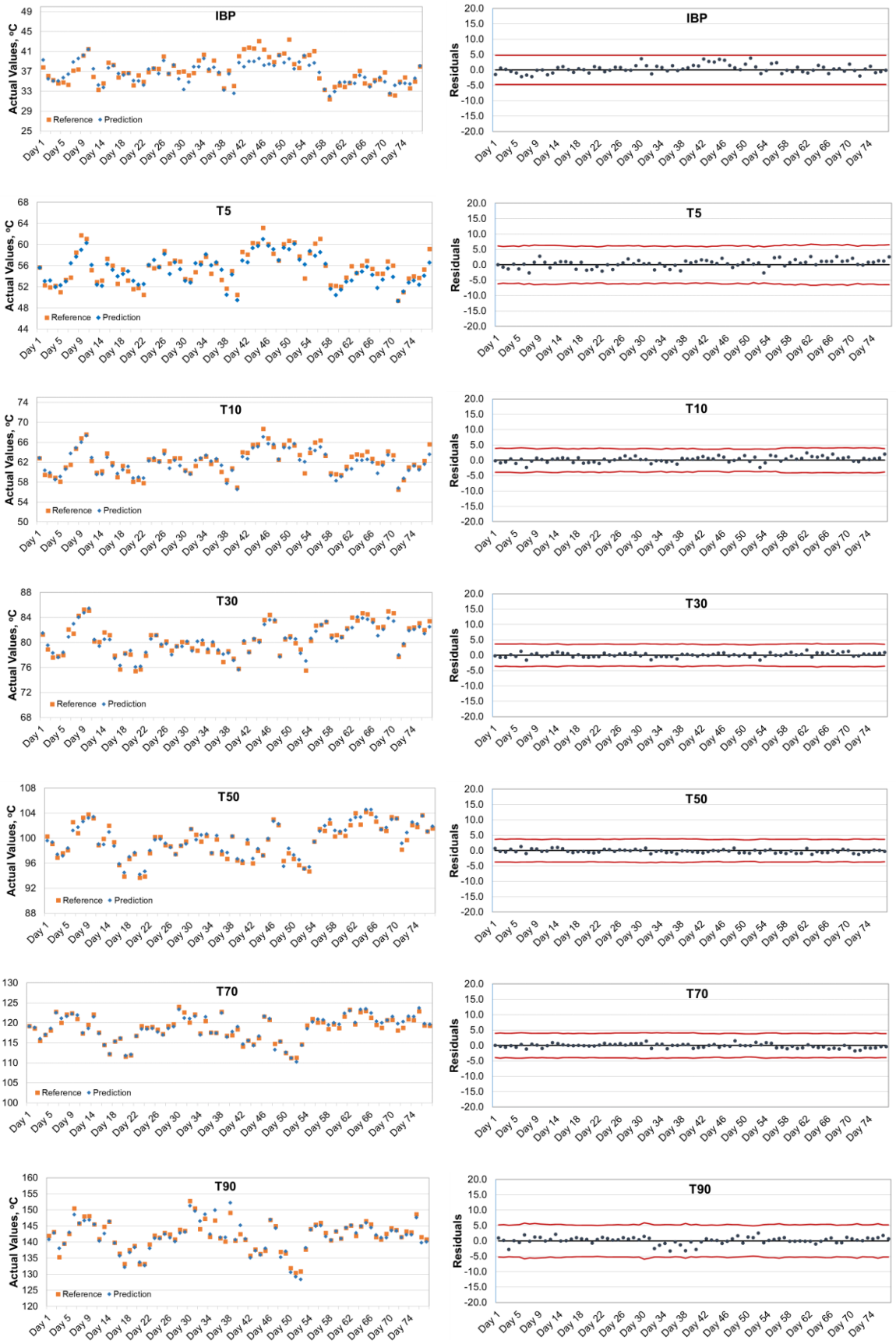

Figure 4.32. Model predictions of LSRN samples belongs to İzmir Refinery.

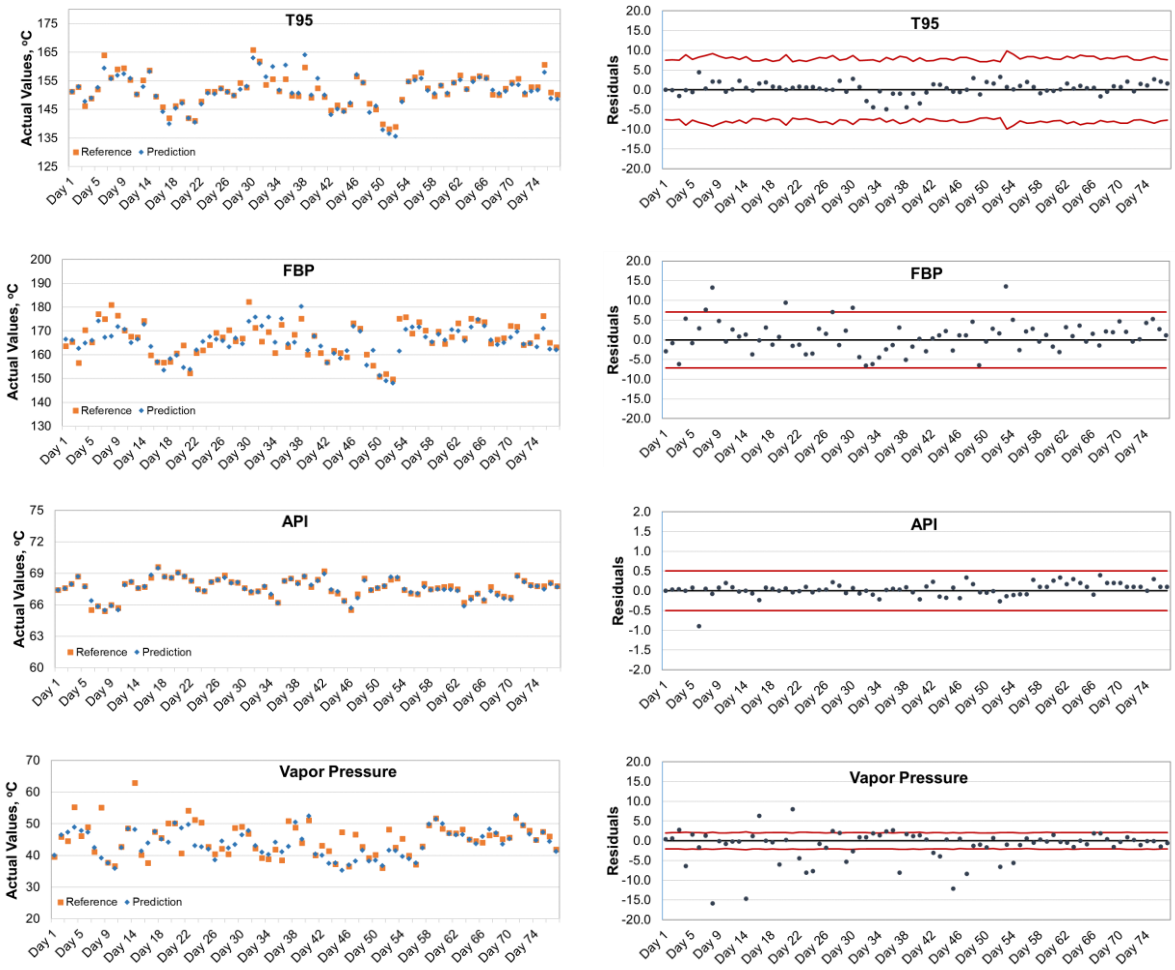

Figure 4.32. Model predictions of LSRN samples belongs to İzmir Refinery (cont'd).

In HSRN sample predictions, except for final boiling point predictions, successful predictions were obtained when difference between reference analysis and model predictions compared to reproducibility value. However, in initial boiling point, it has been observed that systematic errors are obtained, and the results always give positive or negative errors. The model is not completely reliable because the errors do not come in a normal distribution. When final boiling point predictions were observed, high error in some days was observed. Considering high SEP values shown in Table 4.18 for İzmir HSRN multivariate calibration results, low prediction ability for this parameter can be explained.

Similar to HSRN samples, predictions of FBP of LSRN samples also have higher residuals than reproducibility values. Until day 54, reference values and predictions of vapor pressure are not always compatible with each other, and their difference is obtained

more than the reproducibility value. This situation had been studied closely. First, sample spectra were examined, and no significant difference was observed. It was expected since if the wrong spectrum was taken that day, we should have seen high error values on other days as well. Another possibility was that dynamic range of developed models did not cover this range. When max and min values of vapor pressure in Table 4.18it can be seen that reference values were obtained between these values. The other possibility is that unsuccessful multivariate calibration models have been installed. The models were revised by adding samples with high deviation, but these values still gave a high cross validation error in the model. Lastly, laboratory reference analysis was check. Before performing the vapor pressure analysis, the sample bottle needs to be cooled in cooling cabinet for a few hours. It was noticed that on some days the analysis took place without cooling the sample sufficiently. Starting from day 54, before performing laboratory analysis, it was ensured that the sample bottle was cold and as can be seen in the last graph in Figure 4.32, successful predictions had been obtained since then.

4.3.1.3. Kırıkkale Refinery –Sample Predictions

In Kırıkkale refinery, predictions of developed multivariate calibration models for HSRN and LSRN samples had been observed for 60 days and results are shown in Figure 4.33 for HSRN samples and in Figure 4.34 for LSRN samples.

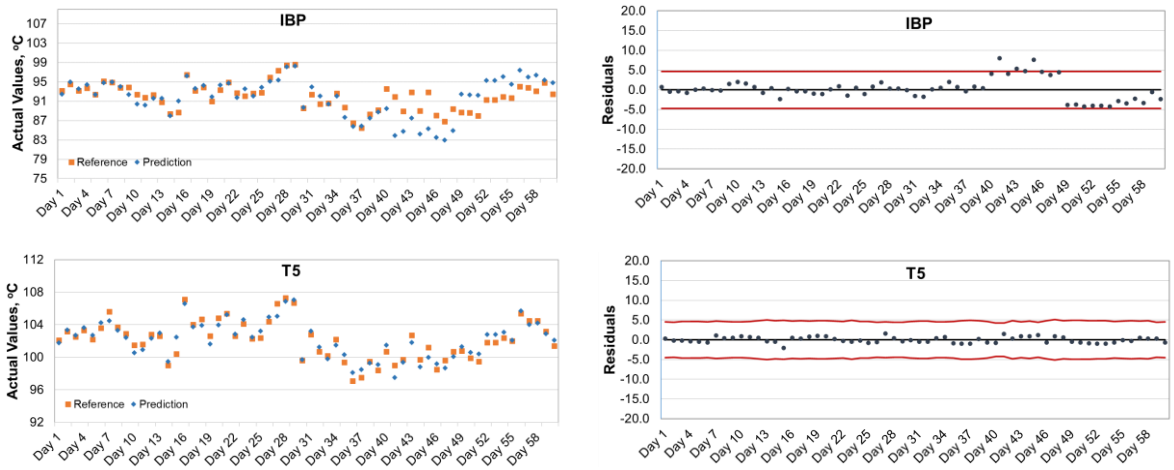

Figure 4.33. Model predictions of HSRN samples belongs to Kırıkkale Refinery.

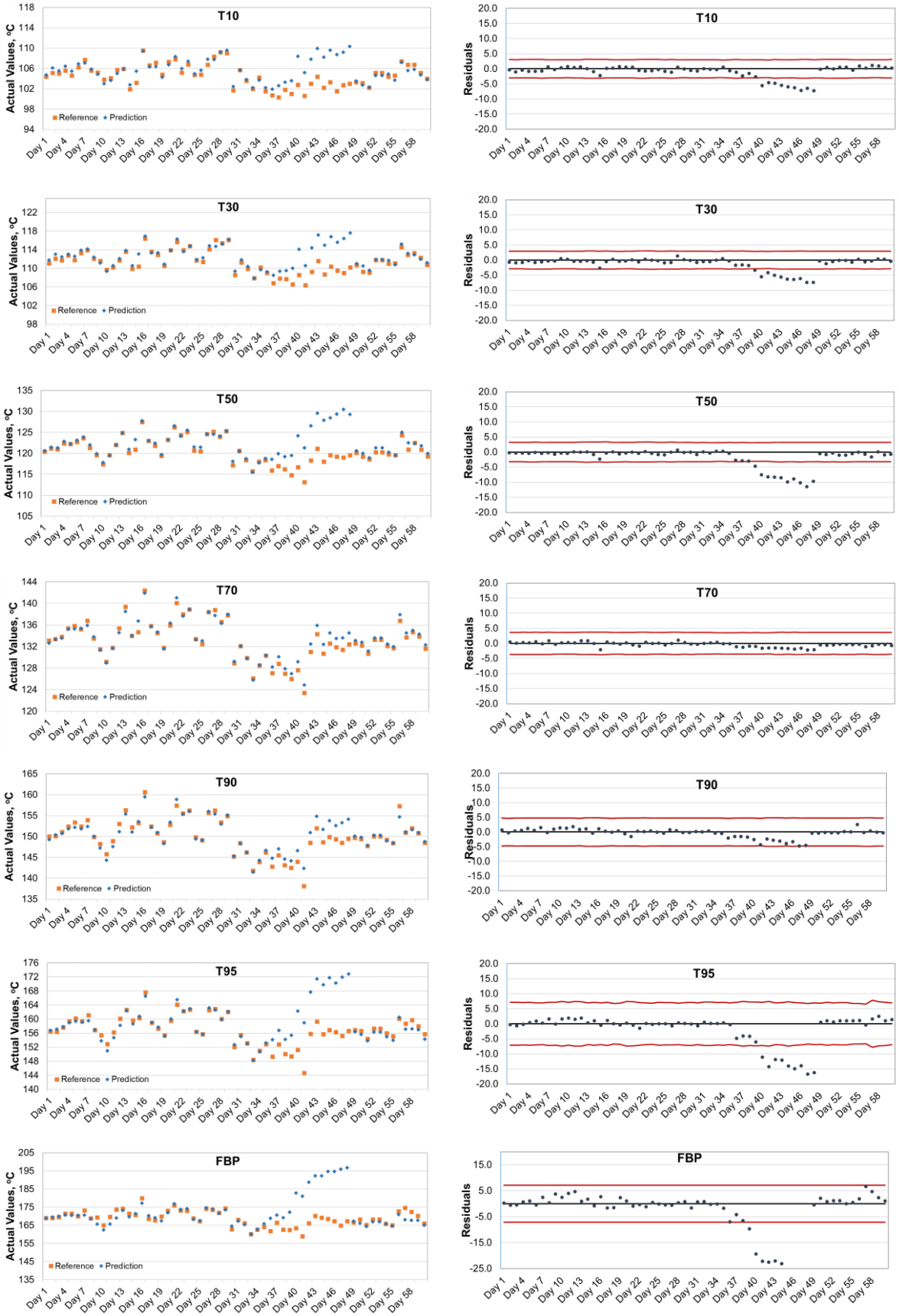

Figure 4.33. Model predictions of HSRN samples belongs to Kırıkkale Refinery (cont'd).

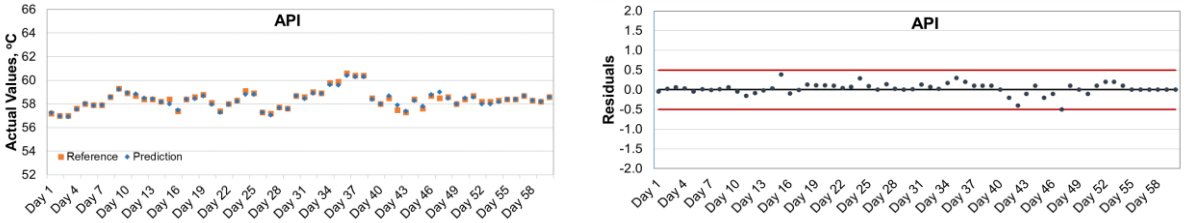

Figure 4.33. Model predictions of HSRN samples belongs to Kırıkkale Refinery (cont'd)

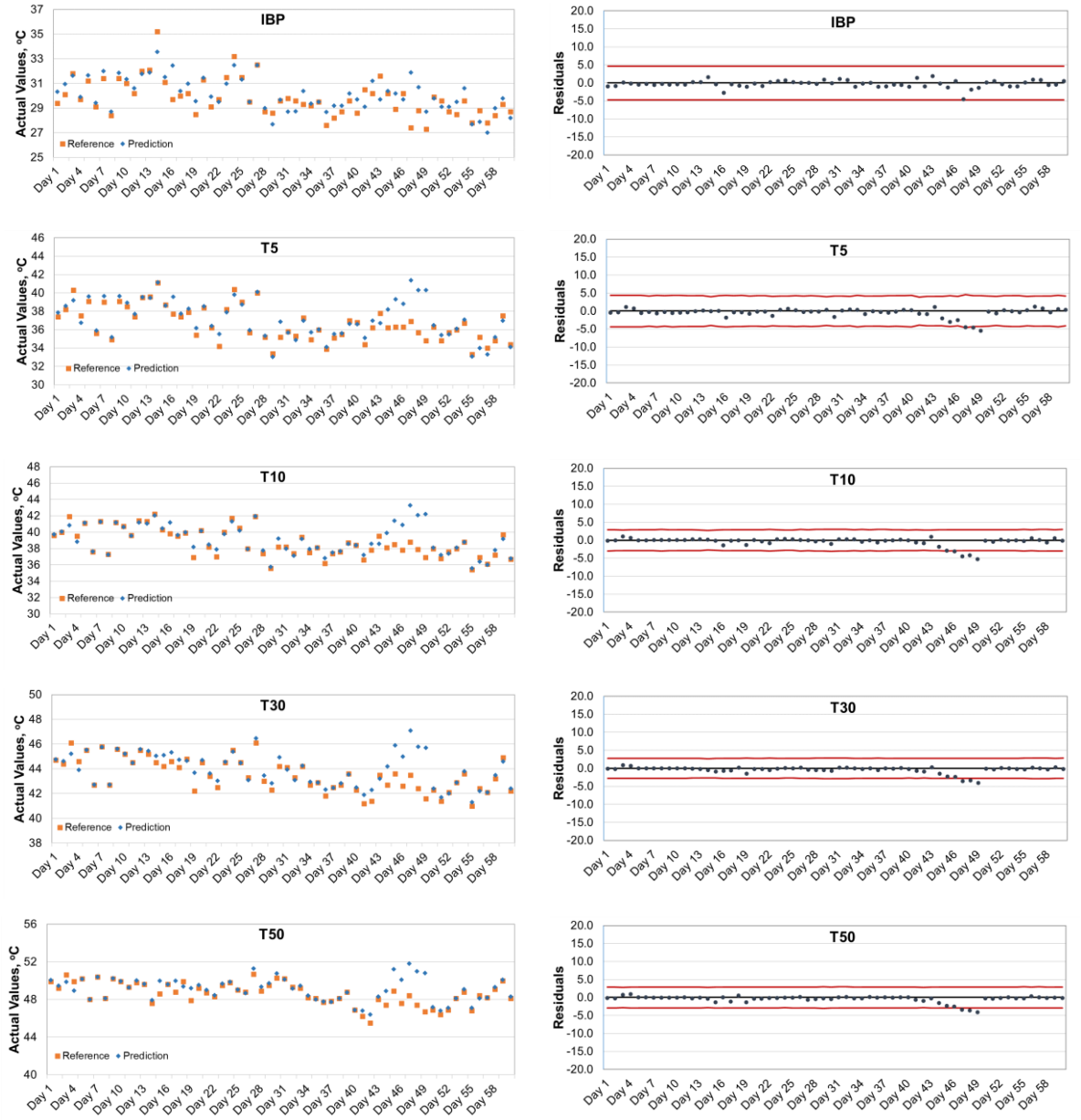

Figure 4.34. Model predictions of LSRN samples belongs to Kırıkkale Refinery.

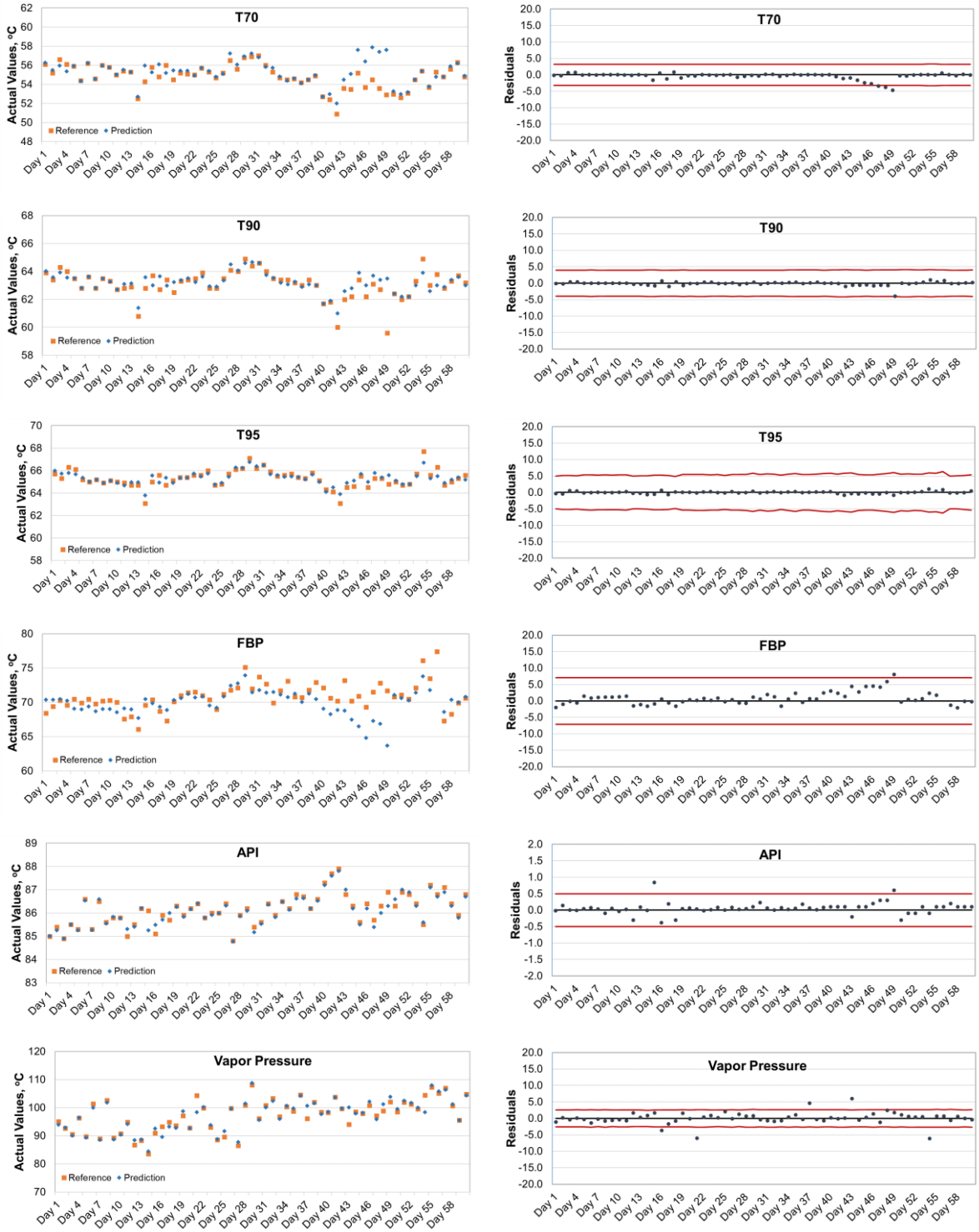

Figure 4.34. Model predictions of LSRN samples belongs to Kırıkkale Refinery (cont'd)

In HSRN samples, successful predictions had been obtained until day 36. Except for T5 and API, residuals between reference values and predictions were started to increase day by day. While the IBP model started to give result in lower values, the other

parameter model results were systematically reported always higher than the reference values. When the same trend is observed in LSRN samples, it was assumed that the product composition has changed due to different crude oil is started to be processed. However, planning instructions were reviewed, and it was found that no new crude oil was processed at the refinery. Sample spectra was also observed and PCA analysis was performed. It was observed that sample spectra of those days were grouped in a different place on the PCA score graph. This proved that these sample spectra were compositionally different. Reason of compositional change was investigated, and it was realized that the solvent, which is used in cell cleaning between sample measurements, has run out and a new one has not been placed. It is also known that, beside naphtha samples, also spectra of kerosene samples are collected in the same FT-NIR spectroscopy. Since the cell cannot be cleaned sufficiently, the kerosene sample remains inside, and the model results report high values because the kerosene sample is a heavier product than naphtha. when the solvent bottle is replaced with a new one, successful predictions began to be obtained again.

As a result of successful model predictions, laboratory reference analyzes were reduced in each refinery and reporting with model results began. Physical properties of HSRN and LSRN samples which are produced daily in crude distillation units are now reported using developed multivariate calibration models.

CHAPTER 5

CONCLUSION

In this thesis, a new methodology to determine physical and chemical properties of petroleum products obtained from actual refining process crude distillation units using different spectroscopic methods along with chemometrics were developed.

Three different spectral analyses were conducted and developed multivariate calibration models were compared for heavy and light diesel samples, separately. Having a more complex chemical composition resulted in successful calibration models that have high predictive ability for most of the quality parameters of UV-Vis spectroscopy for heavy diesel samples. However, when same comparison was performed with light diesel samples, vibrational spectroscopy, FT-NIR and FTIR-ATR, was found to have low prediction errors. In both cases, GILS algorithm showed better predictive ability compared to PLS algorithm for low distillation temperatures and FT-NIR spectral data were selected to develop API gravity.

Fluorescence spectroscopy had been used in two different modes for measurement of light diesel samples obtained from crude distillation unit. It was observed that fluorescence intensity of samples with lower API grade show high fluorescence intensity in both measurement mode. Since there is no information about chemical composition of samples, it can be only assumed that samples with low API value have more fluorescent aromatic compounds. Synchronous fluorescence spectra showed more sharped and detailed information of light diesel samples. Multivariate calibration results were both compared as well as fluorescence measurement modes. PLS multivariate calibration results for both spectral analysis modes have resulted in lower standard error of prediction for most of the parameters. It was also concluded that better calibration models obtained from synchronous fluorescence spectra.

Studies that have been conducted with heavy straight run naphtha (HSRN) and light straight run naphtha (LSRN) samples obtained from three different crude distillation units belonging to three different refineries which are İzmit Refinery, İzmir Refinery and Kırıkkale Refinery have been also presented. Spectroscopic measurements and laboratory

analysis were carried out in refineries where naphtha products are produced. A total of six different data sets were obtained from two different naphtha and 3 different refineries. Two different multivariate calibration approached were applied to EMSC corrected FT-NIR spectra of six different data set and best model for each parameter was chosen according to lower SECV and SEP values. Each developed model was used to predict features of new samples produced in refineries. These predictions, made with blind unknown samples, were followed daily at each refinery and prediction performance of multivariate calibration models were observed. In every refinery, successful model predictions were obtained for HSRN and LSRN samples.

The successful results obtained from the developed models were found suitable for use as an alternative method to routine analyzes in Tüpraş refinery laboratories. While reporting with model results, it is important to monitor whether there is a significant change in the type of crude oil processed in the crude oil distillation unit, the dynamic ranges of the parameters and the operational conditions. Important operational changes may cause the model prediction success to decrease, the models should be revised and the change in the product should be introduced to the new models.

REFERENCES

(1) International Organization for Standardization. Petroleum Products - Determination of Distillation Characteristics at Atmospheric Pressure (ISO 3405:2011). **2011**.

(2) ASTM Standard D2887-04 "Standard Test Method for Boiling Range Distribution of Petroleum Fractions by Gas. *ASTM International 02.04*.

(3) International Organization for Standardization. Crude Petroleum and Liquid Petroleum Products — Laboratory Determination of Density — Hydrometer Method (ISO 3675:1998). **2022**.

(4) International Organization for Standardization. Determination of Flash Point - Pensky-Martens Closed Cup Method (ISO 2719:2016). *International Organization for Standardization*. **2016**.

(5) International Organization for Standardization. IP 394: Liquid Petroleum Products - Vapour Pressure - Part 1: Determination of Air Saturated Vapour Pressure (ASVP) and Calculated Dry Vapour Pressure Equivalent (DVPE). **2019**.

(6) International Organization for Standardization. IP 435: Determination of the Freezing Point of Aviation Turbine Fuels by the Automatic Phase Transition Method. **2016**.

(7) Chung, H. Applications of Near-Infrared Spectroscopy in Refineries and Important Issues to Address. *Appl Spectrosc Rev* **2007**, *42* (3), 251–285. https://doi.org/10.1080/05704920701293778.

(8) Kim, M.; Lee, Y.-H.; Han, C. Real-Time Classification of Petroleum Products Using near-Infrared Spectra. *Comput Chem Eng* **2000**, *24*, 513–517.

(9) Kelly, J. J.; Callis, J. B. Nondestructive Analytical Procedure for Simultaneous Estimation of the Major Classes of Hydrocarbon Constituents of Finished Gasolines. *Anal Chem* **1990**, *62* (14), 1444–1451. https://doi.org/10.1021/ac00213a019.

(10) Fodor, G. E.; Kohl, K. B.; Mason, R. L. Analysis of Gasolines by FT-IR Spectroscopy. *Anal Chem* **1996**, *68* (1), 23–30. https://doi.org/10.1021/ac9507294.

(11) Al-Ghouti, M. A.; Al-Degs, Y. S.; Amer, M. Determination of Motor Gasoline Adulteration Using FTIR Spectroscopy and Multivariate Calibration. *Talanta* **2008**, *76* (5), 1105–1112. https://doi.org/10.1016/j.talanta.2008.05.024.

(12) Özdemir, D. Determination of Octane Number of Gasoline Using near Infrared Spectroscopy and Genetic Multivariate Calibration Methods. *Pet Sci Technol* **2005**, *23* (9–10), 1139–1152. https://doi.org/10.1081/LFT-200035547.

(13) Balabin, R. M.; Safieva, R. Z.; Lomakina, E. I. Gasoline Classification Using near Infrared (NIR) Spectroscopy Data: Comparison of Multivariate Techniques. *Anal Chim Acta* **2010**, *671* (1–2), 27–35. https://doi.org/10.1016/j.aca.2010.05.013.

(14) Lysaght, M. J.; Kelly, J. J.; Callis, J. B. Rapid Spectroscopic Determination of per Cent Aromatics, per Cent Saturates and Freezing Point of JP-4 Aviation Fuel. *Fuel* **1993**, *72* (5), 623–631. https://doi.org/10.1016/0016-2361(93)90574-L.

(15) Westbrook, S. R. Army Use of Near-Infrared Spectroscopy to Estimate Selected Properties of Compression Ignition Fuels. *SAE Technical Paper Series* **2010**, *1*. https://doi.org/10.4271/930734.

(16) Chung, H.; Ku, M. S.; Lee, J. S. Comparison of Near-Infrared and Mid-Infrared Spectroscopy for the Determination of Distillation Property of Kerosene. *Vib Spectrosc* **1999**, *20* (2), 155–163. https://doi.org/10.1016/S0924-2031(99)00034-X.

(17) Ku, M.-S.; Chung, H.; Lee, J. S. Rapid Compositional Analysis of Naphtha by Near Infrared Spectroscopy. *Bull Korean Chem Soc* **1998**, *19* (11), 1189–1193.

(18) Ku, M.-S.; Chung, H. Comparison of Near-Infrared and Raman Spectroscopy for the Determination of Chemical and Physical Properties of Naphtha. *Appl Spectrosc* **1999**, *52* (5), 557–564. https://doi.org/10.1016/j.aca.2008.10.057.

(19) Breitkreitz, M. C.; Raimundo, I. M.; Rohwedder, J. J. R.; Pasquini, C.; Dantas Filho, H. A.; José, G. E.; Araújo, M. C. U. Determination of Total Sulfur in Diesel Fuel Employing NIR Spectroscopy and Multivariate Calibration. *Analyst* **2003**, *128* (9), 1204–1207. https://doi.org/10.1039/b305265f.

(20) Marinović, S.; Krištović, M.; Špehar, B.; Rukavina, V.; Jukić, A. Prediction of Diesel Fuel Properties by Vibrational Spectroscopy Using Multivariate Analysis. *Journal of Analytical Chemistry* **2012**, *67* (12), 939–949. https://doi.org/10.1134/s1061934812120039.

(21) Özdemir, D. Near Infrared Spectroscopic Determination of Diesel Fuel Parameters Using Genetic Multivariate Calibration. *Pet Sci Technol* **2008**, *26* (1), 101–113. https://doi.org/10.1080/10916460600705824.

(22) de Oliveira, I. K.; de Carvalho Rocha, W. F.; Poppi, R. J. Application of near Infrared Spectroscopy and Multivariate Control Charts for Monitoring Biodiesel Blends. *Anal Chim Acta* **2009**, *642* (1–2), 217–221. https://doi.org/10.1016/j.aca.2008.11.003.

(23) Guerrero, A.; Anguebes, F.; Castelán, M.; Morales, V.; García, R.; Córdova, A. V.; Zavala, J. C. FTIR-ATR and Multivariate Calibration for the Prediction of Biodiesel Concentration in Petrodiesel Blends. *Am J Analyt Chem* **2013**, *04* (07), 343–347. https://doi.org/10.4236/ajac.2013.47043.

(24) Insausti, M.; Romano, C.; Pistonesi, M. F.; Band, B. S. F. Simultaneous Determination of Quality Parameters in Biodiesel/Diesel Blends Using

Synchronous Fluorescence and Multivariate Analysis. *Microchemical Journal* **2013**, *108*, 32–37. https://doi.org/10.1016/j.microc.2012.12.007.

(25) Inan, T. Y.; Al-Hajji, A.; Koseoglu, O. R. Chemometrics-Based Analytical Method Using FTIR Spectroscopic Data to Predict Diesel and Diesel/Diesel Blend Properties. *Energy and Fuels* **2016**, *30* (7), 5525–5536. https://doi.org/10.1021/acs.energyfuels.6b00731.

(26) de Fátima Bezerra de Lira, L.; de Vasconcelos, F. V. C.; Pereira, C. F.; Paim, A. P. S.; Stragevitch, L.; Pimentel, M. F. Prediction of Properties of Diesel/Biodiesel Blends by Infrared Spectroscopy and Multivariate Calibration. *Fuel* **2009**, *89* (2), 405–409. https://doi.org/10.1016/j.fuel.2009.05.028.

(27) Guerrero, A.; Anguebes, F.; Castelán, M.; Morales, V.; García, R.; Córdova, A. V.; Zavala, J. C. FTIR-ATR and Multivariate Calibration for the Prediction of Biodiesel Concentration in Petrodiesel Blends. *Am J Analyt Chem* **2013**, *04* (07), 343–347. https://doi.org/10.4236/ajac.2013.47043.

(28) Aleme, H. G.; Barbeira, P. J. S. Determination of Biodiesel Content in Diesel Using Distillation Curves and Multivariate Calibration. *Energy and Fuels* **2012**, *26* (9), 5769–5774. https://doi.org/10.1021/ef3008757.

(29) Aleme, H. G.; Assunção, R. A.; Carvalho, M. M. O.; Barbeira, P. J. S. Determination of Specific Gravity and Kinematic Viscosity of Diesel Using Distillation Curves and Multivariate Calibration. *Fuel Processing Technology* **2012**, *102*, 90–95. https://doi.org/10.1016/j.fuproc.2012.04.016.

(30) Frackowiak, D. *News and Views The Jablonski Diagram*; 1988; Vol. 2.

(31) Skoog, D. A.; West, D. M.; Holler, J. *Fundamentals of Analytical Chemistry, Sixth Edition (Hollar, James F.; Skoog, Douglas A.; West, Donald M.)*; 2007. https://pubs.acs.org/sharingguidelines.

(32) Siesler, H. W. , et al. , eds. . *Near-Infrared Spectroscopy: Principles, Instruments, Applications*; John Wiley & Sons, 2008.

(33) Wilkins, C. Fourier Transform Spectrometry By Sumner P. Davis (University of California, Berkeley), Mark C. Adams (ITT Industries Aerospace/Communicatons, Fort Wayne, Indiana), and James W. Brault (Formerly of National Solar Observatory, Kitt Peak, Arizona). Academic Press: San Diego. 2001. Xiv + 262 Pp. $79.95. ISBN 0-12-042510-6. *J Am Chem Soc* **2002**, *124* (19), 5601–5601. https://doi.org/10.1021/ja0153060.

(34) Kallevik, H.; Hansen, S. B.; Sæther, Ø.; Kvalheim, O. M.; Sjöblom, J. Crude Oil Model Emulsion Characterised by Means of near Infrared Spectroscopy and Multivariate Techniques. *J Dispers Sci Technol* **2000**, *21* (3), 245–262. https://doi.org/10.1080/01932690008913265.

(35) Zimmermann, B.; Kohler, A. Optimizing Savitzky-Golay Parameters for Improving Spectral Resolution and Quantification in Infrared Spectroscopy. *Appl Spectrosc* **2013**, *67* (8), 892–902. https://doi.org/10.1366/12-06723.

(36) Rinnan, Å.; Berg, F. van den; Engelsen, S. B. Review of the Most Common Pre-Processing Techniques for near-Infrared Spectra. *TrAC - Trends in Analytical Chemistry* **2009**, *28* (10), 1201–1222. https://doi.org/10.1016/j.trac.2009.07.007.

(37) Savitzky, A.; Golay, M. J. *Smoothing and Differentiation of Data Dy Simplified Least Squares Procedures*; 1964; Vol. 40. https://doi.org/https://doi.org/10.1021/ac60214a047.

(38) Afseth, N. K.; Kohler, A. Extended Multiplicative Signal Correction in Vibrational Spectroscopy, a Tutorial. *Chemometrics and Intelligent Laboratory Systems* **2012**, *117*, 92–99. https://doi.org/10.1016/j.chemolab.2012.03.004.

(39) Chen, Z. P.; Morris, J.; Martin, E. Extracting Chemical Information from Spectral Data with Multiplicative Light Scattering Effects by Optical Path-Length Estimation and Correction. *Anal Chem* **2006**, *78* (22), 7674–7681. https://doi.org/10.1021/ac0610255.

(40) Martens, H.; Nielsen, J. P.; Engelsen, S. B. Light Scattering and Light Absorbance Separated by Extended Multiplicative Signal Correction. Application to near-Infrared Transmission Analysis of Powder Mixtures. *Anal Chem* **2003**, *75* (3), 394–404. https://doi.org/10.1021/ac020194w.

(41) Santos Panero, P. dos; Santos Panero, F. dos; Santos Panero, J. dos; Bezerra da Silva, H. E. Application of Extended Multiplicative Signal Correction to Short-Wavelength near Infrared Spectra of Moisture in Marzipan. *Journal of Data Analysis and Information Processing* **2013**, *01* (03), 30–34. https://doi.org/10.4236/jdaip.2013.13005.

(42) Martens, H.; Jensen, S. A.; Geladi, P. Multivariate Linearity Transformation for Near-Infrared Reflectance Spectrometry. *Proceedings of the Nordic symposium on applied statistics* **1983**, 205–234.

(43) Geladi, P.; MacDougall, D.; Martens, H. Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat. *Appl Spectrosc* **1985**, *39* (3), 491–500.

(44) Martens, H.; Stark, E. Extended Multiplicative Signal Correction and Spectral Interference Substraction: New Preprocesseing Method for near Infrared Spectroscopy. *J Pharm Biomed Anal* **1991**, *9* (8), 625–635. https://doi.org/10.1002/bate.200810020.

(45) Santos Panero, P. dos; Santos Panero, F. dos; Santos Panero, J. dos; Bezerra da Silva, H. E. Application of Extended Multiplicative Signal Correction to Short-Wavelength near Infrared Spectra of Moisture in Marzipan. *Journal of Data Analysis and Information Processing* **2013**, *01* (03), 30–34. https://doi.org/10.4236/jdaip.2013.13005.

(46) Chen, Z. P.; Morris, J.; Martin, E. Extracting Chemical Information from Spectral Data with Multiplicative Light Scattering Effects by Optical Path-Length Estimation and Correction. *Anal Chem* **2006**, *78* (22), 7674–7681. https://doi.org/10.1021/ac0610255.

(47) Martens, H.; Nielsen, J. P.; Engelsen, S. B. Light Scattering and Light Absorbance Separated by Extended Multiplicative Signal Correction. Application to near-Infrared Transmission Analysis of Powder Mixtures. *Anal Chem* **2003**, *75* (3), 394–404. https://doi.org/10.1021/ac020194w.

(48) Ottestad, S.; Isaksson, T.; Saeys, W.; Wold, J. P. Scattering Correction by Use of a Priori Information. *Appl Spectrosc* **2010**, *64* (7), 795–804. https://doi.org/10.1366/000370210791666381.

(49) Martens, H.; Stark, E. *Extended Multiplicative Signal Correction and Spectral Interference Subtraction: New Preprocessing Methods for near Infrared Spectroscopy**; 1991; Vol. 9.

(50) Alves, J. C. L.; Poppi, R. J. Near-Infrared Spectroscopy in Analysis of Crudes and Transportation Fuels. *Encyclopedia of Analytical Chemistry* **2015**, 1–16. https://doi.org/10.1002/9780470027318.a1817.pub2.

(51) Sotelo, F.; Araujo Pantoja, F. ; López-Gejo, P. ; le Roux, J. ; Quina, G. A. C. ; Nascimento, F. H. ; Sotelo, F. F.; Araujo Pantoja, P.; López-Gejo, J.; le Roux, G. A. C.; Quina, F. H.; Nascimento, C. A. O. BRAZILIAN JOURNAL OF PETROLEUM AND GAS APPLICATION OF FLUORESCENCE SPECTROSCOPY FOR SPECTRAL DISCRIMINATION OF CRUDE OIL SAMPLES. *Brazilian Journal of Petroleum and Gas* **2008**, No. 2, 63–71.

(52) Pantoja, P. A.; López-Gejo, J.; le Roux, G. A. C.; Quina, F. H.; Nascimento, C. A. O. Prediction of Crude Oil Properties and Chemical Composition by Means of Steady-State and Time-Resolved Fluorescence. *Energy and Fuels* **2011**, *25* (8), 3598–3604. https://doi.org/10.1021/ef200567x.

(53) Divya, O.; Mishra, A. K. Multivariate Methods on the Excitation Emission Matrix Fluorescence Spectroscopic Data of Diesel-Kerosene Mixtures: A Comparative Study. *Anal Chim Acta* **2007**, *592* (1), 82–90. https://doi.org/10.1016/j.aca.2007.03.079.

(54) Taksande, A.; Hariharan, C. Synchronous Fluorescence Method to Check Adulteration of Petrol and Diesel by Kerosene. *Spectroscopy Letters* **2006**, *39* (4), 345–356. https://doi.org/10.1080/00387010600781340.

(55) Divya, O.; Mishra, A. K. Combining Synchronous Fluorescence Spectroscopy with Multivariate Methods for the Analysis of Petrol-Kerosene Mixtures. *Talanta* **2007**, *72* (1), 43–48. https://doi.org/10.1016/j.talanta.2006.09.032.

(56) Chung, H.; Choi, H. J.; Ku, M. S. Rapid Identification of Petroleum Products by Near-Infrared Spectroscopy. *Bull Korean Chem Soc* **1999**, *20* (9), 1021–1025.

(57) Macho, S.; Larrechi, M. S. Near-Infrared Spectroscopy and Multivariate Calibration for the Quantitative Determination of Certain Properties in the Petrochemical Industry. *TrAC - Trends in Analytical Chemistry* **2002**, *21* (12), 799–806. https://doi.org/10.1016/S0165-9936(02)01202-5.

(58) Zhu, L.; Lu, S. H.; Zhang, Y. H.; Zhai, H. L.; Yin, B.; Mi, J. Y. An Effective and Rapid Approach to Predict Molecular Composition of Naphtha Based on Raw NIR Spectra. *Vib Spectrosc* **2020**, *109* (March), 103071. https://doi.org/10.1016/j.vibspec.2020.103071.

(59) Da Silva, V. H.; Reboucas, M. V.; Salles, A. R.; Pimentel, M. F.; Pontes, M. J. C.; Pasquini, C. Determination of Naphtha Composition by near Infrared Spectroscopy and Multivariate Regression to Control Steam Cracker Processes. *Fuel Processing Technology* **2015**, *131*, 230–237. https://doi.org/10.1016/j.fuproc.2014.10.035.

(60) Reboucas, M. V.; Santos, E. C.; Vieira, F. S. V. Feasibility of Quality Process Control of a Naphtha Fractioning Unit Based on Near-Infrared Spectroscopic Prediction of Physical and Chemical Properties of Medium Naphtha Streams. *Vib Spectrosc* **2007**, *44* (1), 187–191. https://doi.org/10.1016/j.vibspec.2006.09.004.

(61) Chung, H.; Choi, H. J.; Ku, M. S. Rapid Identification of Petroleum Products by Near-Infrared Spectroscopy. *Bull Korean Chem Soc* **1999**, *20* (9), 1021–1025.

(62) Macho, S.; Larrechi, M. S. Near-Infrared Spectroscopy and Multivariate Calibration for the Quantitative Determination of Certain Properties in the Petrochemical Industry. *TrAC - Trends in Analytical Chemistry* **2002**, *21* (12), 799–806. https://doi.org/10.1016/S0165-9936(02)01202-5.

(63) Chung, H. Applications of Near-Infrared Spectroscopy in Refineries and Important Issues to Address. *Applied Spectroscopy Reviews*. 2007, pp 251–285. https://doi.org/10.1080/05704920701293778.

VITA

PERSONAL INFORMATION

Surname, Name: Meşe Sezen, Ayten Ekin

EDUCATION

Degree	Institution	Year of Graduation
MSc	Izmir Institute of Technology	2016
BSc	Izmir Institute of Technology	2014

WORK EXPERIENCE

Year	Place	Enrollment
2018 – present	Turkish Petroleum Refineries Corporation (TUPRAS)	R&D Researcher

PUBLICATIONS

- Cebi, N., Dogan, C. E., Mese, A. E., Ozdemir, D., Arıcı, M., & Sagdic, O. (2019). A rapid ATR-FTIR spectroscopic method for classification of gelatin gummy candies in relation to the gelatin source. Food chemistry, 277, 373-381. https://doi.org/10.1016/j.foodchem.2018.10.125
- Uçar, Ö. İ., Meşe, A. E., Birbaşar, O., Dündar, M., & Özdemir, D. (2017). Determination of Aluminum Oxide Thickness on the Annealed Surface of 8000 Series Aluminum Foil by Fourier Transform Infrared Spectroscopy. In Light Metals 2017 (pp. 273-278). Springer, Cham. https://doi.org/10.1007/978-3-319-51541-0_36