DETERMINATION OF CRUCIAL PARAMETERS IN GASOLINE BLENDS BY USING INFRARED SPECTROSCOPY COUPLED WITH MULTIVARIATE CALIBRATION METHODS

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

by
Fatma Nur SAKALLI

December 2021
İZMİR

ACKNOWLEDGEMENT

Many people made this thesis possible, and I am pleased to thank them here. I am very much thankful to my advisor, Prof. Dr. Durmuş ÖZDEMİR for his valuable guidance, patience, and encouragement during the whole period of my master thesis and for everything I've learned from him. Besides my advisor, I would like to thank Prof. Dr. Figen TOKATLI and Assoc. Prof. Dr. Hasan ERTAŞ for accepting to be the jury in my defense exam. I would like to thank TUPRAS Izmir Refinery Quality Control Laboratory and TUPRAS R&D group for their great support. I also thank my husband for his support and my family for their encouragement through all my education.

ABSTRACT

DETERMINATION OF CRUCIAL PARAMETERS IN GASOLINE BLENDS BY USING INFRARED SPECTROSCOPY COUPLED WITH MULTIVARIATE CALIBRATION METHODS

In petroleum refineries, converting the manual gasoline blending system to an automatic inline blending system provides the most economical blending in gasoline production, increasing efficiency, and reliability. The most important requirement for an automatic inline blending system is the determination of gasoline parameters in a short time with high reliability. For this purpose, fast and simple analytical methods have been developed to determine crucial parameters of gasoline blends by using Fourier Transform Infrared Spectroscopy (FTIR) coupled with multivariate calibration methods which are Partial Least Squares Regression (PLSR) and Genetic Inverse Least Squares Regression (GILS) for this study. Turkey Petroleum Refinery Incorporated Company (TUPRAS) Izmir Refinery collected all gasoline samples and tested them using reference test methods at Quality Control Laboratory. Since commercial product samples were used in this study, the data ranges of the parameters were quite narrow. The Standard Error of Cross-Validation (SECV) and Standard Error of Prediction (SEP) values were acceptable, although the determintion coefficient (R^2) value of some parameters was below the expectation. It has been observed that the prediction results of GILS are better in these parameters, whose R^2 value is low because the data range is very narrow. In the comparison made with the reproducibility values specified in the reference measurement methods, it was determined that the calibration model results of most parameters were acceptable. Collecting more samples in a longer time interval to expand the data range of the parameters, or preparing a data set with experimental design can improve the prediction performance.

ÖZET

BENZİN KARIŞIMLARININ ÖNEMLİ PARAMETRELERİNİN ÇOK DEĞİŞKENLİ KALİBRASYON METOTLARI İLE DESTEKLENMİŞ KIZILÖTESİ SPEKTROSKOPİSİ YÖNTEMİ İLE TAYİNİ

Petrol rafinerilerinde verimliliğin ve güvenilirliğin arttırılması için sürekli iyileştirme çalışmaları yapılmaktadır. Benzin üretimindeki verimliliğin ve güvenilirliğin arttırılması için ürünlere ait önemli parametrelerin kısa süre içerisinde tayin edilmesi önemlidir. Bu amaçla, bu çalışmada Fourier Dönüşümü Kızılötesi Spektroskopisi (FTIR) ile çok değişkenli kalibrasyon yöntemleri olan Kısmi En Küçük Kareler Regresyon (PLSR) ve Genetik Ters En Küçük Kareler Regresyon (GILS) kullanılarak benzin karışımlarının önemli parametrelerini belirlemek için hızlı ve basit analitik yöntem geliştirilmiştir. Türkiye Petrol Rafinerileri A.Ş. (TÜPRAŞ) İzmir Rafinerisi, tüm benzin numunelerini toplamış ve Kalite Kontrol Laboratuvarı'nda referans test yöntemleri ile test etmiştir. Bu çalışmada ticari ürün numuneleri kullanıldığı için parametrelerin veri aralıkları dar olmasına bağlı olarak, bu parametrelerde Korelasyon Katsayısı (R^2) değeri beklentinin altında olduğu, fakat Standart çapraz doğrulama hatası (SECV) ve standart tahmin hatası (SEP) değerleri kabul edilebilir olduğu tespit edildi. Veri aralığı çok dar olduğu için R^2 değeri düşük olan bu parametrelerde GILS model tahmin sonuçlarının daha iyi olduğu gözlemlenmiştir. Referans ölçüm yöntemlerinde belirtilen tekrar üretilebilirlik değerleri ile yapılan karşılaştırmada çoğu parametrenin kalibrasyon modeli tahminlerinin kabul edilebilir olduğu gözlenmiştir. Parametrelerin veri aralığını genişletmek için daha uzun bir zaman aralığında daha fazla örnek toplamak veya deneysel tasarım ile bir veri seti hazırlamak tahmin performansını iyileştirebilir.

TABLE OF CONTENTS

LIST OF FIGURES viii

LIST OF TABLES xi

CHAPTER 1 INTRODUCTION 1

1.1. Literature Review 4

1.2. Structure and Purpose of The Thesis 5

CHAPTER 2 PETROLEUM AND PETROLEUM REFINING 7

2.1. Petroleum Refining 7

2.2. Petroleum Products 8

2.2.1. LPG 8

2.2.2. Gasoline 8

2.2.3. Jet Fuels 9

2.2.4. Kerosene 9

2.2.3. Diesel And Fuel Oils 9

2.2.5. Asphalt/Tar 10

2.3. Parameters of Gasoline 10

2.3.1. Density 11

2.3.2. Octane Numbers 12

2.3.3. Distillation Points 12

2.3.4. Hydrocarbon Composition 14

2.3.5. Methyl Tertiary-Butyl Ether - MTBE Content 14

CHAPTER 3 CHEMOMETRICS 16

3.1. Infrared Spectroscopy 17

3.2. Calibration 18

3.2.1. Univariate Calibration 18

3.2.2. Multivariate Calibration 19

3.2.2.1. Classical Least Squares 20

3.2.2.2. Inverse Least Squares 21

3.2.2.3. Partial Least Squares Regression 22

3.2.2.4. Genetic Inverse Least Squares 25
3.2.2.4.1. Selection of Initial Genes 26
3.2.2.4.2. Evaluation of the Population 27
3.2.2.4.3. Selection of the Parent Genes for Breeding and Mating 28
3.2.2.4.4. Cross-Over and Mutations 28
3.2.2.4.5. Replacing Parents with Offspring 29
3.3. Preprocessing Techniques 29
3.3.1. Mean Centering And Scaling 30
3.3.1. Removing Variables 30
3.3.2. Baseline Correction 30
CHAPTER 4 EXPERIMENTATION AND INSTRUMENTATION 32
4.1. Experimentation 32
4.1.1. Sample Collection And Sample Preparation 32
4.2. Instrumentation 32
4.2.1. Density 32
4.2.2. Octane Numbers – RON and MON 33
4.2.3. Distillation Points 34
4.2.4. Hydrocarbon Content 35
4.2.5. Fourier Transform Infrared Spectrometer – FTIR 36
4.3. Data Analysis 37
CHAPTER 5 RESULTS AND DISCUSSION 38
5.1. FTIR Spectra 38
5.2. Partial Least Squares Regression 39
5.3. Genetic Inverse Least Square Regression 41
5.4. Multivariate Calibration Results and Comparison 42
5.4.1. Density 42
5.4.2. E70 - Evaporated at 70 °C 45
5.4.3. E100 - Evaporated at 100 °C 49
5.4.4. E150 - Evaporated at 150 °C 52
5.4.5. FBP – Final Boiling Point 55
5.4.6. MTBE Content - Methyl Tertiary Butyl Ether 58
5.4.7. Benzene Content 61

5.4.8. Olefin Content 64
5.4.9. Aromatic Contents 67
5.4.10. RON - Research Octane Number 70
5.4.11. MON - Motor Octane Number 73
CHAPTER 6 CONCLUSION 77
REFERENCES 78

LIST OF FIGURES

Figure **Page**

Figure 1: Atmospheric Distillation of Crude Oil 2
Figure 2: Process involved in the development of the calibration models for fuel properties from IR spectra 3
Figure 3: Distillation curve of gasoline. Correlation of distillation profile ranges with gasoline performance 13
Figure 4: FTIR Simple block diagram for the instrumentation of FTIR Spectroscopy.. 18
Figure 5: PLS Calibration 24
Figure 6: A sample plot of the number of components vs. PRESS values 25
Figure 7: Flow Chart of Typical Genetic Algorithm for Multivariate Calibration..... 26
Figure 8: Example of Roulette Wheel selection method 28
Figure 9: Density meter 33
Figure 10: CFR Engine 34
Figure 11: Atmospheric Distillation Instrument 35
Figure 12: Gas Chromatograph 36
Figure 13: FTIR Analyser 37
Figure 14: FTIR spectra of 110 Gasoline Sample 38
Figure 15: FTIR spectra of 110 Gasoline Samples with Narrowed Range 39
Figure 16: Number of PCs vs. PRESS plot for selecting the optimal number of LVs a) Density, b) E70, c) E100, d) E150, e) FBP, f) MTBE content, g) Benzene content, h) Olefin content, i) Aromatic content, j) RON, k) MON..... 41
Figure 17: Reference Density Value vs Predicted Density Value; a) PLS, B)GILS 42
Figure 18: Reference Density vs. Corresponding Model Prediction Density Residuals; a) PLS, b)GILS 43
Figure 19: Density Residual R-Chart; a)PLS, b)GILS 44
Figure 20: Reference E70 Value vs Predicted E70 Value; a) PLS, B)GILS 45
Figure 21: Reference E70 vs. Corresponding Model Prediction E70 Residuals; a) PLS, b)GILS 47
Figure 22: E70 Residual R-Chart; a)PLS, b)GILS 48
Figure 23: Reference E100 Value vs Predicted E100 Value; a) PLS, B)GILS 49

Figure **Page**

Figure 24: Reference E100 vs. Corresponding Model Prediction E100 Residuals; a) PLS, b)GILS 50
Figure 25: E100 Residual R-Chart; a)PLS, b)GILS 51
Figure 26: Reference E150 Value vs Predicted E150 Value; a) PLS, B)GILS 52
Figure 27: Reference E150 vs. Corresponding Model Prediction E150 Residuals; a) PLS, b)GILS 53
Figure 28: E150 Residual R-Chart; a)PLS, b)GILS 54
Figure 29: Reference FBP Value vs Predicted FBP Value; a) PLS, B)GILS 55
Figure 30: Reference FBP vs. Corresponding Model Prediction FBP Residuals; a) PLS, b)GILS 56
Figure 31: FBP Residual R-Chart; a)PLS, b)GILS 57
Figure 32: Reference MTBE Value vs Predicted MTBE Value; a) PLS, B)GILS 58
Figure 33: Reference MTBE vs. Corresponding Model Prediction MTBE Residuals; a) PLS, b)GILS 59
Figure 34: MTBE Residual R-Chart; a)PLS, b)GILS 60
Figure 35: Reference Benzene Value vs Predicted Benzene Value; a) PLS, B)GILS ... 61
Figure 36: Reference Benzene vs. Corresponding Model Prediction Benzene Residuals; a) PLS, b)GILS 62
Figure 37: Benzene Residual R-Chart; a)PLS, b)GILS 63
Figure 38: Reference Olefin Value vs Predicted Olefin Value; a) PLS, B)GILS 64
Figure 39: Reference Olefin vs. Corresponding Model Prediction Olefin Residuals; a) PLS, b)GILS 65
Figure 40: Olefin Residual R-Chart; a)PLS, b)GILS 66
Figure 41: Reference Aromatic Value vs Predicted Aromatic Value; a) PLS, b)GILS . 67
Figure 42: Reference Aromatic vs. Corresponding Model Prediction Aromatic Residuals; a) PLS, b)GILS 68
Figure 43: Aromatic Residual R-Chart; a)PLS, b)GILS 69
Figure 44: Reference RON Value vs Predicted RON Value; a) PLS, B)GILS 70
Figure 45: Reference RON vs. Corresponding Model Prediction RON Residuals; a) PLS, b)GILS 71
Figure 46: RON Residual R-Chart; a)PLS, b)GILS 72

Figure **Page**

Figure 47: Reference MON Value vs Predicted MON Value; a) PLS, B)GILS 73
Figure 48: Reference MON vs. Corresponding Model Prediction MON Residuals; a) PLS, b)GILS 74
Figure 49: MON Residual R-Chart; a)PLS, b)GILS 75

LIST OF TABLES

Table **Page**

Table 1: Crucial Parameters of Gasoline and Relevant Test Methods 11
Table 2: FTIR Specifications 37
Table 3: Calibration Model Performance for Density 43
Table 4: Paired t-test Results for Density 45
Table 5: Calibration Model Performance for E70 46
Table 6: Paired t-test Results for E70 48
Table 7: Calibration Model Performance for E100 50
Table 8: Paired t-test Results for E100 51
Table 9: Calibration Model Performance for E150 53
Table 10: Paired t-test Results for E150 54
Table 11: Calibration Model Performance for FBP 56
Table 12: Paired t-test Results for FBP 57
Table 13: Calibration Model Performance for MTBE Content 59
Table 14: Paired t-test Results for MTBE 60
Table 15: Calibration Model Performance for Benzene Content 62
Table 16: Paired t-test Results for Benzene Content 63
Table 17: Calibration Model Performance 65
Table 18: Paired t-test Results for Olefin Content 66
Table 19: Calibration Model Performance for Aromatic Content 68
Table 20: Paired t-test Results for Aromatic Content 69
Table 21: Calibration Model Performance for RON 71
Table 22: Paired t-test Results for RON 72
Table 23: Calibration Model Performance for MON 74
Table 24: Paired t-test Results for MON 75

CHAPTER 1

INTRODUCTION

The word petroleum is formed by the combination of the Latin “Petro” (Stone) and “Oleum” (Oil). It means stone oil. Petroleum is a very complex compound consisting mainly of hydrogen and carbon with small amounts of nitrogen, oxygen, and sulfur and has no simple formula. It can exist in gas, liquid, and solid form under normal conditions (National Centre for Catalysis Research, 2006).

Crude oil is a mixture of hydrocarbon molecules containing 1 to 60 carbon atoms (Beşergil, 2009). The properties of hydrocarbons depend on the number and arrangement of carbon and hydrogen atoms in their molecules. Crude oil, for example, is not a single compound like water; some are a mixture of large and some small hydrocarbon molecules. Crude oil starts to boil at room temperature, the steam above the liquid rises as the boiling continues as long as it is heated; the reason is that as the temperature rises, various hydrocarbon molecules in crude oil leave the liquid by evaporation or boiling. Higher temperatures are required for the chemometricmixture to boil. These distillation steps are shown in Figure 1.

The composition and appearance of crude oil differ according to the region and location, its fluidity varies from watery to tar-like solids, and its color varies from light to black. An average crude oil consists of 84% carbon, 14% hydrogen, 1-3% sulfur, and less than 1% nitrogen, oxygen, metals, and salts (Beşergil, 2009).

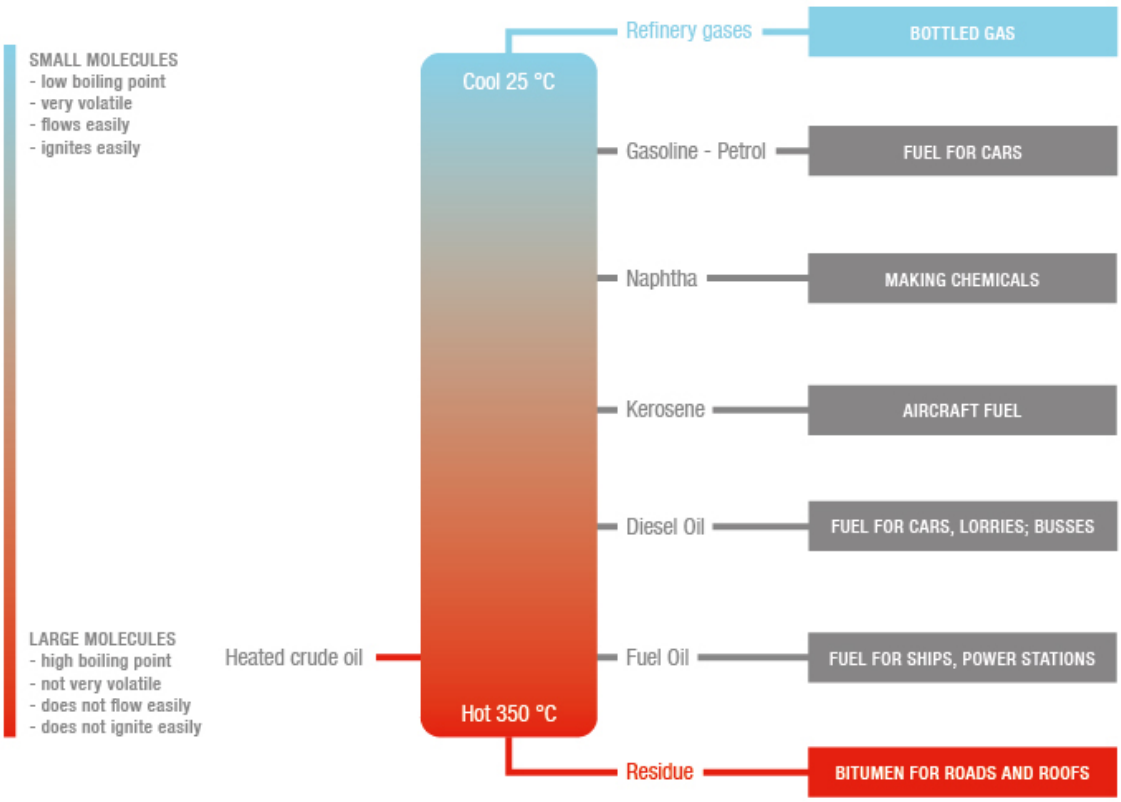


Figure 1: Atmospheric Distillation of Crude Oil

(Anton-Paar, 2009)

As can be seen in Figure 1, a wide variety of products are formed by the distillation of petroleum. Each of these products is used for different purposes due to their different performances. Gasoline, which is formed by the chemical processes of crude oil, is still one of the largest energy sources for automobiles. Gasoline used for vehicles has some properties and these vary according to the structure, characteristics, and operating conditions of the engine.

One of the significant properties of gasoline is specific gravity. High specific gravity will increase the boiling point of gasoline, making it difficult to start, negatively affecting the operation of the engine at low temperatures and speeds. If the evaporation pressure of gasoline is low and the air is cold, it makes it difficult for the engine to start.

The distillation feature of gasoline shows the evaporation ability of gasoline. The distillation temperature is highly effective on the operation of the engine. Starting the engine at 10% distillation temperature and in cold weather affects factors such as the acceleration of engine oil contamination. The higher the octane number, which shows the

resistance of gasoline against knocking, the less chance of self-ignition of gasoline, and there will be no knocking.

For the measurements of all these parameters mentioned, different tests are carried out in the laboratories. These tests require both labor and time. However, with the developing technology, studies have been made on the use of Infrared Spectroscopy for the determination of all these important parameters. It is aimed to save both time and labor, and at the same time, higher measurement accuracy that meets international standards is aimed. Figure 2 shows the basic steps of the calibration models for the determination of fuel properties with IR spectra.

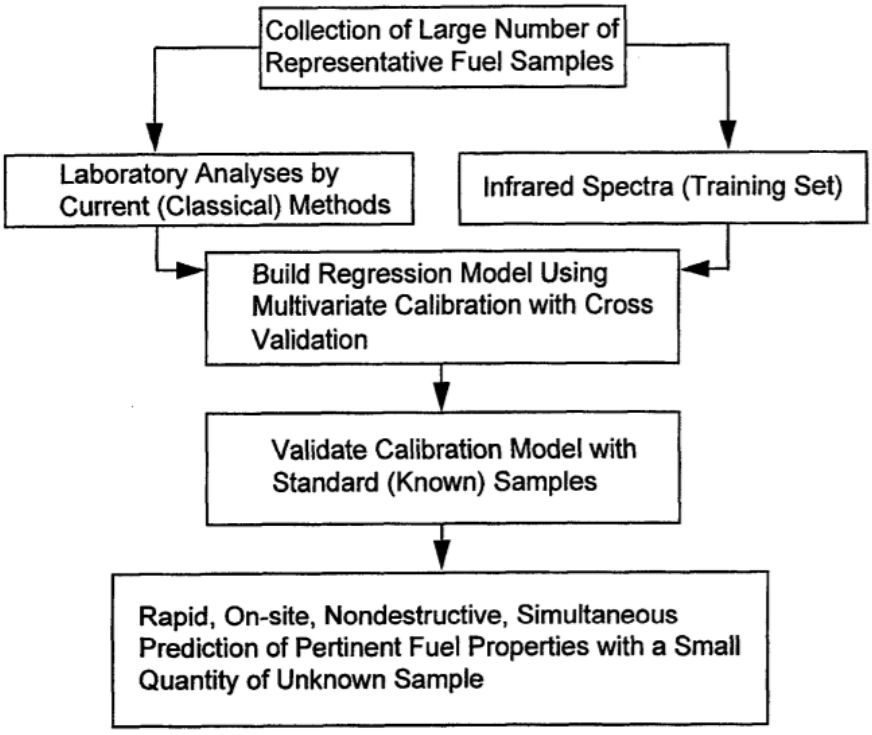


Figure 2: Process involved in the development of the calibration models for fuel properties from IR spectra

(Fodor & Hutzler, 1997)

1.1. Literature Review

Due to the diversity of products, parameters, and methods, there are many studies in the literature containing infrared spectroscopy coupled with multivariate calibration models for the characterization of petroleum products.

In a study by Özdemir, it was shown that near-infrared (NIR) spectroscopy and three different genetic algorithm-based multivariate calibration approaches (GILS, GR, GCLS) can be used to rate the octane number of gasoline (Özdemir, 2005). In this study, a data set consisting of 60 gasoline samples with known octane numbers and the spectrums (900 to 1700 nm spectral region) of those samples, previously created by John H. Kalivas (Kalivas, 1997), was used. By using a genetic algorithm, it was tried to improve the results of the calibration study, which was done with the least square and classical methods before. It was stated that the calibrations using the genetic algorithm were successful. Bohács and friends tried to present a feasible procedure for the prediction of quality parameters of gasoline from its NIR spectrum in a large and very diverse sample set (Bohács, Ovádi, & Salgó, 1998). They developed NIR methods with Partial Least Square Regression(PLSR) to predict four important gasoline properties (RON, MON, benzene, MTBE content, evaporation values, density, Reid vapor pressure, sulfur content) with reproducibilities equivalent to the standard test procedures. They used 350 gasoline samples. 900 to 1700 nm spectral range was used. It is stated research and motor octane number, benzene, and MTBE content correlation are excellent (R^2>0.97), Reid vapor pressure, sulfur content, initial and final boiling point and evaporated at 180°C point correlation as poor (R^2<0.9). Fodor& Hutzler developed to FT-IR method with PLS calibration for the rapid, simultaneous determination of several pertinent fuel properties, using less than 2 mL of sample, as an alternative to established laboratory protocols (Fodor & Hutzler, 1997). They tried to confidently measure by FT-IR: research and motor octane numbers, aromatic, olefinic, and saturated hydrocarbon content, benzene content, and concentrations of ethanol, methyl tert-butyl ether, and total oxygenated compounds. In this study, two different FTIRs were used and compared. A spectral region of 4000-650 cm^{-1} was used. A large number of summer and winter gasoline samples collected from different regions were used. In this study, analyzes were made depending on the parameters of different season product types, different devices, baseline correction,

or not. It has been stated that the prediction error increases when the calibration model prepared in one FTIR device is used for prediction in the other FTIR device. Some gasoline parameters were found to have better correlation coefficients without baseline correction. It has been observed that the correlation of the model made in winter products is better.. Reboucas also tried for prediction of key gasoline properties (density, hydrocarbon composition) by Near-infrared Spectroscopy coupled with the PLS calibration method (Reboucas, Santos, Pimentel, & Teixeira, 2011). They used Doehlert design matrix with three input variables (wavenumber range, preprocessing technique, and regression/validation technique) varied at 5, 7, and 3 levels, respectively, and improved the efficiency of PLS modeling. 10,000 – 3900 cm^{-1} spectral region was used. They have stated that PLS modeling became more efficient because of the use of experimental design, which required less experimental effort and ensured the accurate simultaneous evaluation of important response variables. In a study by Özdemir, the application of genetic algorithm-based multivariate calibration to the near-infrared spectroscopic determination of several diesel fuel parameters was demonstrated (Özdemir, 2008). Genetic Inverse Least Square Regression was used in this study. It is tried to determine: cetane number; boiling point, °C; freezing point, °C; total aromatic, (w/w)%; viscosity, cSt; density, g/mL. It has been managed to keep the percent standard error of prediction (SEP%) below 5%. Breitkreutz tried to determine total sulfur in diesel fuel employing NIR spectroscopy and multivariate calibration (Breitkreitz, et al., 2003). In this study, multiple linear regression (MLR) following variable selection using the genetic algorithm (GA) or the successive projection algorithm was compared to the performance of principal component regression (PCR) and partial least square (PLS) chemometric approaches. 97 diesel samples were used, and they used 4000 to 12,820 cm^{-1} spectral region for NIR analysis. They determined that the use of variable selection methods increases the success of the model prediction and the genetic algorithm with multiple linear regression model gives the better prediction result.

1.2. Structure and Purpose of The Thesis

In the first chapter of the thesis, general information about petroleum products and crucial parameters are given with their common principles of analysis. In the second

chapter, the most common chemometric calibration methods are shared along with their motivation, mathematical background. Additionally, a study of the determination of crucial parameters of blended gasoline was provided in the third chapter in which different calibration techniques were compared.

TUPRAS İzmir Refinery is planning to convert the existing manual Gasoline Blending system to an automated In-Line Blending system for enhancing efficiency and reliability thus ensuring the most economical blending in gasoline production. The in-line blending of continuous blending is the process of mixing crude oil or petrochemical feedstocks (blend components) to produce a homogenous mixture of consistent quality. The first step of this project, which belongs to the TUPRAS İzmir Refinery, is to determine the parameters of the blended gasoline while the process is in progress and to intervene in the process in a short time. In the next steps, the samples of all the lines included in the mixture are modeled and the optimization system is developed and the blended gasoline is automatically prepared to depend on the final and intermediate product parameters.

In this study, it is aimed to develop an alternative method with FTIR coupled PLS and GILS calibration methods to determine the important parameters of gasoline. The developed method was compared with both traditional methods which are CFR engine analysis (RON, MON), multidimensional gas chromatography (MTBE, Olefin, Aromatic, Benzene Content), atmospheric distillation (distillation points), automated densitometer (density), and the reproducibility criteria of the relevant methods specified in the international standards.

CHAPTER 2

PETROLEUM AND PETROLEUM REFINING

Petroleum is a fossil fuel formed over millions of years. Oil has been used for thousands of years in human history. The share of oil in energy consumption is 40%. This is followed by coal with a share of 28% and natural gas with a share of 23% (Akova, 2019). A mixture of liquid hydrocarbons stored in porous rocks deep within the Earth is called crude oil. In addition to hydrocarbon compounds in its structure, petroleum also contains organic components containing nitrogen, oxygen, and sulfur. The term crude oil is used to understand that refined petroleum products are different. In today's technology, many intermediates and fuel products that we use in our daily life are obtained by separating (distilled) crude oil into its components in refineries.

2.1. Petroleum Refining

The purpose of the refining process is to transform the natural raw material, crude oil, into salable products. Products from refineries, fuel for cars, trucks, aircraft, ships, and other vehicles, fuel for heat and power generation for industrial, commercial, and domestic use, raw materials for the petrochemical and chemical industry, specialty products such as lubricating oils, paraffin/waxes, and bitumen Energy as by-products can be grouped into heat (steam) and power (electricity) (Turkey Ministry of Environment and Urbanization, 2019).

For the production of petroleum products, raw materials are processed at different distillation plants. The combination of these processing units that convert crude oil into products with the help of supporting units and facilities is called a refinery. Market demand for the product type, available raw quality, and requirements set by authorities affect the size, configuration, and complexity of a refinery. Since these factors vary from

region to region, the facility structures of refineries also differ (Turkey Ministry of Environment and Urbanization, 2019).

2.2. Petroleum Products

2.2.1. LPG

Propane, butane, and paraffinic light hydrocarbons, which are mixtures of these two, which are released during refinery and natural gas processes, are easily liquefied under pressure. Composed primarily of propane and butane, LPG is produced for fuel purposes and is an intermediate in the production of petrochemicals. Important specification control tests include vapor pressure and pollution tests.

2.2.2. Gasoline

Gasoline product is grouped under two general groups as motor gasoline and aviation gasoline. These two types of gasoline are also divided into various classes or grades.

Motor gasoline, which is the most important refinery product, is a mixture of hydrocarbons used in internal combustion engines (other than aircraft engines) and with a boiling range of around 35-215 °C; Important quality characteristics of gasoline are octane number (anti-knock), volatility (engine running and vapor compression) and vapor pressure. Additives are added to gasoline to meet this performance and also to protect against oxidation and rust. With the development of reciprocating aircraft engines, aviation gasoline with different octane numbers (87, 100/130, and 115/145 octane) blended with lead compound additives was produced in the 1940s.

2.2.3. Jet Fuels

Jet fuels are fuels produced by blending kerosene and/or 'wide-cut' fractions used in turbine aircraft engines to meet the requirements in the specifications.

Wide-cut jet fuel is a mixture of light hydrocarbons with a distillation range of about 100-250 °C; It is prepared by blending kerosene, gasoline, or naphtha fractions in such a way that the content of aromatic hydrocarbons does not exceed 25% by volume and the vapor pressure is between 13.7 kPa - 20.6 kPa (Beşergil, 2009). Necessary additives are also added to increase the stability of the fuel and to increase its good flammability properties. Since wide-cut type jet fuels are a mixture of lighter hydrocarbons with a boiling range between gasoline and kerosene, the risk of ignition and evaporation losses in high flights is higher than in kerosene.

Kerosene-type jet fuel is a medium distillate product; distillation range (150 C - 300 °C. usually not exceeding 250 °C) and flash point are the same as kerosene. Unlike, some critical properties such as freezing point are changed.

2.2.4. Kerosene

Kerosene is a refined medium-distilled petroleum product; Its volatility is between gasoline and diesel fuel (boiling range is around 150-300 °C) (Beşergil, 2009). Kerosene is used as jet fuel and has a wide range of uses for heating purposes. Kerosene, which is defined by a small number of specification values, can also be evaluated by mixing it with diesel.

2.2.3. Diesel And Fuel Oils

Diesel Fuels: Diesel fuels (gas oil or diesel distillate), used as diesel engine fuel and heating oil, contain alkanes with 12 or more carbon atoms, boiling range is 180-380 °C. Among the quality tests, the cetane number of diesel fuel is important, which defines

the flash and pour point, clean combustion, no deposits in storage tanks, and good combustion initiation and combustion properties.

Distillate Fuel Oils: Heavy gas oil (distillate fuel oils) is used as an industrial fuel and as a raw material (starting material) in the production of some products. It contains long-chain (20-70 carbon atoms) alkanes, cycloalkanes, and aromatic hydrocarbons, boiling range 380-540 °C.

Heavy Fuel Oil (Residues): Products in this group are distillation residues; densities > 900 kg/L, flashpoints >50 °C, and kinematic viscosities at 80 °C greater than 10 cSt.

2.2.5. Asphalt/Tar

They are by-products obtained from the refining of petroleum. Asphalt can be found in different forms from highly viscous to solid. It is used in the coating of roads, airports, roof insulation, waterproofing in structures related to water. It is used in the paint industry, battery production, lining water channels, and bonding clay bricks.

Tar, on the other hand, is very similar to asphalt in appearance. There are differences in the proportions of hydrocarbons in their structure. It can be used in areas similar to asphalt. Tar can be obtained from the refining of petroleum, as well as from wood and coal.

2.3. Parameters of Gasoline

Petroleum products have international quality criteria. For this reason, the analyzes to determine these criteria should be carried out by the relevant international standard. All gasoline samples should be tested according to the standard test methods given in TS EN 228 - Automotive Fuels - Unleaded Petrol Requirements and Test Methods – Specification. In this standard, the appropriate test methods and the reliability values of the relevant methods are given in Table 1.

Table 1: Crucial Parameters of Gasoline and Relevant Test Methods

PROPERTY		UNIT	REPRODUCIBILITY	NOTE	TEST METHOD
Density		kg/m^3	0.5	-	TS EN ISO 12185
Distillation	**E 70°C**	v/v %	±0.02(150-X)	X is the measured value.	ASTM D 86
	E 100°C				
	E 150°C				
	FBP	°C	±6.78	-	
Research Octane Number, RON		-	±0.7	-	TS EN ISO 5164
Motor Octane Number, MON		-	±0.9	-	TS EN ISO 5163
Benzene		v/v %	± (0.0777Y-0.025)	Y is the mean of the two results being compared.	TS EN ISO 22854
Olefins			± (0.1176Y+0.5118)		
Aromatics			± (0.045Y+0.1384)		
MTBE			± (0.0251Y+0.3515)		

2.3.1. Density

Density (or specific gravity) is an indicator of density or weight per unit volume of fuel. Density is a basic parameter, it is usually given in kg/L or kg/m^3 for petroleum products (TS EN ISO 12185, 2007). As the density increases, the amount of energy per unit volume increases (Rand, 2010). Given the amount of fuel injected at a constant rate, the energy supplied to the engine increases with intensity, which increases engine performance. However, exhaust emissions, and especially particulates, increase under full load due to the richer mixture. On the other hand, as density decreases, volumetric fuel consumption increases.

Relative density is frequently used instead of absolute density. The mass of a given volume of fuel at a certain temperature divided by the mass of an identical volume of water at the same temperature is known as relative density or specific gravity. At 15.6°C, most vehicle fuels have relative densities between 0.70 and 0.78 (Rand, 2010).

2.3.2. Octane Numbers

The octane number is a characteristic that defines the knock-free combustion characteristics of gasoline; It is the ability of the engine to resist knocking when it is burned in the combustion chamber of the engine. There are two laboratory test methods for determining the octane number of gasoline; Research Octane Number (RON) and Motor Octane Number (MON). RON is associated with low velocity and moderate knock conditions, MON with high velocity, high-temperature knock conditions. In a gasoline example, the value of RON is always greater than MON; The difference between them is called “sensitivity”. The chemical structure affects the knocking tendency. Aromatics have a higher octane number than other types of hydrocarbons. (Sivasankar, 2008).

2.3.3. Distillation Points

Each hydrocarbon boils at a specific temperature called the "boiling point", which increases as the size of the molecule increases. For this reason, the distillation profile of gasoline, which is a mixture, shows the distribution of hydrocarbons it contains against temperature. The temperature limits of the distillation profile determine the boiling range, excluding low- and high-boiling compounds.

Gasoline is a mixture of hundreds of hydrocarbon molecules with different boiling points, and so it boils or distills over a range of temperatures, not at a single temperature like pure substances. A distillation profile (or distillation curve) is drawn by measuring the volumes of liquid that gasify and condense against increasing temperatures when gasoline is heated under specific conditions. Figure 3 shows the distillation profile of gasoline. Various ranges of a distillation profile are associated with gasoline performance. Gasoline must contain a light vaporizable fraction to provide a good start in a cold engine and must not contain fractions with boiling points above 200°C, as they would evaporate and dilute the oil in the engine.

The values of E70, E100, and E150 (evaporated at temperatures of 70, 100, and 150°C) determine the standard range for gasoline quality according to European standard EN 228. Refineries try to move T95 for gasoline at temperatures 160–170 °C in the future

(Chevron Corporation, 2009). These features have an impact on engine starting, warm-up, and vapor lock at high operating temperatures and altitudes in the case of fuels. Furthermore, the outcome is critical for the sample's safety and handling rules.

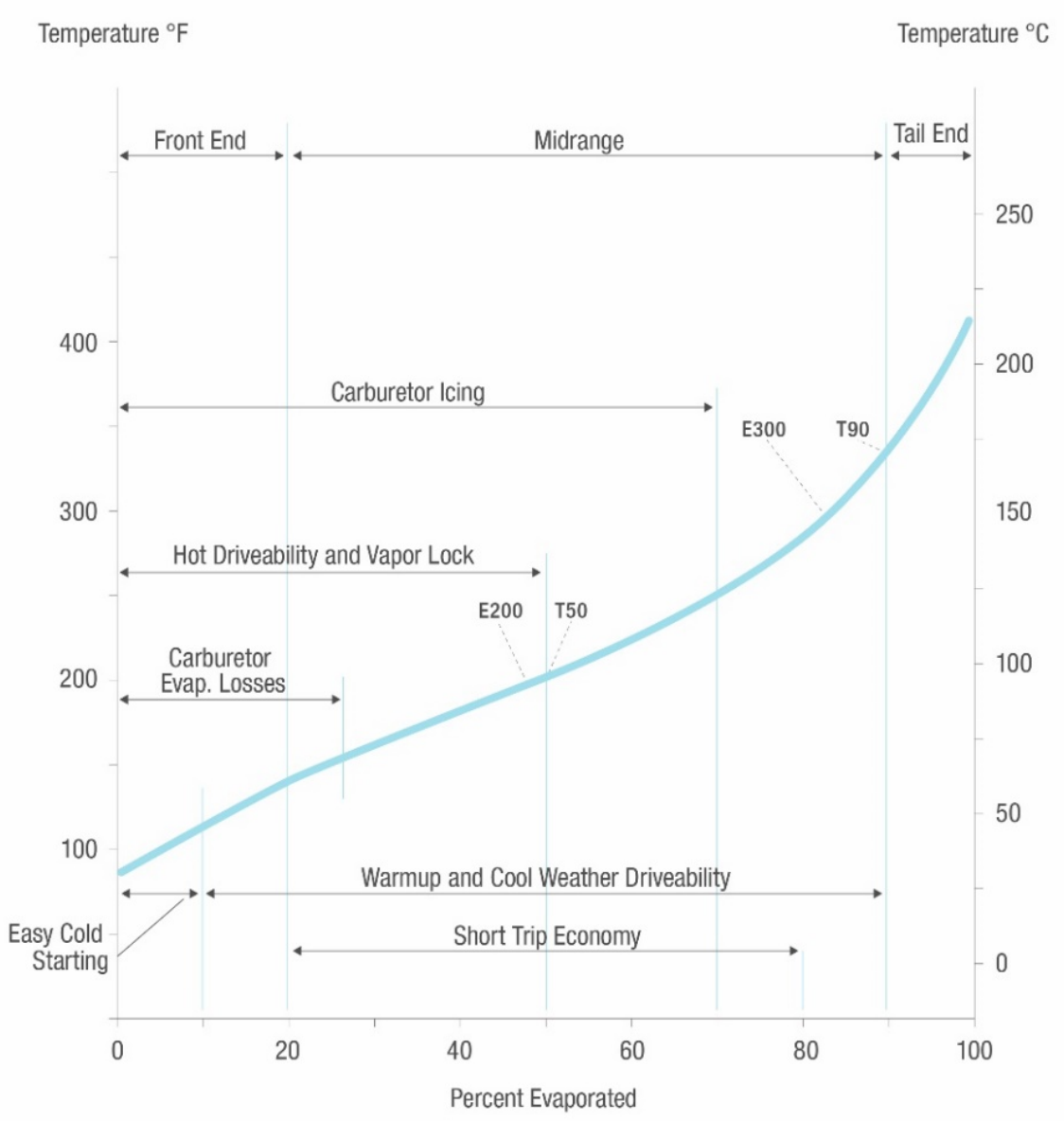


Figure 3: Distillation curve of gasoline. Correlation of distillation profile ranges with gasoline performance

(Chevron Corporation, 2009)

2.3.4. Hydrocarbon Composition

The physical and chemical properties and performance of gasoline are determined by the hydrocarbons and additives in their composition. Different carbon numbers and different hydrocarbon types are present in different proportions in the composition of gasoline. These hydrocarbon groups have different physical and chemical properties and their amounts in gasoline directly affect the fuel properties, performance, and emissions of gasoline. Although aromatic hydrocarbons and olefins also increase the knock resistance of the fuel, the ratio of these hydrocarbons in gasoline is limited because aromatic hydrocarbons increase irregular emissions (formaldehyde, acetaldehyde, benzene, toluene, and xylene), cause carbon deposits in the engine, and olefins increase the volatility of the fuel. The molecular composition of the gasoline is connected to octane. Aromatics, branched paraffin, and olefins have high octane values, but straight-chain paraffin and naphthenes (saturated cyclic hydrocarbons) have low octane values.

2.3.5. Methyl Tertiary-Butyl Ether - MTBE Content

The chemical compound MTBE (methyl tertiary-butyl ether) is produced by the reaction of methanol with isobutylene. MTBE is manufactured in enormous quantities and used nearly exclusively as a fuel additive in gasoline. Because it raises the oxygen content of gasoline, it is one among a group of compounds known as "oxygenates." MTBE is a volatile, flammable, colorless liquid that dissolves readily in water at ambient temperature.

Since 1979, MTBE has been used as an octane enhancer in U.S. gasoline at low levels to replace lead (helps prevent the engine from "knocking"). Between 1992 and 2005, higher quantities of MTBE were used in some gasoline to meet the oxygenate standards established by Congress in the 1990 Clean Air Act Amendments (Concawe, 2002)

Oxygen aids in the complete combustion of fuel, lowering hazardous tailpipe emissions from automobiles. In one way, oxygen dilutes or displaces gasoline constituents like aromatics (benzene, for example) and sulfur. Oxygen, on the other hand,

optimizes oxidation during burning. Most refiners prefer MTBE to other oxygenates because of its blending properties and cost.

CHAPTER 3

CHEMOMETRICS

Advances in computers, software, statistics, and applied mathematics have led to the birth of a new discipline called chemometrics for the solution of complex systems in the field of chemistry, especially analytical chemistry. These developments have created new fields of study with chemometric methods using multidimensional and multivariate parameters, which provide new possibilities for analytical chemistry and researchers in neighboring branches to solve analytical problems.

Chemometrics is a chemistry discipline that encompasses the processing of chemical data using computers, along with statistics and mathematics. This discipline is a powerful tool in chemical analysis that allows the extraction of real information from chemical data or the revealing of hidden information. One of the main application areas of chemometrics in analytical chemistry.

In a word, chemometrics started to be mentioned in the 1970s for its applications in chemistry, where computers and software are used, along with statistical and mathematical methods. The concept of chemometrics was put forward by Swedish Svante Wold and American Bruce R. Kowalski in 1972, and the first official explanation of this discipline was made by the International Chemometrics Association in 1974. In the following years, it is observed that national and international chemometrics conferences were also organized in the world (Dinç, 2007).

Chemometrics includes concepts and applications such as Descriptive and Inference Statistics, Signal Processing, Experimental Design, Modeling, Calibration, Optimization, Pattern Recognition, Classification, Artificial Intelligence Methods, Image Processing, Information and System Theory.

3.1. Infrared Spectroscopy

Infrared (IR) spectroscopy is one of the spectroscopic methods developed based on the interaction between light rays and the analyzed chemical molecule. In IR spectroscopy, radiation in a certain part of the electromagnetic spectrum (at a certain frequency) is used. The region between 12800-10 cm^{-1} of the electromagnetic spectrum is the infrared area, the region between 12800-4000 cm^{-1} is known as the near IR, the region between 4000-200 cm^{-1} is known as the mid-IR, and the region between 200-10 cm^{-1} is known as the far IR (Thomson, 2007).

The IR spectrum gives characteristic peaks for many groups. Thus, it contributes to the analysis of which characteristic groups are present in the substance whose spectrum we have taken, and therefore to the analysis of the structure of the substance. It is also important for us that the characteristic group peaks shift with the change in the molecular structure. For example, the C=O group gives a peak between 1900-1600 cm^{-1} in the IR, but the side of the peak in this region depends on the structure of the molecule (Ankara University, 2021).

It can be used IR spectroscopy for quantitative analysis. The analysis application can be carried out in two ways;

- According to Lambert-Beer law: In such an application, the cell thickness must be known exactly to be able to calculate. Measuring this is both very difficult and not very sensitive.
- By drawing a calibration curve: This method is more sensitive but time-consuming. In this method, firstly, solutions of many different concentrations are prepared from the substance whose concentration will be found (as in UV spectroscopy), and the observed absorption for each concentration at a characteristic peak of this substance is plotted against the concentration. The graphing equivalent of the absorption of the solution, whose concentration we do not know, at the same frequency and under the same conditions, gives us the concentration of this substance.

Fourier transform infrared (FTIR) spectrometry was developed to overcome the limitations encountered with dispersive instruments. FTIR is a chemical analytical method that measures the number of waves against the infrared intensity of light using

the mathematical Fourier transform method. Figure 4 shows typical FTIR instrumentation.

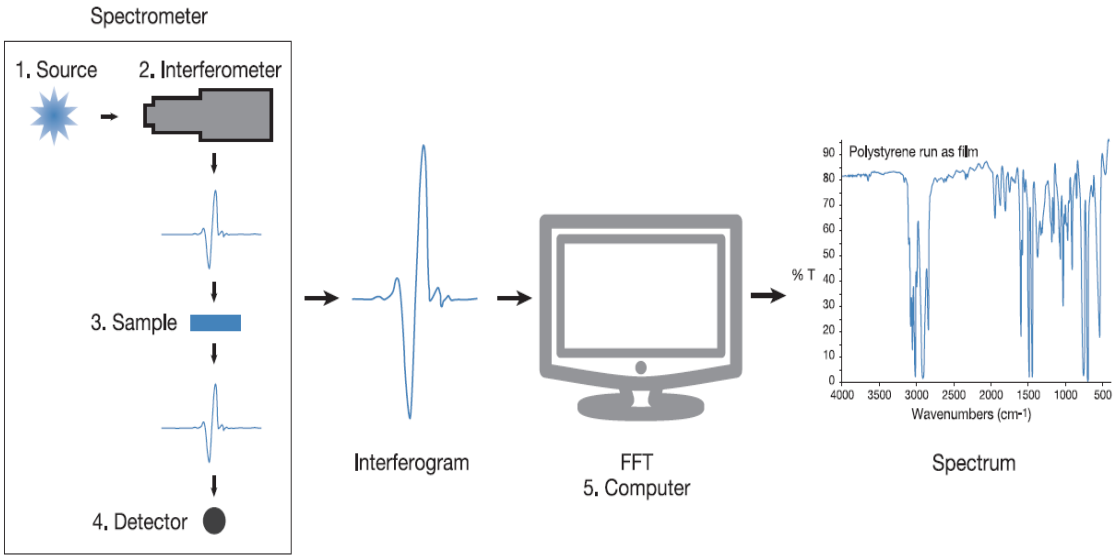


Figure 4: FTIR Simple block diagram for the instrumentation of FTIR Spectroscopy (Thermo Fisher Scientific Inc., 2013)

3.2. Calibration

Calibration is the process of transforming the signals received from the instruments into meaningful data in both quantitative and qualitative measurements. It is done by mathematical and statistical processing of signals.

3.2.1. Univariate Calibration

In univariate calibration, the aim is to correlate the property of matter with the spectral density. It is defined by the Beer-Lambert Law (Eq. 1). The law states that the concentration of a chemical is directly proportional to the absorbance of a solution. The relationship can be used to determine the concentration of a chemical species in a solution

using a colorimeter or spectrophotometer. The relationship is most commonly used in UV-visible absorption spectroscopy.

$$\mathbf{A = \varepsilon bc} \qquad \text{Eq. (1)}$$

Where **A** is absorbance, **ε** is molar absorption coefficient, **b** is optic path length **c** is molar concentration. There is a linear relationship between the concentration and the absorbance of the solution, which enables the concentration of a solution to be calculated by measuring its absorbance.

3.2.2. Multivariate Calibration

Multivariate calibrations are obtained by applying chemometric algorithms to the concentration set and corresponding multivariate measurement data. In chemometrics, various mathematical algorithms are used for multivariate calibrations.

The model may appear to be very optimistic for the data used for modeling, even though the actual performance is far worse, this is a big risk with mathematical assumptions and parameter tuning (Brereton 2003). Overfitting is the term for this phenomenon. On the other hand, the opposite situation can occur on occasion. The model fails to capture relevant data, in this case, resulting in a poor fit and poor prediction performance. As a result, validating the model against data that isn't used for modeling is critical for determining model performance. This data set is known as a validation set, an independent data set, or a test set, while the data used for modeling is known as a calibration set or a training set. Splitting the data into calibration and validation sets is common practice rather than using the entire data set for calibration. While the splitting ratio is optional, the calibration set and validation set should cover as much variance as possible.

The most common performance metric is the root-mean-squared-error (Brereton,2003). Root-Mean-Squared-Error of Calibration (RMSEC) is defined in Eq.(2) and is used for the calibration set.

$$\mathbf{RMSEC} = \sqrt{\frac{\sum_{\mathbf{i=1}}^{\mathbf{n}}(\hat{\mathbf{y}}_\mathbf{i} - \mathbf{y}_\mathbf{i})^\mathbf{2}}{\mathbf{n-2}}} \quad \text{Eq. (2)}$$

Above in Eq.(2), $\mathbf{y_i}$ is the actual variable of an observation belonging to the calibration set (i.e., concentration, activity, or another property) and $\hat{\mathbf{y}}_\mathbf{i}$ is the predicted variable that is obtained by the model. For the validation set, Root-Mean-Squared-Error of Prediction (RMSEP) is defined in Eq.(3).

$$\mathbf{RMSEP} = \sqrt{\frac{\sum_{\mathbf{i=1}}^{\mathbf{n}}(\hat{\mathbf{y}}_\mathbf{i} - \boldsymbol{y}_\mathbf{i})^\mathbf{2}}{\mathbf{n}}} \quad \text{Eq. (3)}$$

In Eq.(3), $\mathbf{y_i}$ now refers to the actual variable of a sample from the validation set and $\hat{\mathbf{y}}_\mathbf{i}$ is its model prediction. The overfitting can now be defined as having very low RMSEC and high RMSEP value whereas underfitting is evident with low RMSEC value. In a perfect scenario, both RMSEC and RMSEP values are expected to be small and close.

3.2.2.1. Classical Least Squares

The classical least squares calibration method is the application of the Beer-Lambert law to linear equation systems consisting of measurement data obtained from spectrophotometric or another analytical instrument. Explanations here are made for spectrophotometric studies.

$$\mathbf{A} = \mathbf{K}\mathrm{x}\mathbf{C} + \mathbf{E_A} \quad \text{Eq. (4)}$$

Where A is the *nxm* matrix of absorbance values, **C** is the *nxl* matrix of concentration values of the calibration set. Here, *l* is the number of parameters that is predicted, *n* is the number of samples in the calibration set and *m* is the number of

wavelengths in a given spectrum. The unknown in this equation is the **K** matrix which is the *lxm* matrix of absorptivity coefficients that relates the absorbance values at *m* number of wavelengths to the concentrations of *l* parameters of the *n* number of calibration samples. The term $\mathbf{E_A}$ stands for the absorbance residuals that are not fitted by the model equation. The least-square solution of **K** is given as Eq.(5)

$$\mathbf{K} = (\mathbf{C}'.\mathbf{C})^{-1}.\mathbf{C}'.\mathbf{A} \qquad \text{Eq. (5)}$$

Here **C**′ is the transpose of **C** matrix and superscript **-1** on the upper right corner of the parenthesis stands for matrix inversion. Once the **K** matrix is obtained concentrations of the components in a given unknown sample can be predicted by Eq.(6) as:

$$\mathbf{c} = (\mathbf{K}.\mathbf{K}')^{-1}.\mathbf{K}.\mathbf{a} \qquad \text{Eq. (6)}$$

Here **c** is the *lx1* vector of component concentrations in a given unknown sample and **a** is an *mx1* matrix of absorbance values from the spectrum of the unknown sample.

3.2.2.2. Inverse Least Squares

In spectroscopy of the system of linear equations, the ILS method involves applying the inverse of the Beer-Lambert law to the system of linear equations:

$$\mathbf{C} = \mathbf{A}.\mathbf{P} + \mathbf{EC} \qquad \text{Eq. (7)}$$

C and **A** are the same as they are in the CLS. **EC** is the matrix of concentration residuals, and **P** is the *mxl* matrix of regression coefficients that relate the absorbance values to the concentrations of the components in the calibration set. **P** is calculated as Eq.(8) using pseudo-inverse.

$$\mathbf{P} = (\mathbf{A}'.\mathbf{A})^{-1}.\mathbf{A}'.\mathbf{C} \qquad \text{Eq. (8)}$$

Unlike CLS, the ILS model has the advantage of modeling one component at a time as given in Eq.(9)

$$\mathbf{p} = (\mathbf{A}'.\mathbf{A})^{-1}.\mathbf{A}'.\mathbf{c} \qquad \text{Eq. (9)}$$

Where **p** is an *mx1* vector of the regression coefficient for the component being modeled and **c** is an *nx1* vector of concentrations of the component model in the calibration set. Once the **p** vector is obtained, the predictions of an unknown sample are determined by using Eq.(10).

$$c = \mathbf{a}.\mathbf{p} \qquad \text{Eq. (10)}$$

Here the concentration of the unknown sample and **a** is an *mx1* matrix of absorbance values from the spectrum of the unknown sample.

3.2.2.3. Partial Least Squares Regression

Because of the way the multicollinearity problem is handled, projection methods are very popular among chemometricians. They also allow for multi-compound predictions. The data is projected to a new space using principal components analysis (PCA) in such a way that there is no correlation between the new variables in this space. This decomposes the original absorbance data matrix **(A)** into two smaller matrices known as scores **(T)** and loadings **(B)**. The variance in independent variables is defined by principal components, which can be described as Eq. (11).

$$\boldsymbol{A} = \boldsymbol{T}.\boldsymbol{B} \qquad \text{Eq. (11)}$$

The variables in PCA projected data are also sorted by how much variance they explain from the original data. This allows, for the most part, to be accounted for by only a few first projected variables (scores) while leaving others out helps to remove instrumental noise. The first few columns of scores can then be used in a regression model with these variables. The loadings, which have the same number of columns as scores, are used to project the new data into this space.

Partially least squares (PLS), like PCA, is a projection method that not only projects absorbances but also concentrations in such a way that their covariance is maximized. PLS also eliminates the multicollinearity problem because the number of latent variables can be adjusted as easily as the number of principal components in PCA. The maximization of the covariance ensures that the projected variables contain the information for response prediction, whereas PCA may not account for the information that explains the responses even if the majority of the variance is explained. PLS assumes that errors are distributed evenly across absorbances and concentrations.

Eqs. (12) and (13) for absorbances and concentrations, respectively, are the PLS model equations.

$$\mathbf{A} = \mathbf{T}.\mathbf{B} + \mathbf{E} \qquad \text{Eq. (12)}$$

$$\mathbf{c} = \mathbf{T}.\mathbf{r} + \mathbf{e_c} \qquad \text{Eq. (13)}$$

A matrix has the same dimensions as CLS and ILS in this case. **T** is an *nxh* score matrix, and **B** is a *hxm* loading matrix derived from PCA decomposition. The E matrix is now different from the one in CLS because it comes from PCA, but the size is the same as the one in CLS. The term **c** is an *nx1* vector of calibration concentrations for the component being modeled, and the term **r** is a *hx1* vector of PLS regression coefficients obtained by iteratively solving Eqs. (12) and (13). Figure 5 shows a schematic diagram of Eq. (12) and Eq. (13).

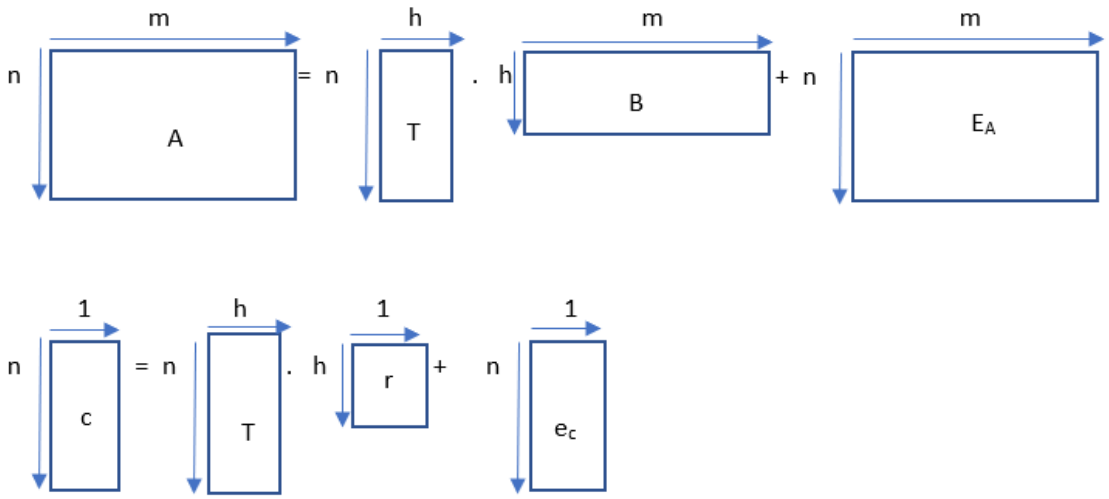


Figure 5: PLS Calibration

A successful model requires the selection of the optimal number of components (Number of PC: h). This number is optimized using the Predicted Residual Error Sum of Squares (PRESS) value. As a general rule, the number of components are chosen before the increment in PRESS is observed or when the PRESS stops decreasing. The addition of extra latent variables may cause modelling of interferences.PRESS can be calculated as Eq.(14).

$$\mathbf{PRESS} = \sum_{i=1}^{n} (\hat{c}_i - \boldsymbol{c_i})^2 \qquad \text{Eq. (14)}$$

Where $\hat{\mathbf{c}}$ is the predicted component concentration. An example for the plot of PRESS values vs. number of principal components is illustrated in Figure 6.

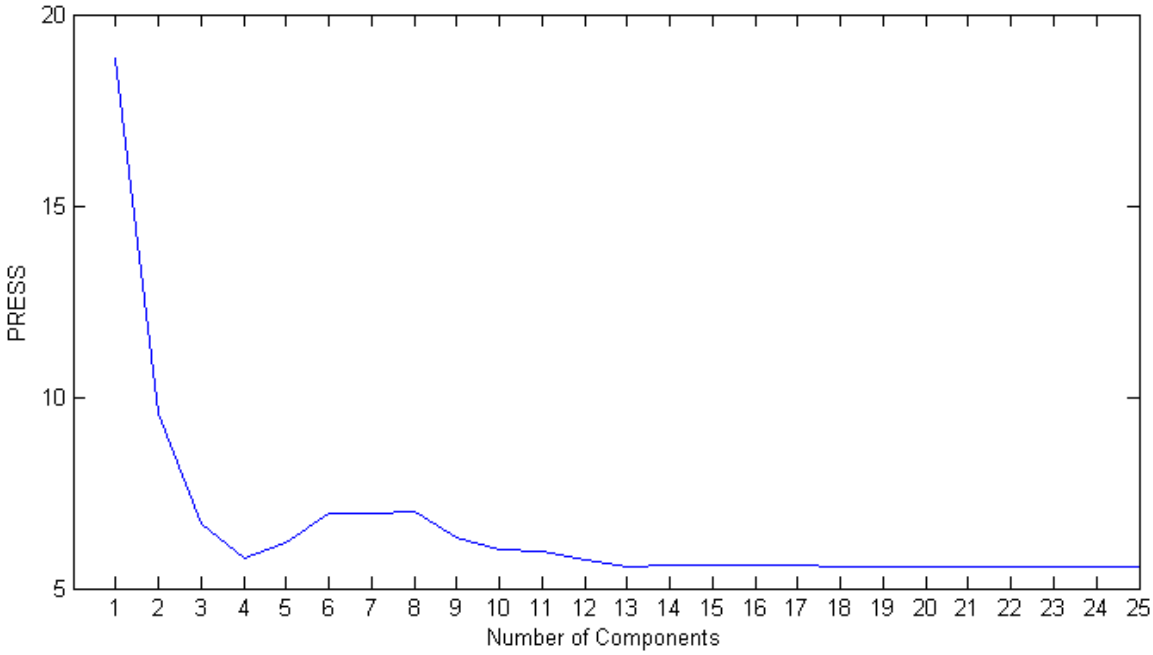


Figure 6: A sample plot of the number of components vs. PRESS values

3.2.2.4. Genetic Inverse Least Squares

Genetic Algorithms (GAs) are a type of research and optimization method based on Darwin's principles of natural evolution and selection. Genetic regression (GR) is a technique for creating linear calibration models that use GA to select wavelengths and mathematical operators. GR is a hybrid calibration method that optimizes simple linear regression models using evolving wavelength selection and simple mathematical operators, combining univariate and multivariate calibration techniques (Özdemir & Öztürk, 2003).

Inverse Least Squares (ILS) and Genetic Algorithms method (GA) are combined in the Genetic Inverse Least Squares (GILS). The purpose of combining the genetic algorithm (GA) with the inverse least squares (ILS) method is to avoid multicollinearity, which can lead to overfitting and wavelength selection issues if the ILS method is used alone without GA.

The GA consists of 5 basic steps, namely initialization of a gene population, evaluation of the population, selection of the parent genes for breeding and mating,

crossover and mutation, and replacing parents with their offspring. These steps are named after the biological basis of the algorithm (Özdemir & Öztürk, 2003).

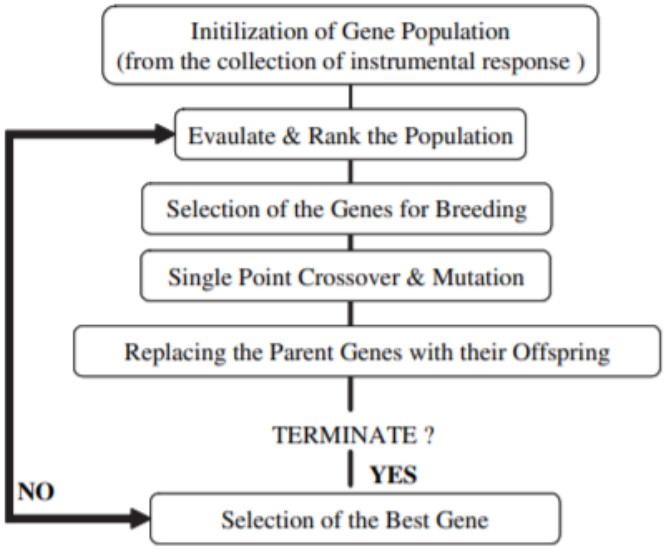


Figure 7: Flow Chart of Typical Genetic Algorithm for Multivariate Calibration (Özdemir & Öztürk, 2006)

The terms used in GILS and their definitions are listed below:

- Gene: Combination of absorbances at specific wavenumbers
- Fitness: Inverse of the standard error of cross-validation
- Cross-over: Exchange of half of the variables between two genes.
- Mutation: Excessive or deficient number of variables in a gene that may result after many iterations.

3.2.2.4.1. Selection of Initial Genes

The first step in GA is to populate a gene pool. A gene is a collection of variables whose counts are randomized within a specific range to minimize bias and maximize the number of possible recombination (Özdemir & Öztürk, 2003). A gene is shown below as an example of spectral data.

G_1: [A_{514}, A_{2585}, A_{855}, A_{941}, A_{1190}, A_{3650}]

G_1 is made up of six variables, and **A** is the variable that represents absorbance at a specific wavelength. A specified number of genes are chosen at random with the condition that their squared Pearson Correlation Coefficient (R^2) is greater than a certain threshold (e.g 0.5). R^2 can be calculated as follows:

$$\mathbf{R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}} \qquad \text{Eq. (15)}$$

Above in Eq.(15), the prediction ŷi is obtained by Cross Validation.

3.2.2.4.2. Evaluation of the Population

A fitness definition of genes is required, as it is in all genetic algorithms. The standard error of cross-validation (SECV) is calculated by comparing the real values and the values obtained from ILS models during cross-validation where the selected variables are contained in the gene.

$$\mathbf{Fitness = \frac{1}{SECV}} \qquad \text{Eq. (16)}$$

And SECV is defined as:

$$\mathbf{SECV = \sqrt{\frac{\sum_{i=1}^{m}(c_i - \hat{c}_i)^2}{n - 2}}} \qquad \text{Eq. (17)}$$

Where **n** is the number of samples, $\boldsymbol{c_i}$ is the property of the sample while $\hat{\mathbf{c}}_\mathbf{i}$ is the prediction. The fitness of each gene is saved and the genes are sorted in ascending order by their finesses.

3.2.2.4.3. Selection of the Parent Genes for Breeding and Mating

While there are a variety of methods for selecting genes for breeding, including top-down and tournament selection, the roulette wheel method is used. This method uses a roulette wheel to assign each gene an area proportional to its fitness. After that, the wheel is spun a number of friable times. The genes selected by this random spinning are then paired top-down, with the first gene paired with the second, the third gene with the fourth gene, and so on. Figure 8 shows the roulette wheel selection method.

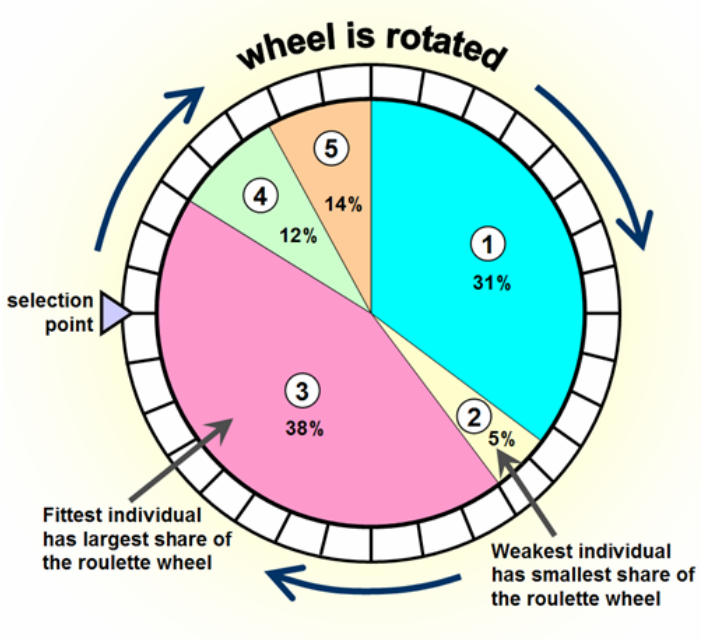


Figure 8: Example of Roulette Wheel selection method

(Reguant, 2021)

3.2.2.4.4. Cross-Over and Mutations

The selected parent gene pairs are cut from the middle in this step to exchange variables and create a new pair of offspring

Parent Gene Pair:

G_1: [A_{514}, A_{2585}, A_{855} # A_{941}, A_{1190}, A_{3650}]

G_1: [A_{677}, A_{2001} # A_{1955}, A_{555}]

Offspring:

$NEWG_1$ = [A_{514}, A_{2585}, A_{855}, A_{1955}, A_{555}]

$NEWG_2$ = [A_{677}, A_{2001}, A_{941}, A_{1190}, A_{3650}]

The # symbol indicates where the genes are broken. The fitness of each gene is determined after each pair of parent genes is subjected to cross-over, as explained in the evaluation section. There is a small chance that a gene will exceed the given number of variable thresholds or become too small to model the data during the cross-over. In this case, the process described in the initialization of the gene population part is used to replace this gene with a new one.

3.2.2.4.5. Replacing Parents with Offspring

The parents are replaced with their offspring after all gene pairs have crossed over. This effectively eliminates unfit genes, allowing fitter genes to take over the population. Even after the parent genes are replaced, the one with the best fitness is always kept and used in the final model. Because of the stochastic nature of breeding, GA can suddenly move away from the minimum.

3.3. Preprocessing Techniques

Data preprocessing is used to make the data to be used more suitable for processing and to obtain more reliable and precise results. Before constructing a calibration model, preprocessing of the data is usually performed to not only address apparent or potential problems but also to improve interpretability.

3.3.1. Mean Centering And Scaling

For numerical variables, it is common to either normalize or standardize your data. Normalization means scaling a dataset so that its minimum is 0 and its maximum is 1. To achieve this we transform each data point **x** to;

$$\mathbf{x_{normalized}} = \mathbf{x} - \boldsymbol{\mu} \qquad \text{Eq. (18)}$$

Standardization is slightly different; its job is to center the data around 0 and to scale concerning the standard deviation:

$$\mathbf{x_{standardized}} = \frac{\mathbf{x} - \mu}{\boldsymbol{\sigma}} \qquad \text{Eq. (19)}$$

where **μ** and **σ** are the mean and standard deviation of the dataset, respectively.

3.3.1. Removing Variables

In many cases, even if a wide variety of variables are obtained, some may be unnecessary. Some parts of the data may contain large and distinctly noisy parts. Removing such unnecessary parts can improve model performance.

3.3.2. Baseline Correction

Detrend is one of the baseline correction methods. A fixed-order polynomial is fitted into the spectrum and then subtracted from the polynomial spectrum (R. J. BARNES, 1989).

The derivative can be used for baseline correction. A fixed baseline offset is obtained by taking the first-degree derivative, and the baseline offset, and baseline slope are subtracted from the spectrum by taking the second-degree derivative. The Savitzky-

Golay algorithm is frequently applied by taking the derivatives of the fit polynomials since derivatization can reduce the signal-to-noise ratio (A. Savitzky, 1964).

There is a certain amount of noise in almost every analytical technique. To remove the noise, the signal is corrected using some algorithms. The Savitzky-Golay algorithm is one of the most frequently used for this purpose. Here, the polynomials are fitted to each of the plurality of data windows, one polynomial to one data window (Temiz, 2019).

CHAPTER 4

EXPERIMENTATION AND INSTRUMENTATION

4.1. Experimentation

4.1.1. Sample Collection And Sample Preparation

In this study, sample collection, reference analyzes, and FTIR spectra of the same samples were performed at TUPRAS Izmir Refinery. 110 commercial gasoline samples were collected. To preserve the stability of the sample, it was stored in the dark and at 4 °C. Reference analyzes were also carried out in accordance with the criteria specified in the international standards of the relevant analysis methods. These methods are specified in TS EN 228 - Automotive Fuels - Unleaded Petrol Requirements and Test Methods – Specification.

4.2. Instrumentation

TUPRAS Izmir Refinery's Quality Control Laboratory performed all reference analyses. The information about the instruments used in the reference analyzes and the FTIR used for the model development are given in detail in the following.

4.2.1. Density

Density measurements were performed using a density meter (DMA 4500M/Anton Paar) by the standard test method TS EN ISO 12185. As a reference temperature, the measurement temperature was set to 15°C. Figure 9 shows the example of an instrument.

Figure 9: Density meter

(Anton Paar)

4.2.2. Octane Numbers – RON and MON

There are two laboratory test methods (TS EN ISO 5164 and TS EN ISO 5163) to measure octane numbers called Research Octane Number (RON) and Motor Octane Number (MON). In accordance with these test methods, Cooperative Fuel Research (CFR) engines (Waukesha, CFR Engines Inc.), were used to test RON and MON capability in the 40-120 octane number range. Variable compression ratio cylinder (4:1 to 18:1) and sleeve assembly, four-bowl falling level carburetor, CFR crankcase, intake air humidity equipment, exhaust surge system, and knock meter are the main components of the CFR engine. Figure 10 shows the example of an instrument.

Figure 10: CFR Engine

(Waukesha CFR)

4.2.3. Distillation Points

The distillation analysis was carried out using atmospheric distillation instruments (OptiDist/PAC) in accordance with the standard test method TS EN ISO 3405 (ASTM D 86). While performing distillation analysis, Sample Temperature was set below 10°C, Temperature of Cooling Bath was set at 0.5°C, Temperature of Bath Around Receiving Cylinder was set at 13°C, and Distillation Rate was set 4.5mL/min. Figure 11 shows the example of an instrument.

Figure 11: Atmospheric Distillation Instrument (PAC)

4.2.4. Hydrocarbon Content

Multidimensional gas chromatography (7890B Reformulyzer/AGILENT) was used to analyze hydrocarbon types, aromatics, olefins, and benzene, according to the TS N ISO 22854 standard test method. The auto-injector, FID detector, capillary/micro packed columns and traps all contribute to a good separation and reduced analysis time. Figure 12 shows the example of an instrument.

Figure 12: Gas Chromatograph

(Agilent, 2018)

4.2.5. Fourier Transform Infrared Spectrometer – FTIR

The FTIR (ANALECTS, Diamond 20) analyzer is a specific instrument manufactured for private use. Figure 13 shows the example of an instrument. It is used for the calibration and application development of the online FTIR analyzer used in the process. It is an exact copy of the FTIR analyzer used in the process, as it is used for calibration. The specifications of the infrared spectrometer used in the analysis are given in Table 2.

Figure 13: FTIR Analyser

(Schneider Electric, 2021)

Table 2 FTIR Specifications

Spectral Range	6000 cm^{-1} - 1000 cm^{-1}
Sample Cell	Fixed cells – 0.1 mm thickness
Detector	Standard DTGS pyroelectric
Interferometer	Patented Transept® Interferometer
Infrared Source	Internal, air-cooled, high-efficiency Reflex Sphere
Resolution	8 cm^{-1}

4.3. Data Analysis

The collected spectra in ASF format were transferred to Microsoft® Excel® in 2016. Then the data analysis is performed by chemometric calibration toolbox 29(OBA Quantifier, OBA Kemometri Inc. Turkey) which is developed in the MATLAB R2018b (Math Works Inc., MA) environment. Genetic Inverse Least Squares (GILS) and Partial Least Squares Regression (PLSR) were performed for this study. In this study, as mentioned before, 110 commercial gasoline samples were used and 11 important parameters of gasoline were tried to be determined. For calibration, 70 gasoline samples, 70x862 absorbance matrix of these samples, and reference measurements were used.

CHAPTER 5

RESULTS AND DISCUSSION

5.1. FTIR Spectra

Fourier Transform Near-Infrared spectra of 110 gasoline samples were collected for 3 months. Figure 14 shows the raw FTIR absorbance spectra, which were recorded in the 6000 cm^{-1} – 1000 cm^{-1} wavenumber region for gasoline samples. This spectral region studied is different from the common range. Between 6000 cm^{-1} – 4000 cm^{-1} is in the near-infrared region, between 4000 cm^{-1} – 1000 cm^{-1} is in the middle infrared region. This region is defined as an "extended mid-IR region" by the device manufacturer.

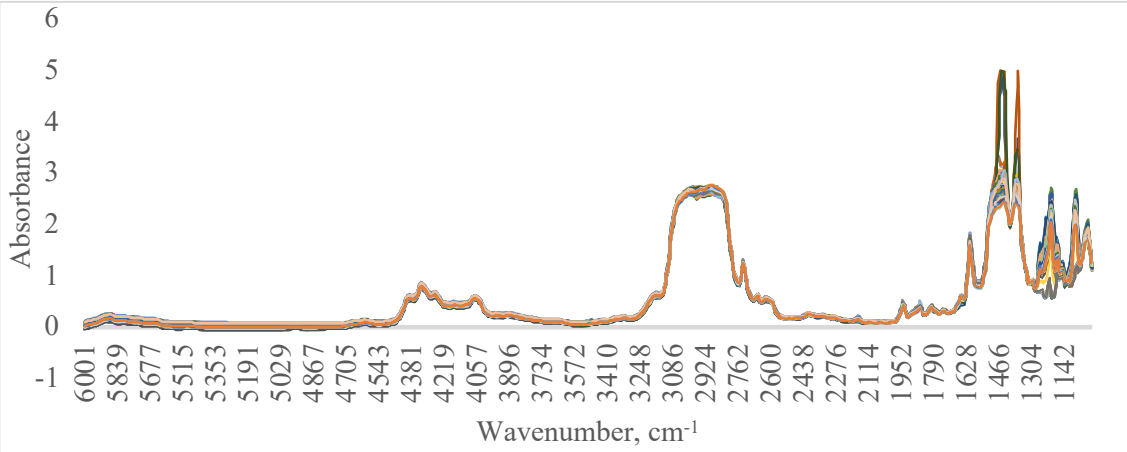


Figure 14: FTIR spectra of 110 Gasoline Sample

Absorbance value between 3150 cm^{-1} - 2670 cm^{-1} and 1600 cm^{-1} – 1000 cm^{-1} spectral region was removed because the absorbance value is found to be greater than 2 which can cause nonlinearity problems. In addition, this removed region shows some noisy futures. In Figure 15, FTIR spectra total of 110 gasoline samples with narrowed intervals are shown.

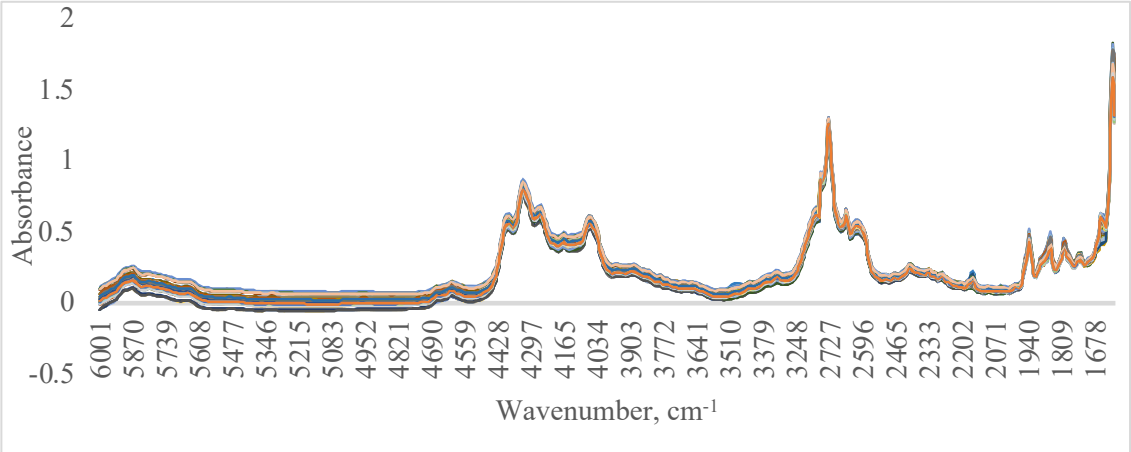


Figure 15: FTIR spectra of 110 Gasoline Samples with Narrowed Range

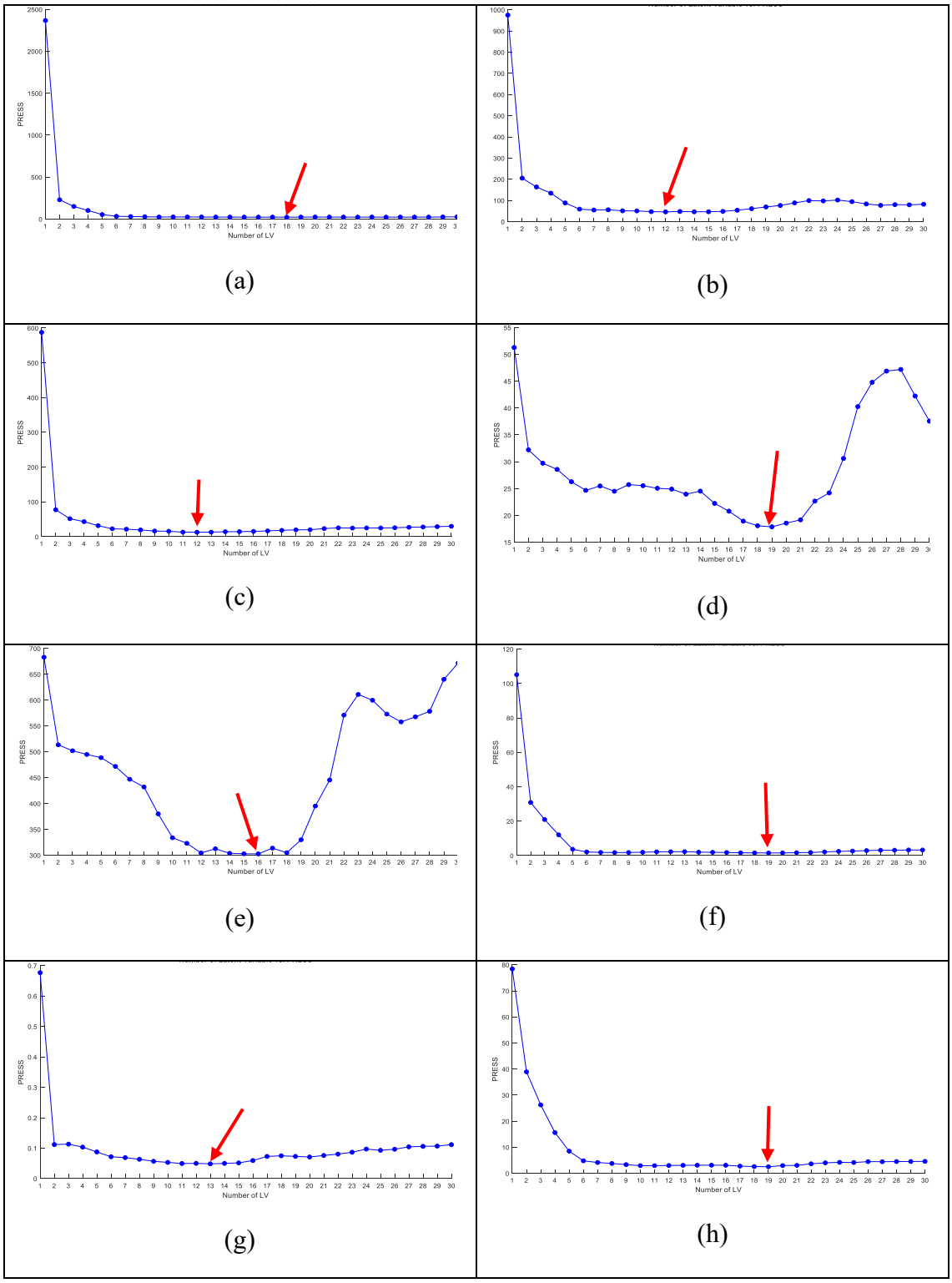
When using the FTIR spectrometer to analyze samples, there is a lot of chemical information that may be used to identify the sample. However, because the overtones in the IR spectra have such a broad structure, it's difficult to link these bands with the chemical bonds in the sample. Furthermore, in spectral analyses, deviations are known as noise frequently occur, which contain no information about the sample and are generated by scattering from the light source, sample cell, or particles of heterogeneous mixtures in the sample during measurement. As a result, without doing any mathematical operations on the spectrum, qualitative and quantitative analysis of the sample will be impossible.

There was no preprocessing was performed in this study, except for the extraction of a small spectral region. For each parameter of the gasoline sample, a model was created with PLS and GILS methods and performance analysis was made. In each model, 70 gasoline samples were used for calibration and 40 gasoline samples for validation.

5.2. Partial Least Squares Regression

For finding the best fitting number of LVs, Predicted residual error sum of squares (PRESS) values were calculated for the first 30 LVs and the results are given in Figure 16. The coefficient of determination (R^2) of the calibration data set, the root mean square error of cross validation (SECV), and the root mean square of validation errors (SEP) data were used to evaluate the prediction performances of the developed model. For the following part, the results are Density, E70, E100, E150, FBP, MTBE content, Benzene

content, Olefin content, Aromatic contents, RON, MON PLS model results will be given for a clear explanation.



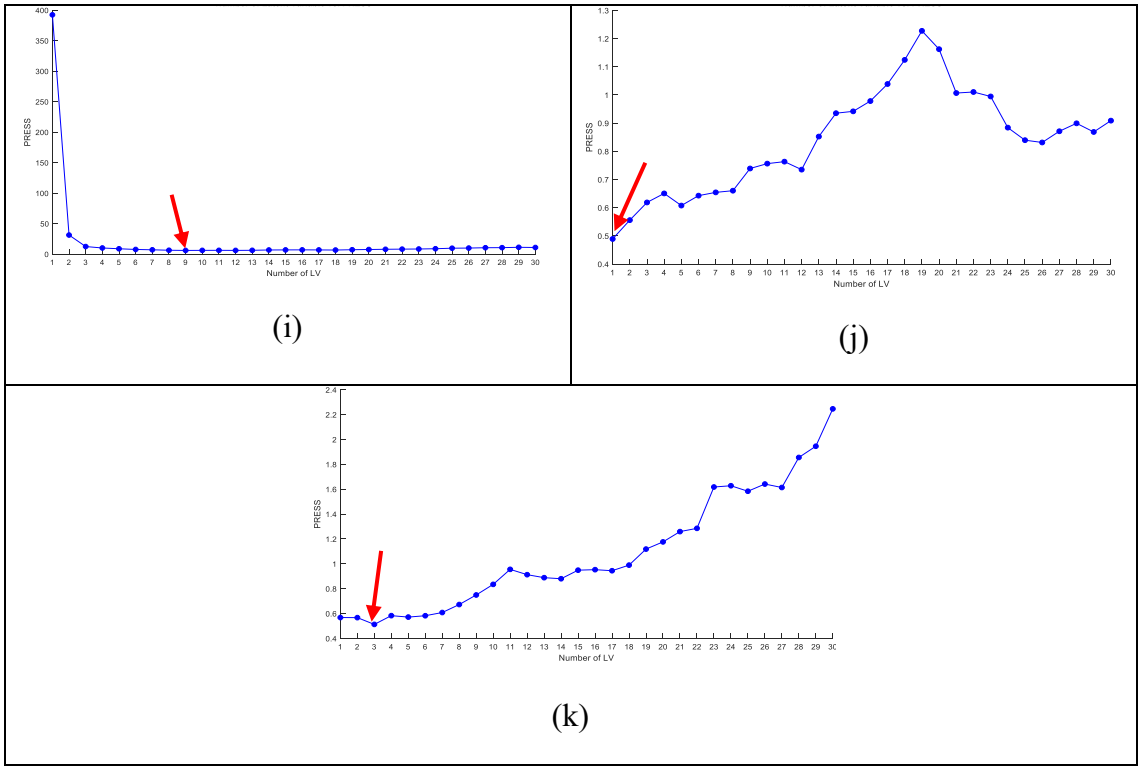


Figure 16: Number of PCs vs. PRESS plot for selecting the optimal number of LVs a) Density, b) E70, c) E100, d) E150, e) FBP, f) MTBE content, g) Benzene content, h) Olefin content, i) Aromatic content, j) RON, k) MON

By using Figure 16, 18 LVs, 12 LVs, 12 LVs, 19 LVs, 16 LVs, 19 LVs, 13 LVs, 19 LVs, 9 LVs, 1 LVs, 3 LVs were selected respectively for Density, E70, E100, E150, FBP, MTBE content, Benzene content, Olefin content, Aromatic content, RON, MON.

5.3. Genetic Inverse Least Square Regression

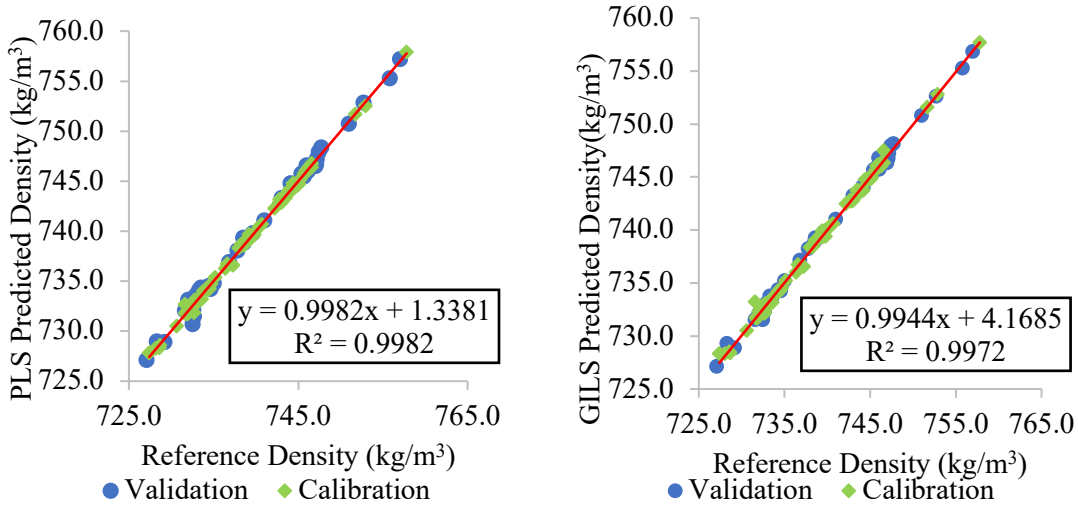
Genetic Inverse Least Square (GILS) was used using 10 genes, 20 iterations, and 8 runs, with an R^2 threshold of 0.5 for initial gene selection and a 1-fold CV for fitness determination. The coefficient of determination (R^2) of the calibration data set, the root mean square error of cross validation (SECV), and the root mean square of validation errors (SEP) data were used to evaluate the prediction performances of the developed model. For the following part, the results are Density, E70, E100, E150, FBP, MTBE content, Benzene content, Olefin content, Aromatic content, RON, MON PLS model results will be given for a clear explanation.

5.4. Multivariate Calibration Results and Comparison

Two calibration methods (PLS and GILS) were used and compared for the determination of each parameter. The results are given in detail for each parameter separately.

5.4.1. Density

Reference values obtained from density meter (DMA 4500M/Anton Paar) vs model predicted values (PLS and GILS) of FTIR spectra of gasoline sample are given in Figure 17 for Density parameter.



(a) **(b)**

Figure 17: Reference Density Value vs Predicted Density Value; a) PLS, B)GILS

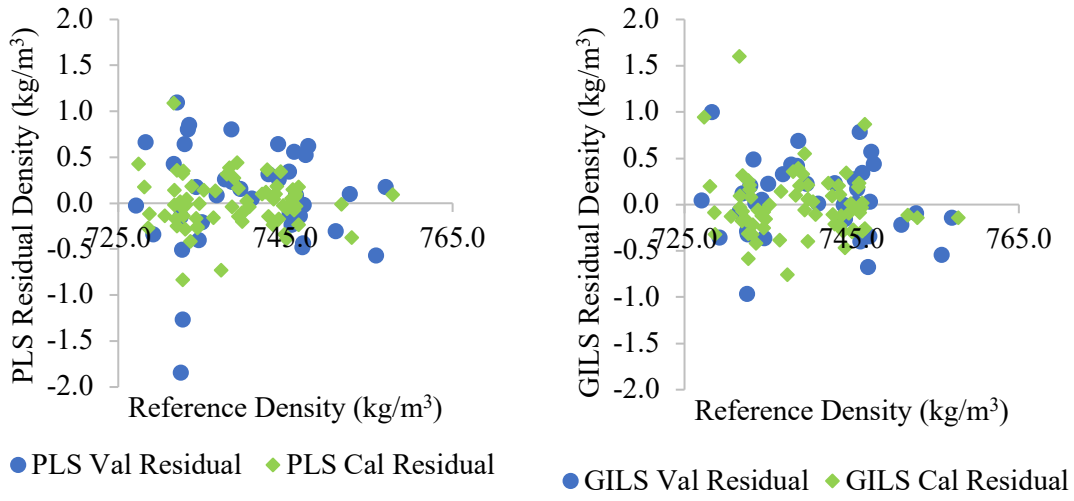
As can be seen in Figure 17, the calibration predictions and validation set predictions of both models (PILS, GILS) are quite close. For PLS model, SECV and SEP values are found to be 0.2803 (kg/m^3) and 0.5700 (kg/m^3), respectively. The R^2 value for calibration set predictions are calculated as 0.9982, and the R^2 value for the validation set is 0.9948. For GILS model, SECV and SEP values are found to be 0.3518 (kg/m^3) and

0.4067 (kg/m^3), respectively. The R^2 value for calibration set predictions are calculated as 0.9972, and the R^2 value for the validation set is 0.9973. Prediction performances of the developed model for Density parameters are given in Table 3.

Table 3: Calibration Model Performance for Density

PLS Calibration Model Results		**GILS Calibration Model Results**		**Data Range (kg/m^3)**	
SECV (kg/m^3)	0.2803	SECV (kg/m^3)	0.3518	Max	757.8
SEP (kg/m^3)	0.5700	SEP (kg/m^3)	0.4067	Min	727.4
R^2 calibration	0.9982	R^2 calibration	0.9972	Interval	30.4
R^2 validation	0.9948	R^2 validation	0.9973		
Number of LVs	18				

To determine the error range and possible residual trends, the residuals for both calibration and validation samples are given in Figure 18. All validation data are very close to calibration data which shows model prediction efficiency.



(a) **(b)**

Figure 18: Reference Density vs. Corresponding Model Prediction Density Residuals; a) PLS, b)GILS

As mentioned earlier, the determination of gasoline parameters is made in accordance with certain international standards. Measurement performance must meet the

quality and reliability criteria specified in these standards. For this reason, the reproducibility values in the standards to which the parameters are subject are accepted as limits and the absolute differences (ranges) between the model predictions and reference analysis are examined.

The Reproducibility, R, for Density given in TS EN ISO 12185 is described as 0.5 (kg/m^3). The R-chart with absolute differences between GILS and PLS model predictions and reference analysis is given in Figure 19.

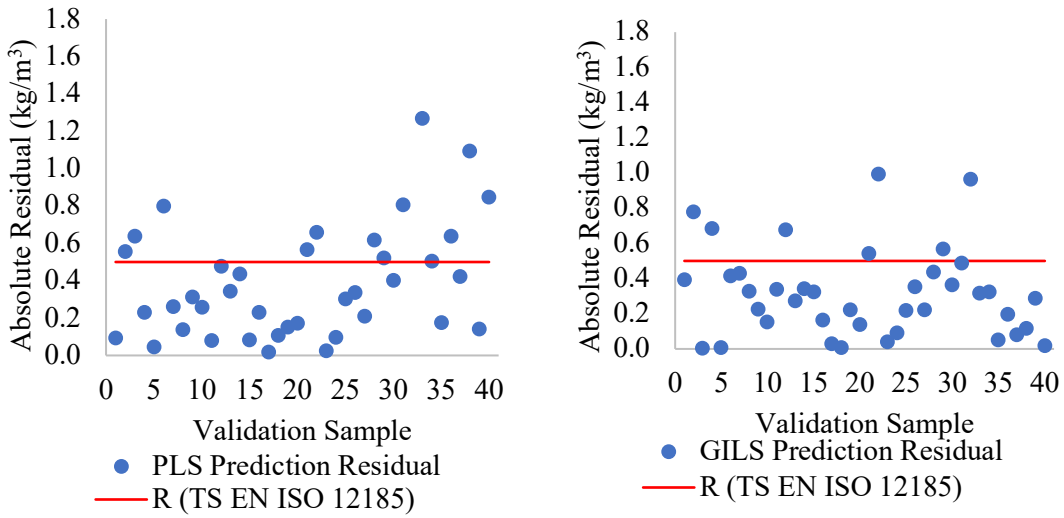


Figure 19: Density Residual R-Chart; a)PLS, b)GILS

As seen in Figure 19, 7 of the GILS prediction values were out of the limit, while 14 of the PLS prediction values were out of the limit. Although the model performance criteria (SEP, SEC, R^2) gave successful results, the PLS model gave 35% and the GILS model 17.5% results out of the limit for 40 samples.

Paired t-test was performed to evaluate whether there was a significant difference between both the PLS and GILS model predicted values and between the reference measurements of each. The results are given in Table 3.

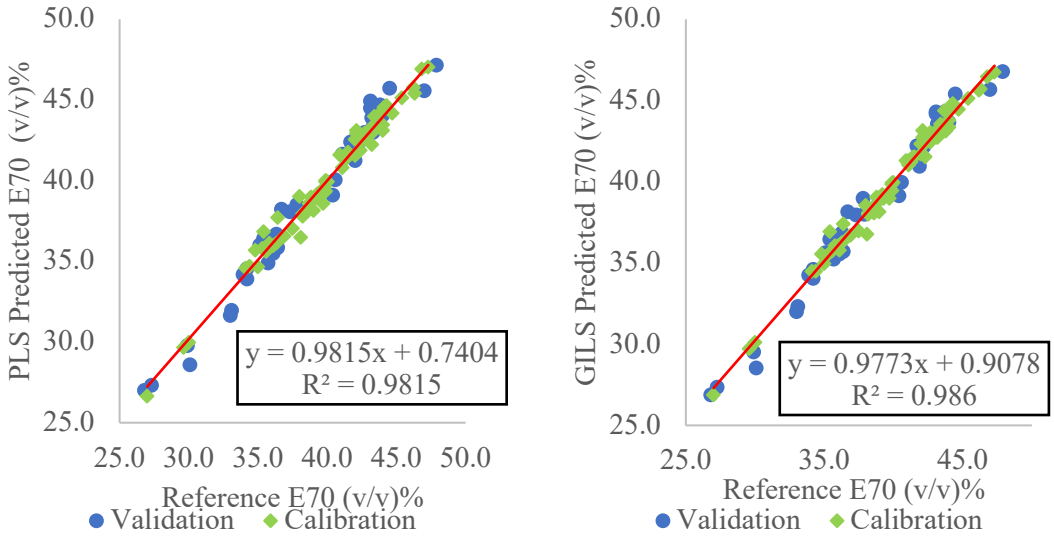
Table 4: Paired t-test Results for Density

	Reference-PLS Comparison	Reference-GILS Comparison	PLS-GILS Comparison
Observations	40		
df	39		
t Stat	-0.762	-0.573	0.512
p-value two-tail	0.451	0.570	0.611
t Critical two-tail	2.023		

According to Table 4, it can be said that, with 95% confidence, there is no statistically significant difference between reference density values and model prediction, nor between GILS and PLS prediction.

5.4.2. E70 - Evaporated at 70 °C

Reference values obtained from distillation instruments (OptiDist/PAC) vs model predicted values (PLS and GILS) of FTIR spectra of gasoline sample are given in Figure 20 for the E70 parameter.



(a) **(b)**

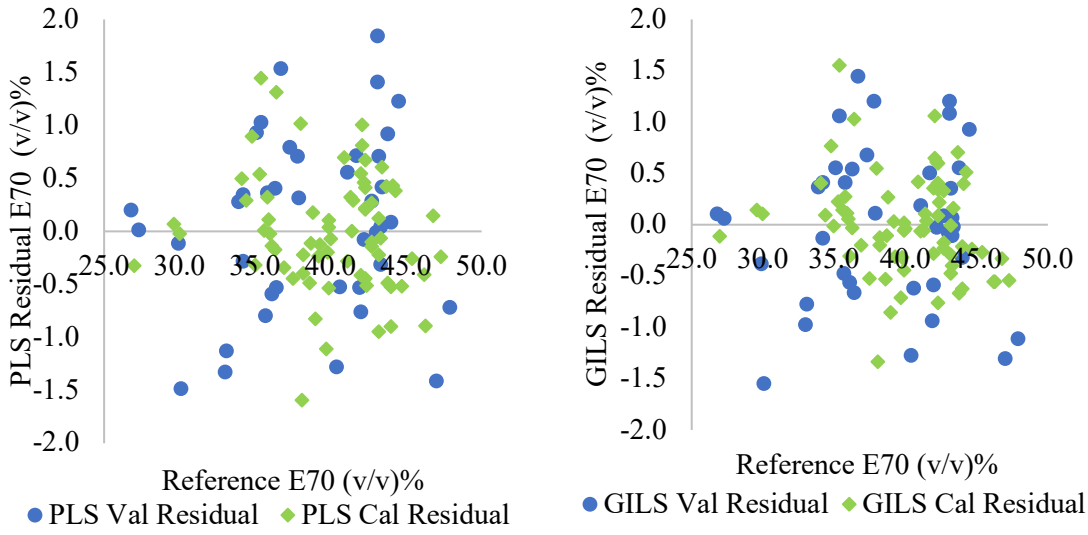
Figure 20: Reference E70 Value vs Predicted E70 Value; a) PLS, B)GILS

As can be seen in Figure 20, the calibration predictions and validation set predictions of both models (PILS, GILS) are quite close. For PLS model, SECV and SEP values are found to be 0.5589 (v/v%) and 0.8380 (v/v%), respectively. The R^2 value for calibration set predictions are calculated as 0.9815, and the R^2 value for the validation set is 0.9768. For GILS model, SECV and SEP values are found to be 0.4884 (v/v%) and 0.7476 (v/v%) respectively. The R^2 value for calibration set predictions are calculated as 0.9860, and the R^2 value for the validation set is 0.9802. Prediction performances of the developed model for the E70 parameter are given in Table 5.

Table 5: Calibration Model Performance for E70

PLS Calibration Model Results		**GILS Calibration Model Results**		**Data Range (v/v %)**	
SECV (v/v %)	0.5589	SECV (v/v %)	0.4884	Max	47.3
SEP (v/v %)	0.8380	SEP (v/v %)	0.7476	Min	27.0
R^2 calibration	0.9815	R^2 calibration	0.9860	Interval	20.3
R^2 validation	0.9768	R^2 validation	0.9802	-	
Number of LVs	12	-		-	

In order to determine the error range and possible residual trends, the residuals for both calibration and validation samples are given in Figure 21. All validation data are very close to calibration data which shows model prediction efficiency.



(a) **(b)**

Figure 21: Reference E70 vs. Corresponding Model Prediction E70 Residuals; a) PLS, b)GILS

The Reproducibility, R, for E70 given in ASTM D 86 is described as Eq. (20). The R-chart with absolute differences between GILS and PLS model predictions and reference analysis is given in Figure 22.

$$\pm \mathbf{0.02} * (\mathbf{150} - \mathbf{X}) \qquad \text{Eq. (20)}$$

Where X is the measured value.

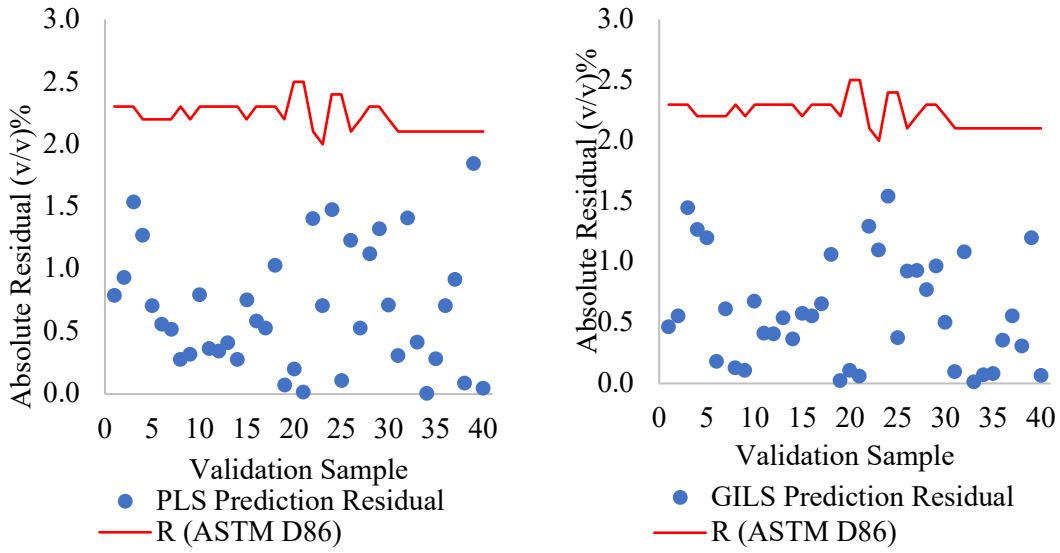


Figure 22: E70 Residual R-Chart; a)PLS, b)GILS

As seen in Figure 22, there is no out-of-limit predicted value for 40 samples.

Paired t-test was performed to evaluate whether there was a significant difference between both the PLS and GILS model predicted values and between the reference measurements of each. The results are given in Table 6.

Table 6: Paired t-test Results for E70

	Reference-PLS Comparison	**Reference-GILS Comparison**	**PLS-GILS Comparison**
Observations	40		
df	39		
t Stat	-0.642	-0.031	1.995
p-value two-tail	0.524	0.975	0.053
t Critical two-tail	2.023		

According to Table 6, it can be said that, with 95% confidence, there is no statistically significant difference between reference E70 values and model prediction, nor between GILS and PLS prediction.

5.4.3. E100 - Evaporated at 100 °C

Reference values obtained from distillation instruments (OptiDist/PAC) vs model predicted values (PLS and GILS) of FTIR spectra of gasoline sample are given in Figure 23 for the E100 parameter.

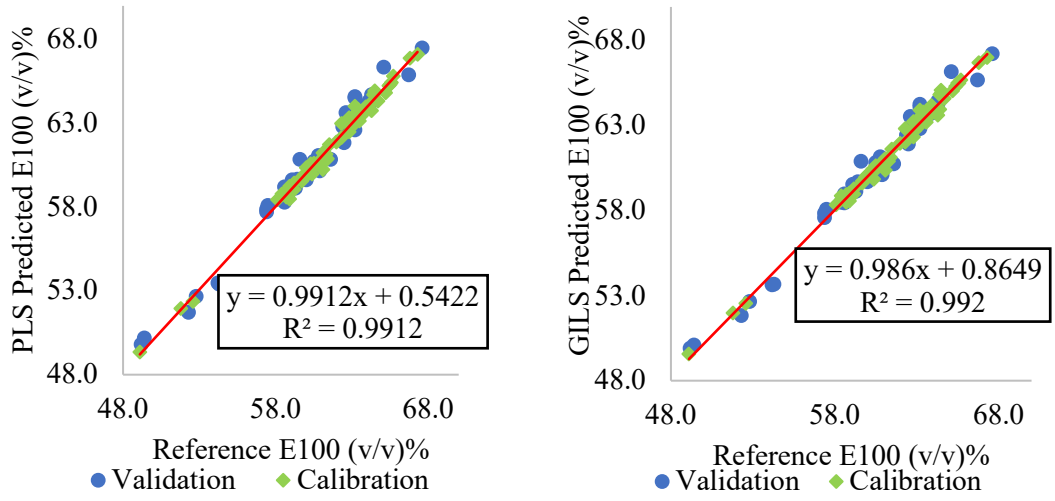


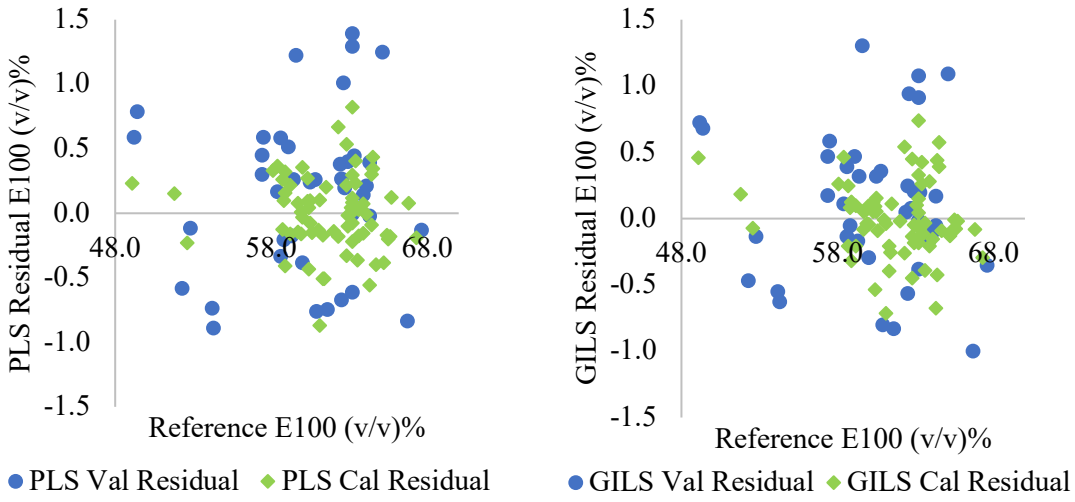
Figure 23: Reference E100 Value vs Predicted E100 Value; a) PLS, B)GILS

As can be seen in Figure 23, the calibration predictions and validation set predictions of both models (PILS, GILS) are quite close. For PLS model, SECV and SEP values are found to be 0.5589 (v/v%) and 0.8380 (v/v%), respectively. The R^2 value for calibration set predictions are calculated as 0.9815, and the R^2 value for the validation set is 0.9768. For GILS model, SECV and SEP values are found to be 0.4884 (v/v%) and 0.7476 (v/v%) respectively. The R^2 value for calibration set predictions are calculated as 0.9860, and the R^2 value for the validation set is 0.9802. Prediction performances of the developed model for the E100 parameter are given in Table 7.

Table 7: Calibration Model Performance for E100

PLS Calibration Model Results		GILS Calibration Model Results		Data Range (v/v %)	
SECV (v/v %)	0.2989	SECV (v/v %)	0.2856	Max	67.3
SEP (v/v %)	0.6326	SEP (v/v %)	0.5606	Min	49.1
R^2 calibration	0.9912	R^2 calibration	0.9920	Interval	18.2
R^2 validation	0.9807	R^2 validation	0.9838	-	
Number of LVs	12	-		-	

In order to determine the error range and possible residual trends, the residuals for both calibration and validation samples are given in Figure 24. All validation data are very close to calibration data which shows model prediction efficiency.



(a) **(b)**

Figure 24: Reference E100 vs. Corresponding Model Prediction E100 Residuals; a) PLS, b)GILS

The Reproducibility, R, for E100 given in ASTM D 86 is described as Eq. (20). The R-chart with absolute differences between GILS and PLS model predictions and reference analysis is given in Figure 25.

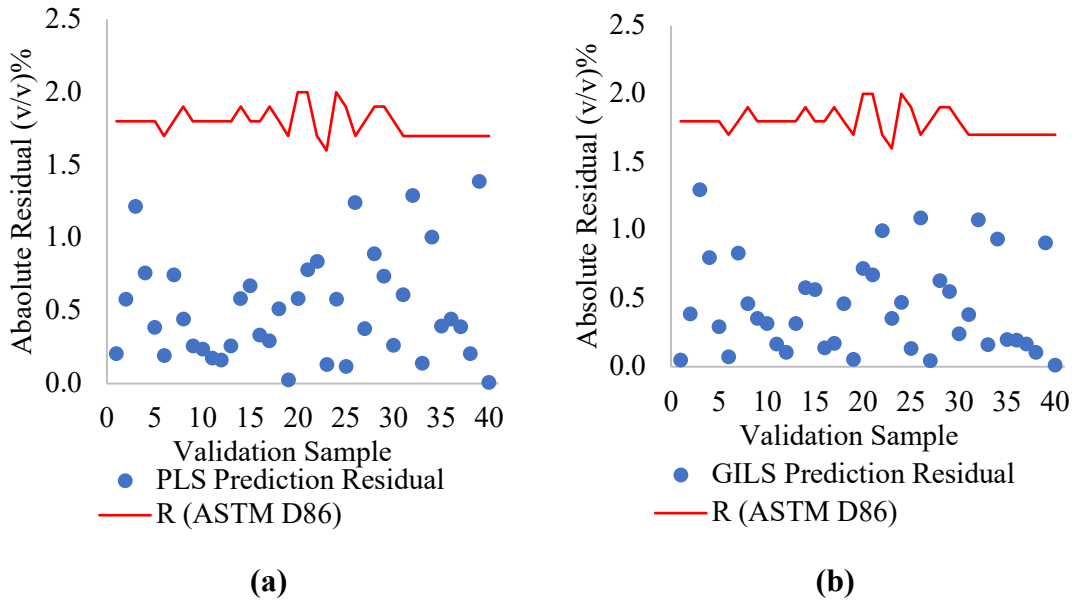


Figure 25: E100 Residual R-Chart; a)PLS, b)GILS

As seen in Figure 25, there is no out-of-limit predicted value for 40 samples.

Paired t-test was performed to evaluate whether there was a significant difference between both the PLS and GILS model predicted values and between the reference measurements of each. The results are given in Table 8.

Table 8: Paired t-test Results for E100

	Reference-PLS Comparison	**Reference-GILS Comparison**	**PLS-GILS Comparison**
Observations	40		
df	39		
t Stat	-1.555	-1.171	1.820
p-value two-tail	0.128	0.248	0.076
t Critical two-tail	2.023		

According to Table 8, it can be said that, with 95% confidence, there is no statistically significant difference between reference E100 values and model prediction, nor between GILS and PLS prediction.

5.4.4. E150 - Evaporated at 150 °C

Reference values obtained from distillation instruments (OptiDist/PAC) vs model predicted values (PLS and GILS) of FTIR spectra of gasoline sample are given in Figure 26 for the E100 parameter.

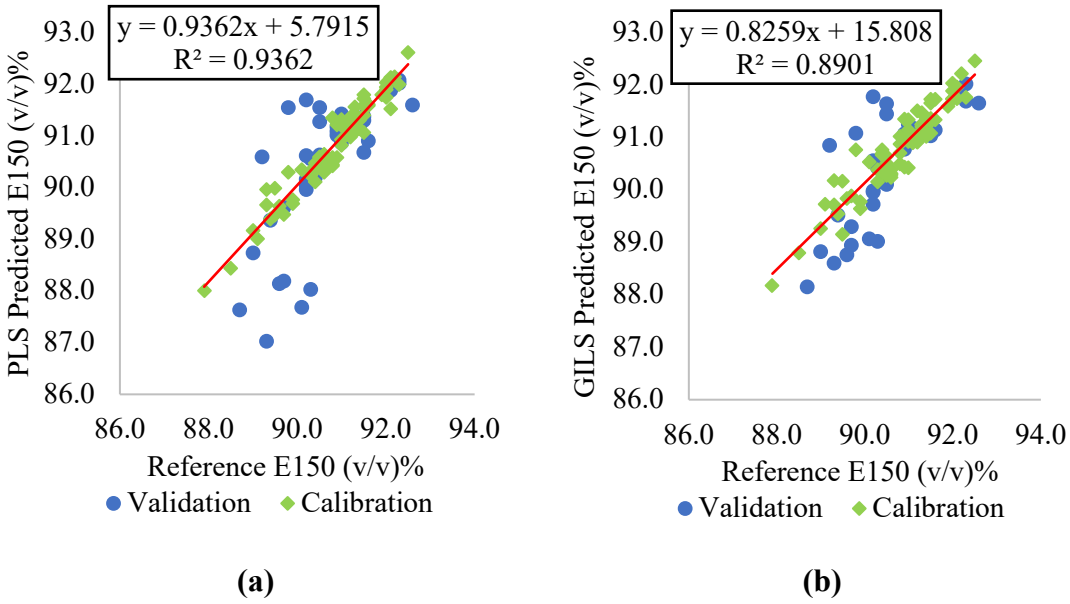


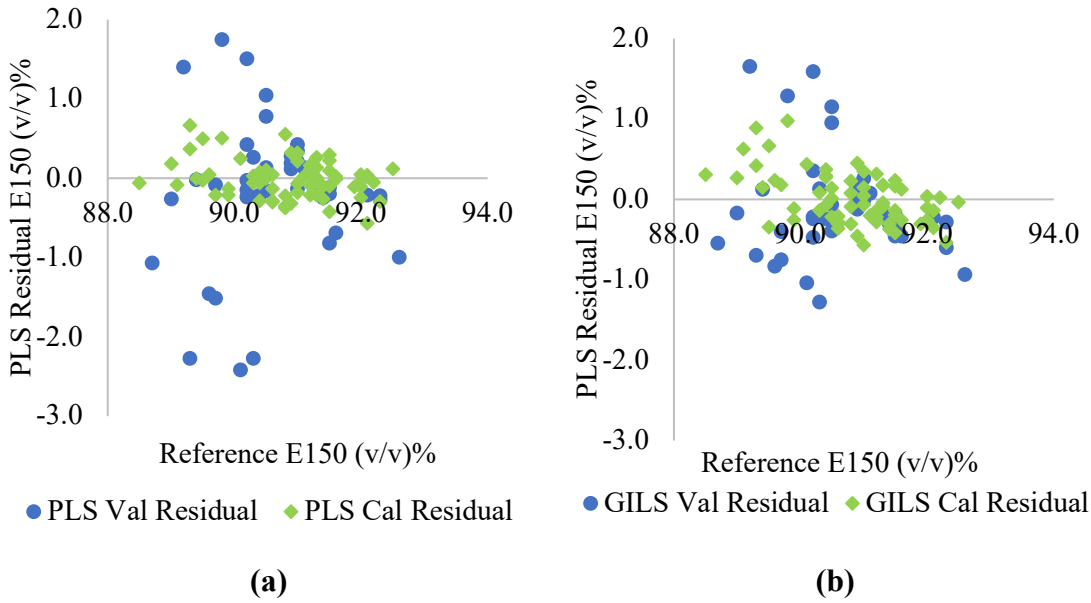
Figure 26: Reference E150 Value vs Predicted E150 Value; a) PLS, B)GILS

As can be seen in Figure 26, both models (PILS, GILS) have differences in validation and calibration predictions. When the R^2 values of the validation and calibration series in the table are examined, it is observed that there are more differences in the validation prediction. For PLS model, SECV and SEP values are found to be 0.2400 (v/v%) and 0.9358 (v/v%), respectively. The R^2 value for calibration set predictions are calculated as 0.9362, and the R^2 value for the validation set is 0.5292. For GILS model, SECV and SEP values are found to be 0.3216 (v/v%) and 0.6604 (v/v%) respectively. The R^2 value for calibration set predictions are calculated as 0.8901, and the R^2 value for the validation set is 0.6132. Prediction performances of the developed models for the E150 parameter are given in Table 9. It is seen that R^2 of the calibration of the PLS model, is better, while R^2 of the validation of the GILS model is better.

Table 9: Calibration Model Performance for E150

PLS Calibration Model Results		**GILS Calibration Model Results**		**Data Range (v/v %)**	
SECV (v/v %)	0.2400	SECV (v/v %)	0.3216	Max	92.5
SEP (v/v %)	0.9358	SEP (v/v %)	0.6604	Min	87.9
R^2 calibration	0.9362	R^2 calibration	0.8901	Interval	4.6
R^2 validation	0.5292	R^2 validation	0.6132	-	
Number of LVs	19	-		-	

In order to determine the error range and possible residual trends, the residuals for both calibration and validation samples are given in Figure 27. Some validation data is far from calibration data. This may indicate that the model performance is not good enough. As seen in Figure 28, there are also values outside the reliability limits specified in ASTM D86.



(a) **(b)**

Figure 27: Reference E150 vs. Corresponding Model Prediction E150 Residuals; a) PLS, b)GILS

The Reproducibility, R, for E150 given in ASTM D 86 is described as Eq. (20). The R-chart with absolute differences between GILS and PLS model predictions and reference analysis is given in Figure 28.

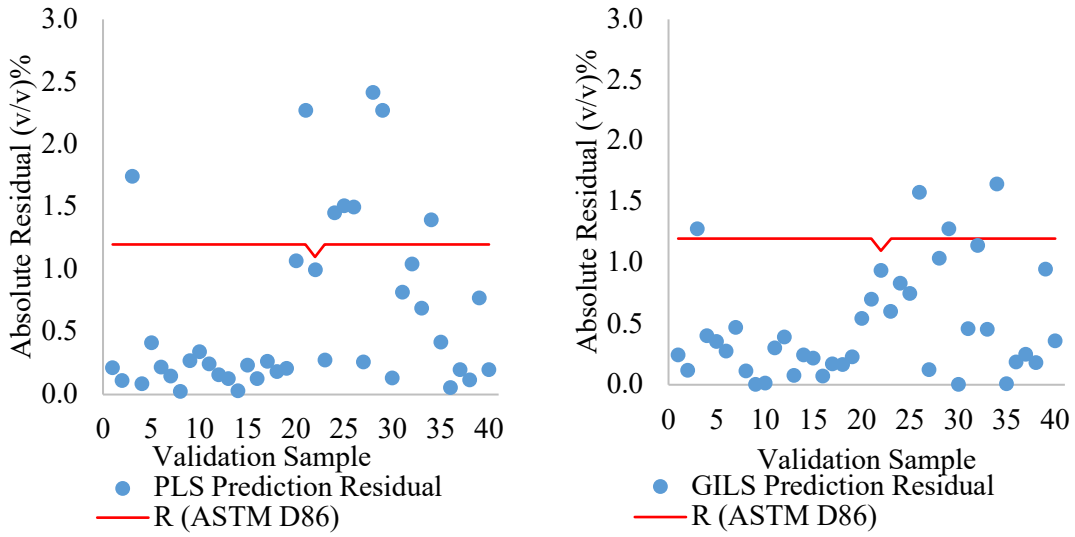


Figure 28: E150 Residual R-Chart; a)PLS, b)GILS

As seen in Figure 28, 4 of the GILS prediction values were out of the limit, while 8 of the PLS prediction values were out of the limit.

Paired t-test was performed to evaluate whether there was a significant difference between both the PLS and GILS model predicted values and between the reference measurements of each. The results are given in Table 10.

Table 10: Paired t-test Results for E150

	Reference-PLS Comparison	**Reference-GILS Comparison**	**PLS-GILS Comparison**
Observations	40		
df	39		
t Stat	1.245	0.927	-1.196
p-value two-tail	0.221	0.359	0.239
t Critical two-tail	2.023		

According to Table 10, it can be said that, with 95% confidence, there is no statistically significant difference between reference E150 values and model prediction, nor between GILS and PLS prediction.

5.4.5. FBP – Final Boiling Point

Reference values obtained from distillation instruments (OptiDist/PAC) vs model predicted values (PLS and GILS) of FTIR spectra of gasoline sample are given in Figure 29 for FBP parameter.

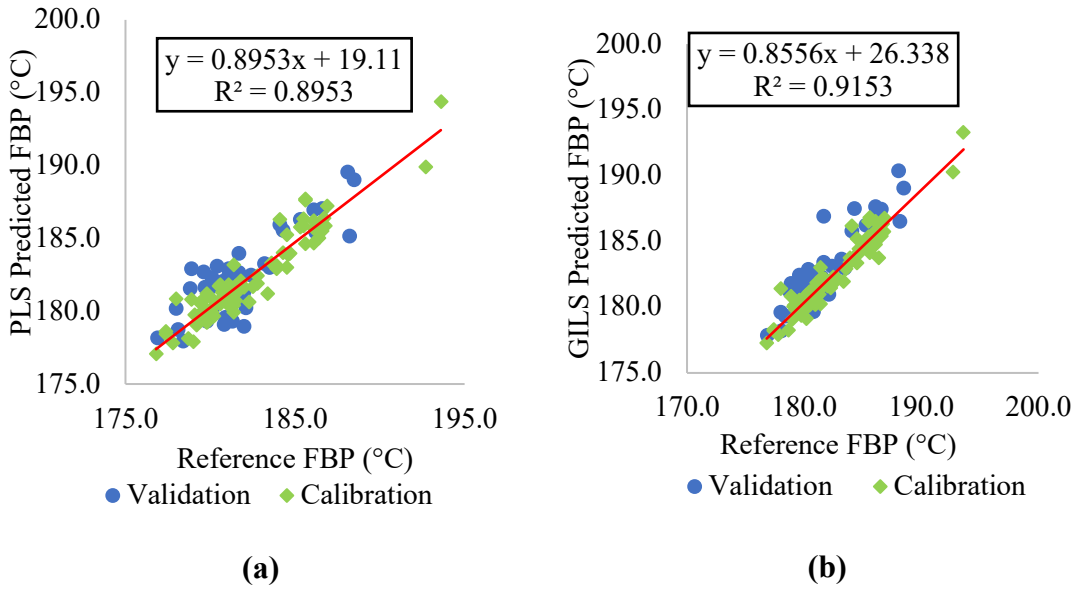


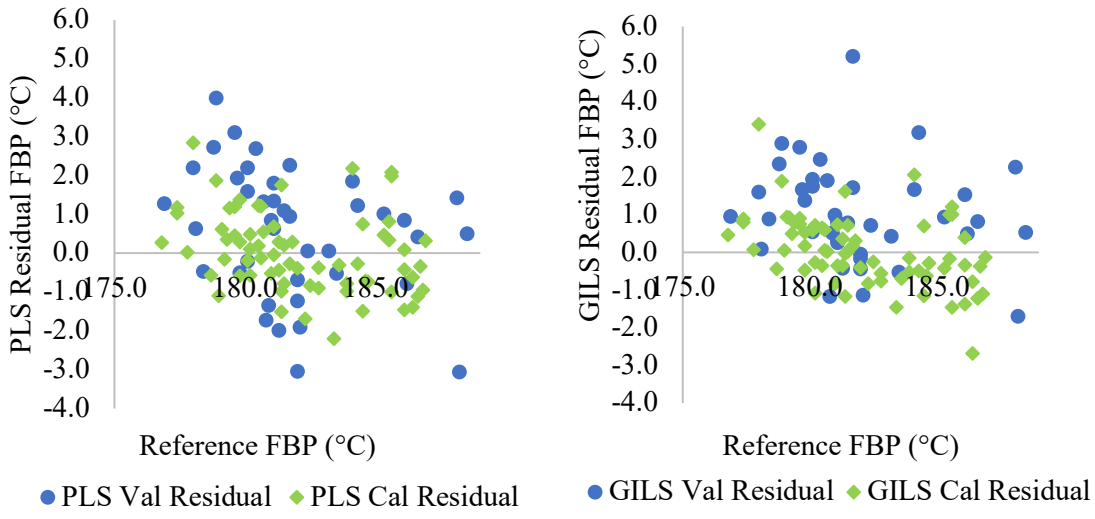
Figure 29: Reference FBP Value vs Predicted FBP Value; a) PLS, B)GILS

As can be seen in Figure 29, the calibration predictions and validation set predictions of both models (PILS, GILS) are quite close. For the PLS model, SECV and SEP values are found to be 1.0792 (°C) and 1.7317 (°C), respectively. The R^2 value for calibration set predictions are calculated as 0.8953, and the R^2 value for the validation set is 0.7031. For the GILS model, SECV and SEP values are found to be 0.9926 (°C) and 1.6801 (°C), respectively. The R^2 value for calibration set predictions are calculated as 0.9153, and the R^2 value for the validation set is 0.8065. Prediction performances of the developed model for the FBP parameter are given in Table 11.

Table 11: Calibration Model Performance for FBP

PLS Calibration Model Results		**GILS Calibration Model Results**		**Data Range (°C)**	
SECV (°C)	1.0792	SECV (°C)	0.9926	Max	193.6
SEP (°C)	1.7317	SEP (°C)	1.6801	Min	176.8
R^2 calibration	0.8953	R^2 calibration	0.9153	Interval	16.8
R^2 validation	0.7031	R^2 validation	0.8065	-	
Number of LVs	16	-		-	

In order to determine the error range and possible residual trends, the residuals for both calibration and validation samples are given in Figure 30. As seen in Figure 30, some validation residuals are not close to calibration residual. However, both SEP and SEC values meet the deviation criteria of the reference measurement method.



(a) **(b)**

Figure 30: Reference FBP vs. Corresponding Model Prediction FBP Residuals; a) PLS, b)GILS

The Reproducibility, R, for FBP given in ASTM D 86 is described as 6.8 (°C). The R-chart with absolute differences between GILS and PLS model predictions and reference analysis is given in Figure 31.

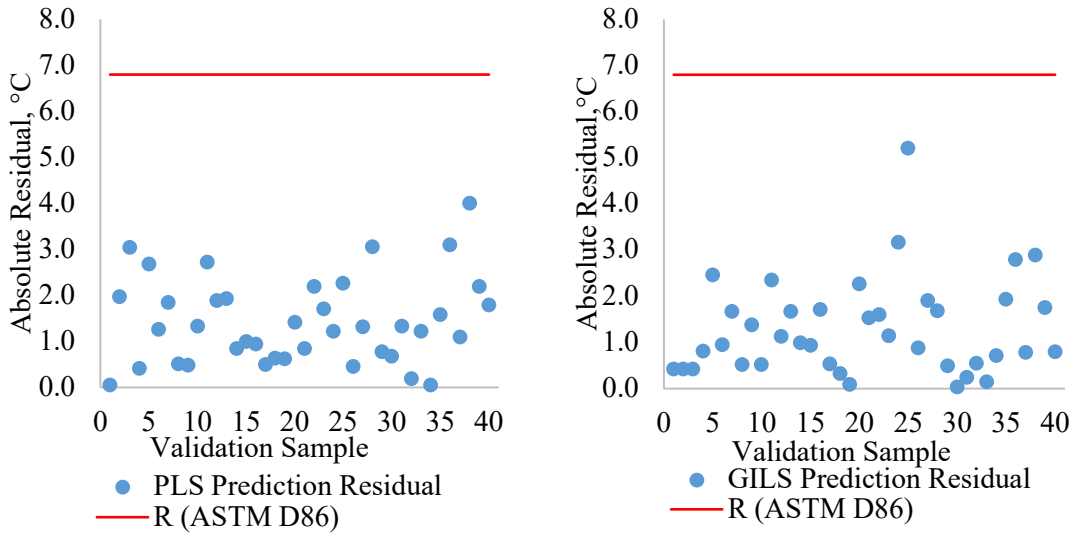


Figure 31: FBP Residual R-Chart; a)PLS, b)GILS

As seen in Figure 31, there is no out-of-limit predicted value for 40 samples.

Paired t-test was performed to evaluate whether there was a significant difference between both the PLS and GILS model predicted values and between the reference measurements of each. The results are given in Table 12.

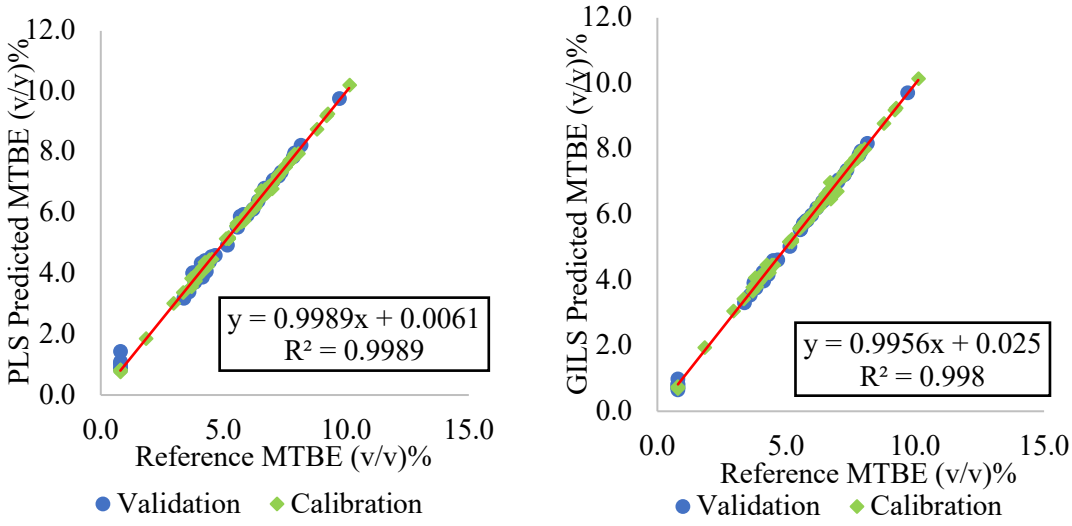
Table 12: Paired t-test Results for FBP

	Reference-PLS Comparison	**Reference-GILS Comparison**	**PLS-GILS Comparison**
Observations	40		
df	39		
t Stat	**-2.192**	**-4.909**	**-3.000**
p-value two-tail	**0.034**	**0.00001**	**0.005**
t Critical two-tail	**2.023**		

According to Table 12, it can be said that, with 95% confidence, there is a statistically significant difference between reference FBP values and model predictions, and between GILS and PLS prediction.

5.4.6. MTBE Content - Methyl Tertiary Butyl Ether

Reference values obtained from multidimensional gas chromatography (7890B Reformulyzer/AGILENT) vs model predicted values (PLS and GILS) of FTIR spectra of gasoline sample are given in Figure 32 for MTBE content.



(a) (b)

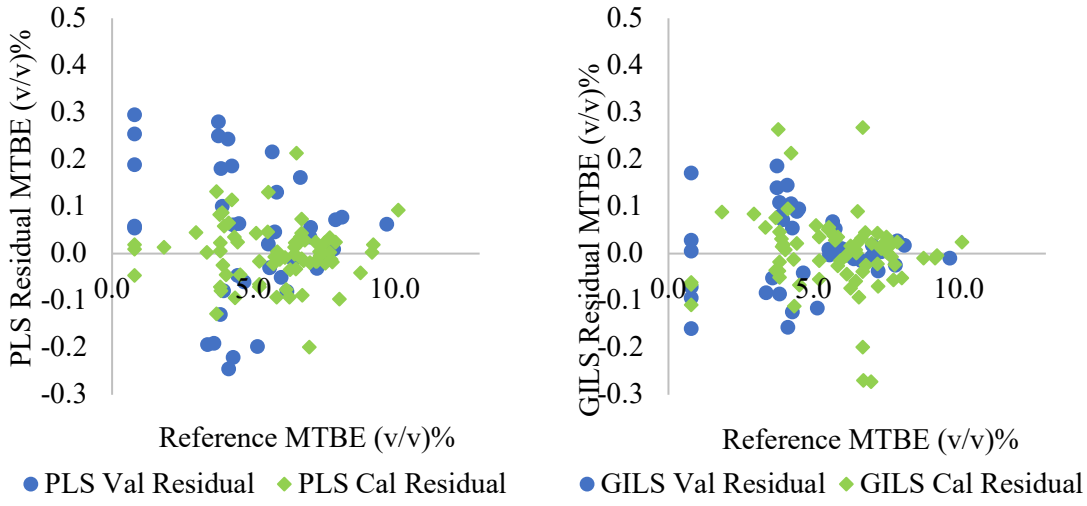
Figure 32: Reference MTBE Value vs Predicted MTBE Value; a) PLS, B)GILS

As can be seen in Figure 32, the performance of both models (PILS, GILS) is quite close to the calibration and validation set predictions. For PLS model, SECV and SEP values are found to be 0.0648 (v/v%) and 0.1791 (v/v%), respectively. The R^2 value for calibration set predictions are calculated as 0.9989, and the R^2 value for the validation set is 0.9946. For GILS model, SECV and SEP values are found to be 0.0876 (v/v%) and 0.0852 (v/v%) respectively. The R^2 value for calibration set predictions are calculated as 0.9980, and the R^2 value for the validation set is 0.9986. Prediction performances of the developed model for MTBE content are given in Table 13.

Table 13: Calibration Model Performance for MTBE Content

PLS Calibration Model Results		**GILS Calibration Model Results**		**Data Range (v/v %)**	
SECV (v/v %)	0.0648	SECV (v/v %)	0.0876	Max	10.12
SEP (v/v %)	0.1791	SEP (v/v %)	0.0852	Min	0.8
R^2 calibration	0.9989	R^2 calibration	0.9980	Interval	9.32
R^2 validation	0.9946	R^2 validation	0.9986	-	
Number of LVs	19	-		-	

In order to determine the error range and possible residual trends, the residuals for both calibration and validation samples are given in Figure 33. All validation data are very close to calibration data which shows model prediction efficiency.



(a) **(b)**

Figure 33: Reference MTBE vs. Corresponding Model Prediction MTBE Residuals; a) PLS, b)GILS

The Reproducibility, R, for MTBE content given in TS EN ISO 22854 is described as Eq. (21). The R-chart with absolute differences between GILS and PLS model predictions and reference analysis is given in Figure 34.

$$\pm \mathbf{0.0251Y + 0.3515} \qquad \text{Eq. (21)}$$

Where Y is the mean of the two results being compared.

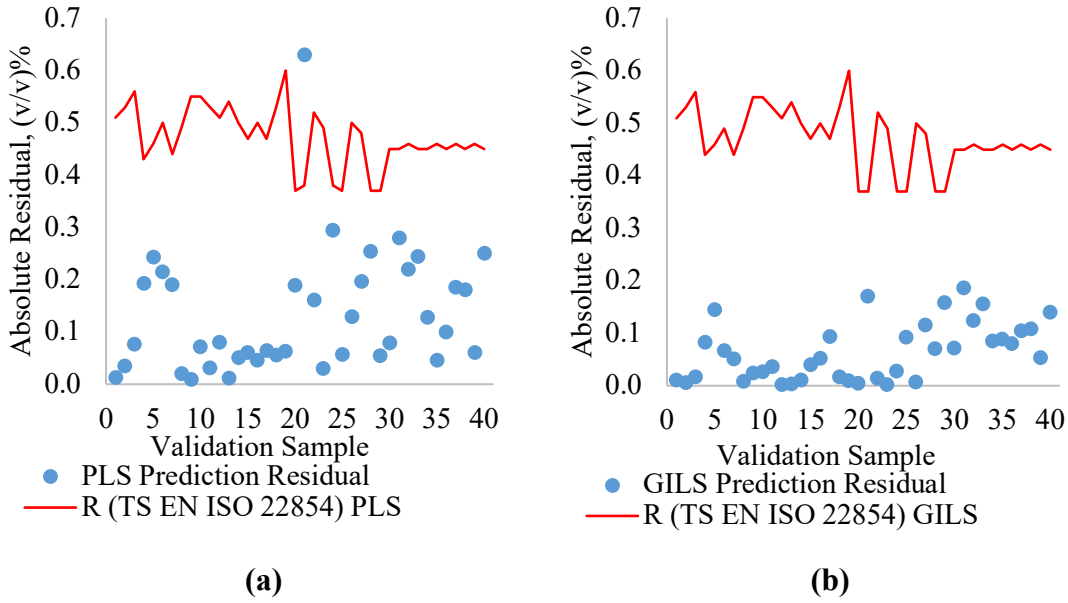


Figure 34: MTBE Residual R-Chart; a)PLS, b)GILS

As seen in Figure 34, there is only one out-of-limit predicted value for 40 samples in PLS predictions.

Paired t-test was performed to evaluate whether there was a significant difference between both the PLS and GILS model predicted values and between the reference measurements of each. The results are given in Table 14.

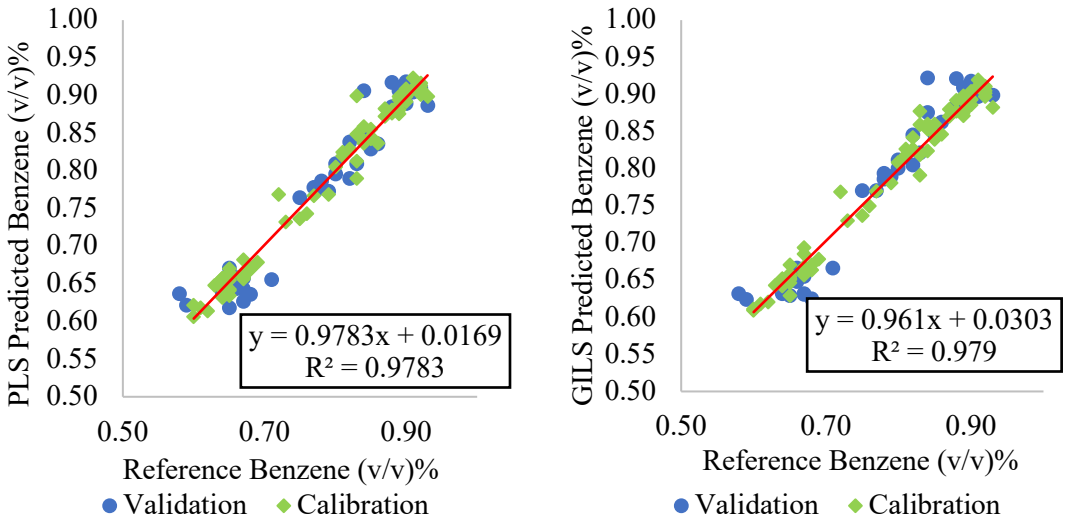
Table 14: Paired t-test Results for MTBE

	Reference-PLS Comparison	**Reference-GILS Comparison**	**PLS-GILS Comparison**
Observations	40		
df	39		
t Stat	-2.000	-0.763	**2.141**
p-value two-tail	0.052	0.450	**0.039**
t Critical two-tail	**2.023**		

According to Table 14, it can be said that, with 95% confidence, there is no statistically significant difference between reference measurements values and both model predictions. However, with the same confidence, there is a statistical difference between PLS and GILS predictions.

5.4.7. Benzene Content

Reference values obtained from Multidimensional gas chromatography (7890B Reformulyzer/AGILENT) vs model predicted values (PLS and GILS) of FTIR spectra of gasoline sample are given in Figure 35 for Benzene content.



(a) (b)

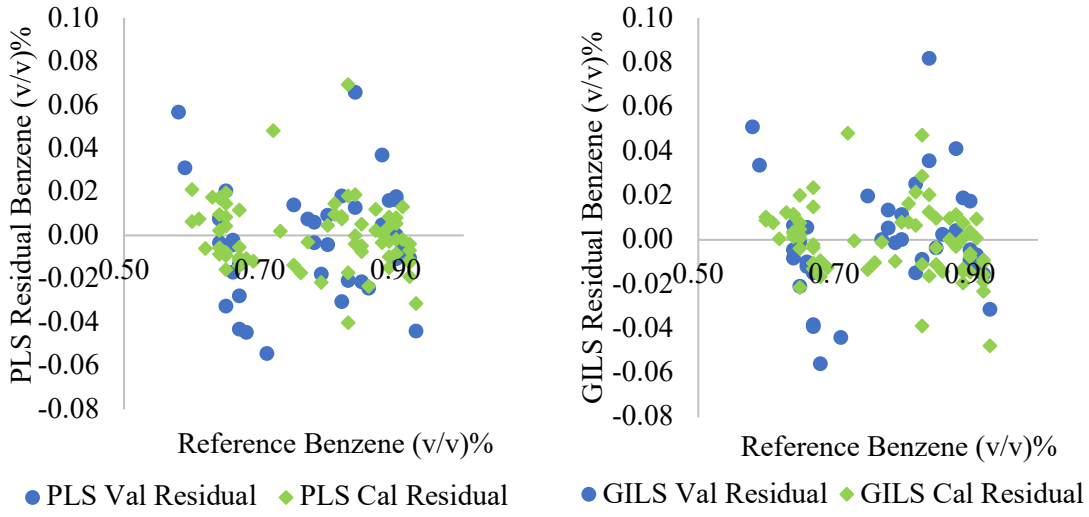
Figure 35: Reference Benzene Value vs Predicted Benzene Value; a) PLS, b)GILS

As can be seen in Figure 35, the calibration predictions and validation predictions of both models (PILS, GILS) are quite close. For PLS model, SECV and SEP values are found to be 0.0164 (v/v%) and 0.0259 (v/v%), respectively. The R^2 value for calibration set predictions are calculated as 0.9783, and the R^2 value for the validation set is 0.9461. For GILS model, SECV and SEP values are found to be 0.0163 (v/v%) and 0.0261 (v/v%) respectively. The R^2 value for calibration set predictions are calculated as 0.9790, and the R^2 value for the validation set is 0.9478. Prediction performances of the developed model for Benzene content are given in Table 15.

Table 15: Calibration Model Performance for Benzene Content

PLS Calibration Model Results		**GILS Calibration Model Results**		**Data Range (v/v %)**	
SECV (v/v %)	0.0164	SECV (v/v %)	0.0163	Max	0.93
SEP (v/v %)	0.0259	SEP (v/v %)	0.0261	Min	0.6
R^2 calibration	0.9783	R^2 calibration	0.9790	Interval	0.33
R^2 validation	0.9461	R^2 validation	0.9478	-	
Number of LVs	13	-		-	

In order to determine the error range and possible residual trends, the residuals for both calibration and validation samples are given in Figure 36. Most of the validation data are very close to calibration data which shows model prediction efficiency.



(a) **(b)**

Figure 36: Reference Benzene vs. Corresponding Model Prediction Benzene Residuals; a) PLS, b)GILS

The Reproducibility, R, for Benzene content given in TS EN ISO 22854 is described as Eq. (22). The R-chart with absolute differences between GILS and PLS model predictions and reference analysis is given in Figure 37.

$$\pm \mathbf{0.0777Y - 0.025} \qquad \text{Eq. (22)}$$

Where Y is the mean of the two results being compared.

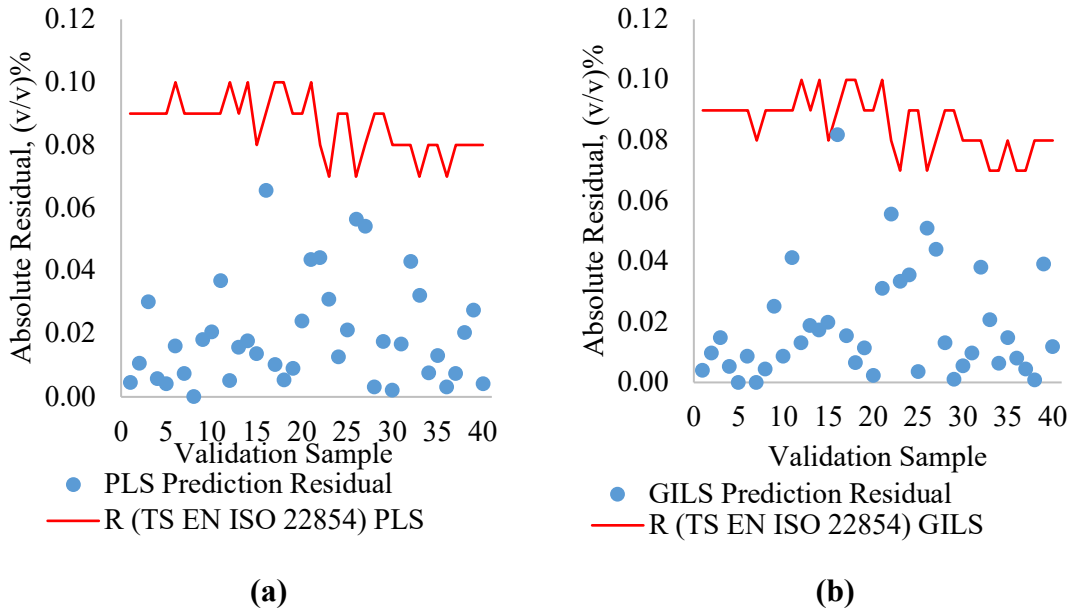


Figure 37: Benzene Residual R-Chart; a)PLS, b)GILS

As seen in Figure 37, there is no out-of-limit predicted value for 40 samples.

Paired t-test was performed to evaluate whether there was a significant difference between both the PLS and GILS model predicted values and between the reference measurements of each. The results are given in Table 16.

Table 16: Paired t-test Results for Benzene Content

	Reference-PLS Comparison	**Reference-GILS Comparison**	**PLS-GILS Comparison**
Observations	40		
df	39		
t Stat	0.551	-0.044	-1.364
p-value two-tail	0.585	0.965	0.180
t Critical two-tail	2.023		

According to Table 16, it can be said that, with 95% confidence, there is no statistically significant difference between reference measurements values and both model predictions, and also between the PLS and GILS predictions.

5.4.8. Olefin Content

Reference values obtained from multidimensional gas chromatography (7890B Reformulyzer/AGILENT) vs model predicted values (PLS and GILS) of FTIR spectra of gasoline sample are given in Figure 38 for Olefin content.

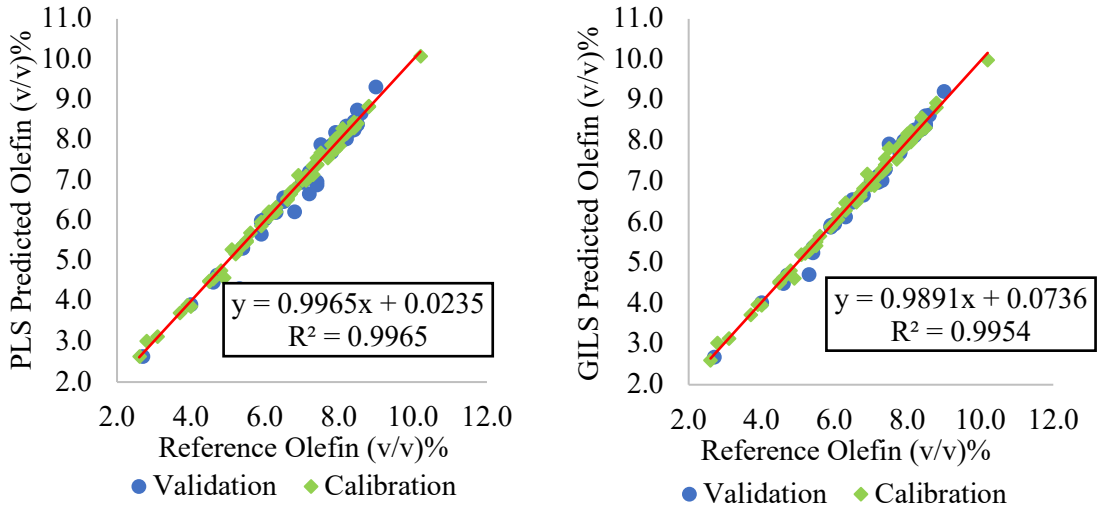


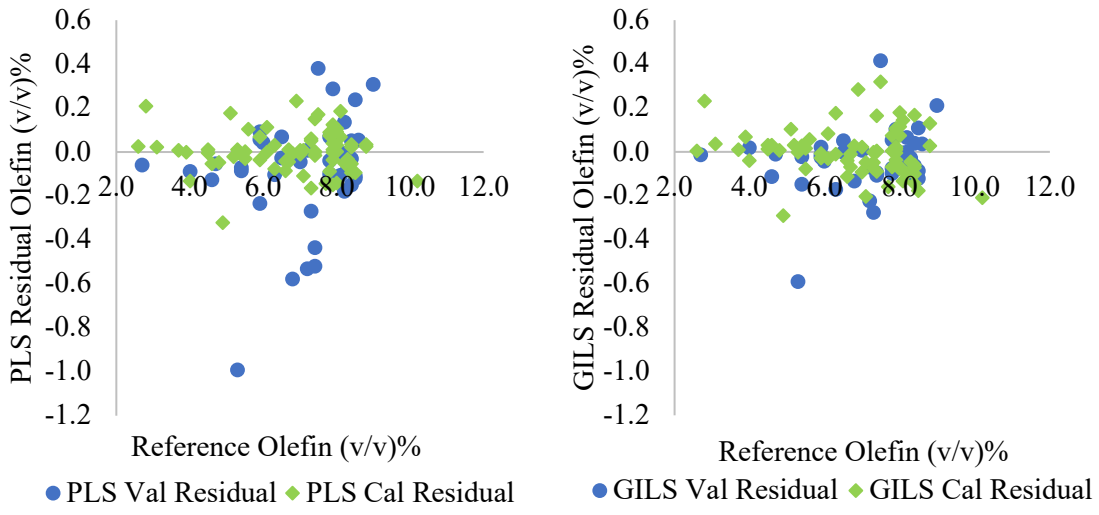
Figure 38: Reference Olefin Value vs Predicted Olefin Value; a) PLS, B)GILS

As can be seen in Figure 38, the calibration predictions and validation predictions of both models (PILS, GILS) are quite close. For PLS model, SECV and SEP values are found to be 0.0953 (v/v%) and 0.2675 (v/v%), respectively. The R^2 value for calibration set predictions are calculated as 0.9965, and the R^2 value for the validation set is 0.9745. For GILS model, SECV and SEP values are found to be 0.1102 (v/v%) and 0.1523 (v/v%) respectively. The R^2 value for calibration set predictions are calculated as 0.9954, and the R^2 value for the validation set is 0.9910. Prediction performances of the developed model for Olefin content are given in Table 17.

Table 17: Calibration Model Performance

PLS Calibration Model Results		**GILS Calibration Model Results**		**Data Range (v/v %)**	
SECV (v/v %)	0.0953	SECV (v/v %)	0.1102	Max	10.2
SEP (v/v %)	0.2675	SEP (v/v %)	0.1523	Min	2.6
R^2 calibration	0.9965	R^2 calibration	0.9954	Interval	7.6
R^2 validation	0.9745	R^2 validation	0.9910	-	
Number of LVs	19	-		-	

In order to determine the error range and possible residual trends, the residuals for both calibration and validation samples are given in Figure 39. Most of the validation data are very close to calibration data which shows model prediction efficiency. It is observed that the residual values increase as you move away from the points where the reference values are concentrated.



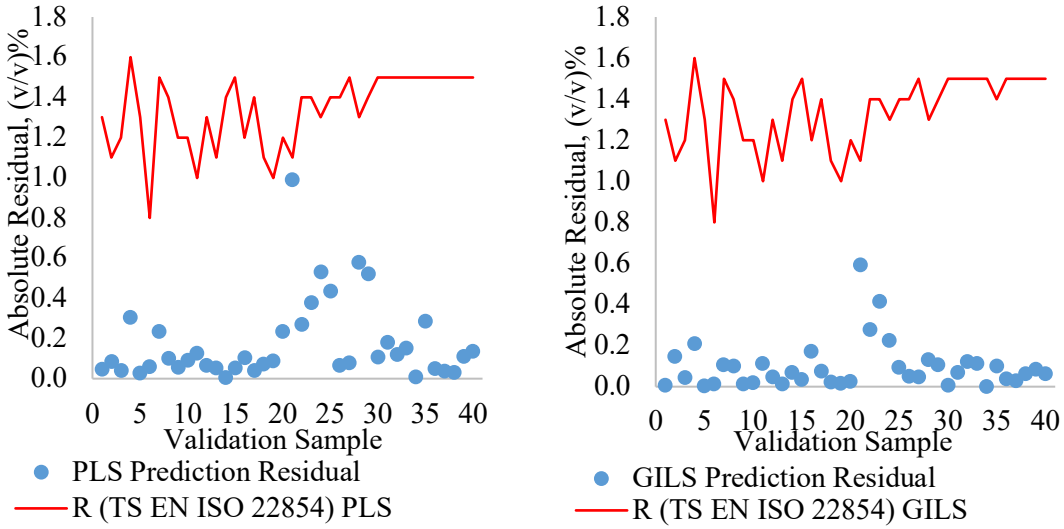
(a) **(b)**

Figure 39: Reference Olefin vs. Corresponding Model Prediction Olefin Residuals; a) PLS, b)GILS

The Reproducibility, R, for Olefin content given in TS EN ISO 22854 is described as Eq. (23). The R-chart with absolute differences between GILS and PLS model predictions and reference analysis is given in Figure 40.

$$\pm 0.1176Y + 0.5118 \quad \text{Eq. (23)}$$

Where Y is the mean of the two results being compared.



(a) **(b)**

Figure 40: Olefin Residual R-Chart; a)PLS, b)GILS

As seen in Figure 40, there is no out-of-limit predicted value for 40 samples.

Paired t-test was performed to evaluate whether there was a significant difference between both the PLS and GILS model predicted values and between the reference measurements of each. The results are given in Table 18.

Table 18: Paired t-test Results for Olefin Content

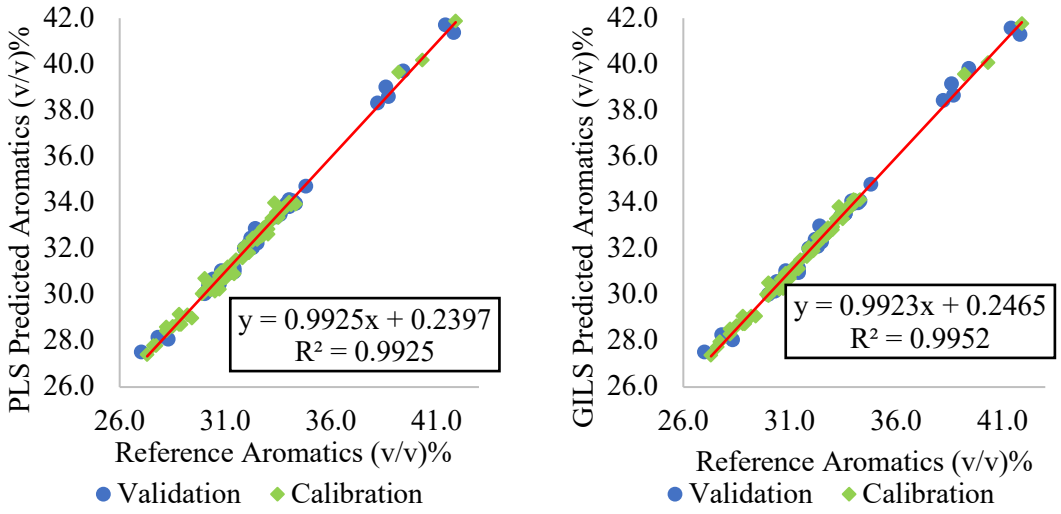
	Reference-PLS Comparison	**Reference-GILS Comparison**	**PLS-GILS Comparison**
Observations	40		
df	39		
t Stat	**2.150**	1.818	-1.878
p-value two-tail	**0.038**	0.077	0.068
t Critical two-tail	**2.023**		

According to Table 18, it can be said that, with 95% confidence, there is no statistically significant difference between reference measurements values and GILS model predictions, and also between the PLS and GILS predictions. However, with the

same confidence, there is a statistical difference between reference measurements and PLS predictions.

5.4.9. Aromatic Contents

Reference values obtained from multidimensional gas chromatography (7890B Reformulyzer/AGILENT) vs model predicted values (PLS and GILS) of FTIR spectra of gasoline sample are given in Figure 41 for Aromatic content.



(a) (b)

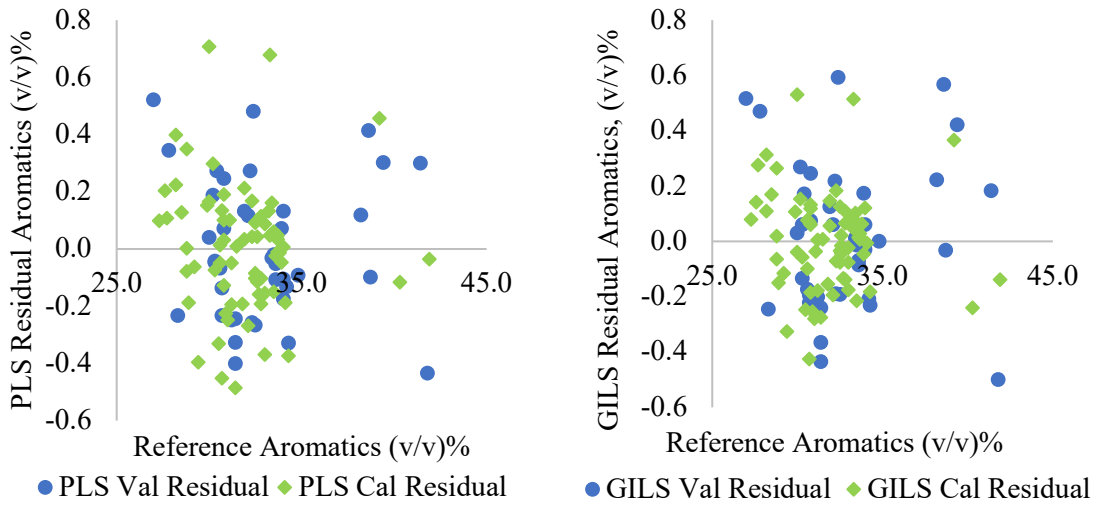
Figure 41: Reference Aromatic Value vs Predicted Aromatic Value; a) PLS, B)GILS

As can be seen in Figure 41, the calibration predictions and validation predictions of both models (PILS, GILS) are quite close. For PLS model, SECV and SEP values are found to be 0.2262 (v/v%) and 0.2421 (v/v%), respectively. The R^2 value for calibration set predictions are calculated as 0.9925, and the R^2 value for the validation set is 0.9949. For GILS model, SECV and SEP values are found to be 0.1801 (v/v%) and 0.2619 (v/v%) respectively. The R^2 value for calibration set predictions are calculated as 0.9952, and the R^2 value for the validation set is 0.9940. Prediction performances of the developed model for Aromatic content are given in Table 19.

Table 19: Calibration Model Performance for Aromatic Content

PLS Calibration Model Results		**GILS Calibration Model Results**		**Data Range (v/v %)**	
SECV (v/v %)	0.2262	SECV (v/v %)	0.1801	Max	41.9
SEP (v/v %)	0.2421	SEP (v/v %)	0.2619	Min	27.3
R^2 calibration	0.9925	R^2 calibration	0.9952	Interval	14.6
R^2 validation	0.9949	R^2 validation	0.9940	-	
Number of LVs	9	-		-	

In order to determine the error range and possible residual trends, the residuals for both calibration and validation samples are given in Figure 42. Most validation data are very close to calibration data which shows model prediction efficiency. It is observed that the residual values increase as you move away from the points where the reference values are concentrated.



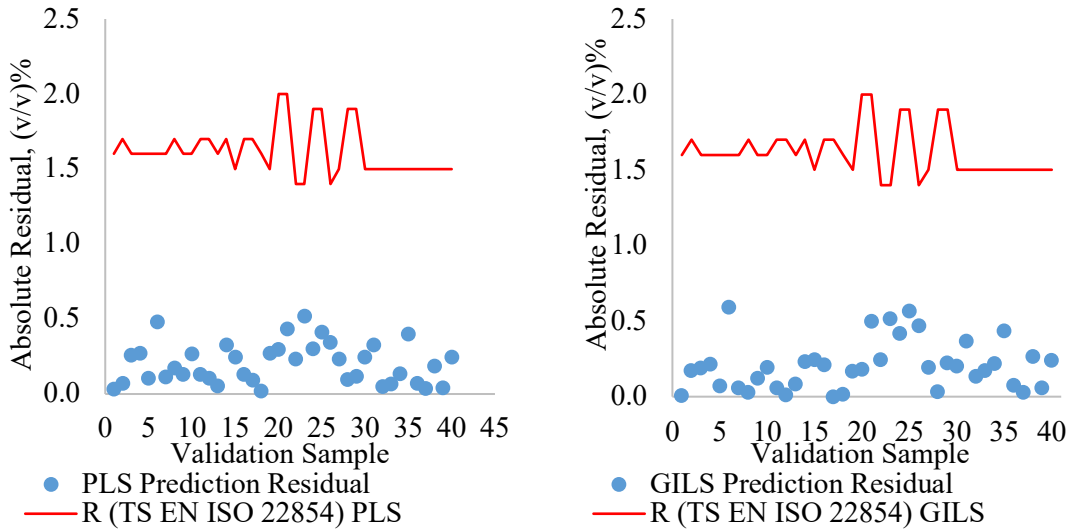
(a) **(b)**

Figure 42: Reference Aromatic vs. Corresponding Model Prediction Aromatic Residuals; a) PLS, b)GILS

The Reproducibility, R, for Aromatic content given in TS EN ISO 22854 is described as Eq. (24). The R-chart with absolute differences between GILS and PLS model predictions and reference analysis is given in Figure 43.

$$\pm 0.045Y + 0.1384 \quad \text{Eq. (24)}$$

Where Y is the mean of the two results being compared.



(a) **(b)**

Figure 43: Aromatic Residual R-Chart; a)PLS, b)GILS

As seen in Figure 43, there is no out-of-limit predicted value for 40 samples.

Paired t-test was performed to evaluate whether there was a significant difference between both the PLS and GILS model predicted values and between the reference measurements of each. The results are given in Table 20.

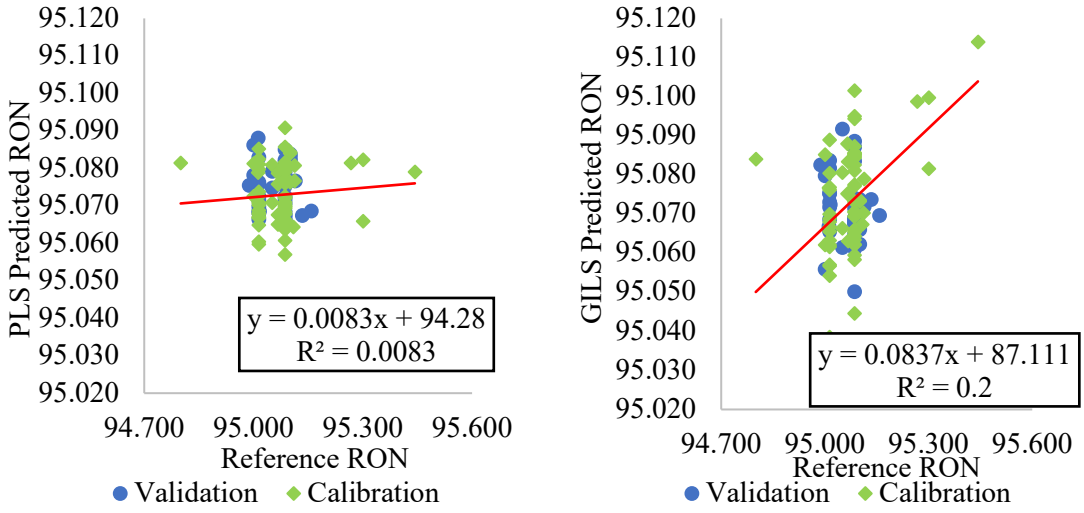
Table 20: Paired t-test Results for Aromatic Content

	Reference-PLS Comparison	**Reference-GILS Comparison**	**PLS-GILS Comparison**
Observations	40		
df	39		
t Stat	0.054	-0.399	-1.529
p-value two-tail	0.958	0.692	0.134
t Critical two-tail	2.023		

According to Table 20, it can be said that, with 95% confidence, there is no statistically significant difference between reference measurements values and both model predictions, and also between the PLS and GILS predictions.

5.4.10. RON - Research Octane Number

Reference values obtained from Cooperative Fuel Research (CFR) engines vs model predicted values (PLS and GILS) of FTIR spectra of gasoline sample are given in Figure 44 for RON parameter.



(a) (b)

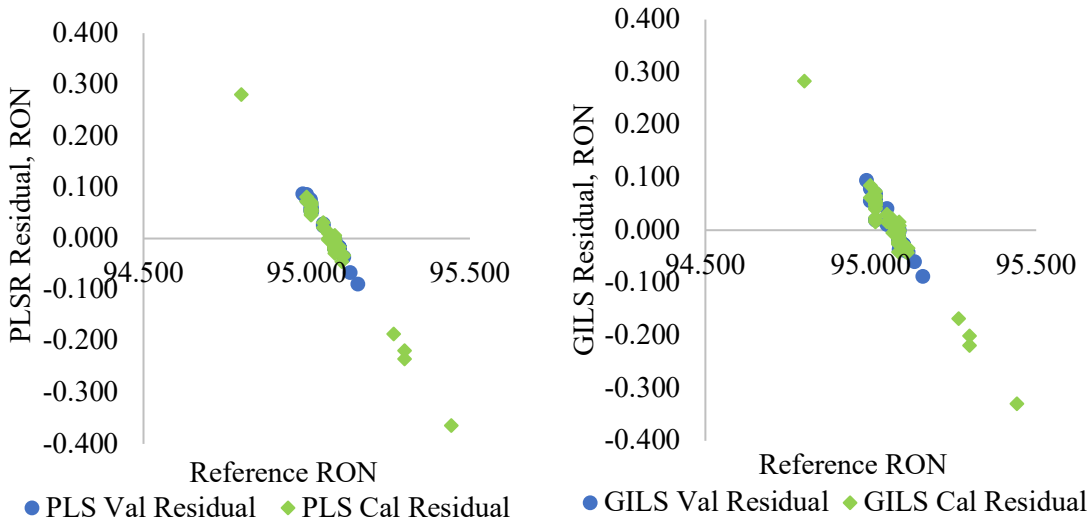
Figure 44: Reference RON Value vs Predicted RON Value; a) PLS, B)GILS

For PLS model, SECV and SEP values are found to be 0.0802 (v/v%) and 0.0453 (v/v%), respectively. The R^2 value for calibration set predictions are calculated as 0.0083, and the R^2 value for the validation set is $9*10^{-8}$. For GILS model, SECV and SEP values are found to be 0.0750 (v/v%) and 0.0453 (v/v%) respectively. The R^2 value for calibration set predictions are calculated as 0.2000, and the R^2 value for the validation set is 0.0007. Prediction performances of the developed model for the RON parameter are given in Table 21.

Table 21: Calibration Model Performance for RON

PLS Calibration Model Results		GILS Calibration Model Results		Data Range	
SECV	0.0802	SECV	0.0750	Max	95.443
SEP	0.0453	SEP	0.0453	Min	94.800
R^2 calibration	0.0083	R^2 calibration	0.2000	Interval	0.643
R^2 validation	0.00000009	R^2 validation	0.0007	-	
Number of LVs	1	-		-	

In order to determine the error range and possible residual trends, the residuals for both calibration and validation samples are given in Figure 45. When the residuals are examined, there seems to be a systematic error, as in Figure 45. Since the octan number is one of the most important parameters of gasoline, it is not a very variable parameter from product to product. It is seen in Table 21 that, the reference value range is very narrow. The narrow reference value range in the calibration set causes the R^2 value to be very low and a systematic error to appear, although the residual values are proportionally very small.



(a) **(b)**

Figure 45: Reference RON vs. Corresponding Model Prediction RON Residuals; a) PLS, b)GILS

The Reproducibility, R, for RON given in TS EN ISO 5164 is described as 0.7. Since the octane number parameters directly represent the product quality, precision in

its measurement uncertainty is important. For this reason, TS EN ISO 5164 repeatability, r, value (described as 0.2), and measurement differences were also compared. The R-chart with absolute differences between GILS and PLS model predictions and reference analysis is given in Figure 46.

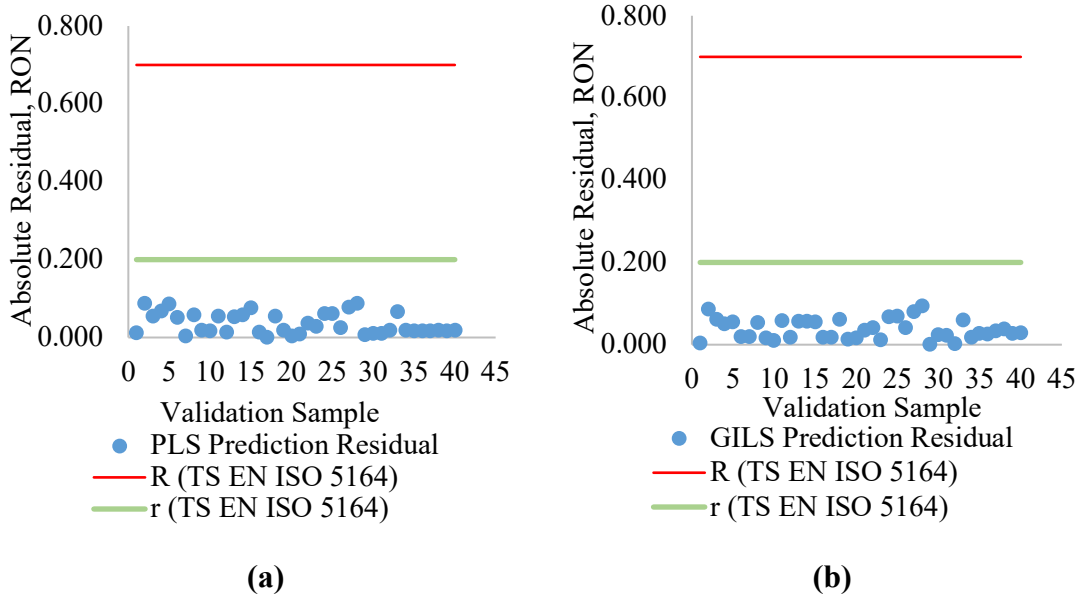


Figure 46: RON Residual R-Chart; a)PLS, b)GILS

As seen in Figure 46, there is no out-of-limit predicted value for 40 samples.

Paired t-test was performed to evaluate whether there was a significant difference between both the PLS and GILS model predicted values and between the reference measurements of each. The results are given in Table 22.

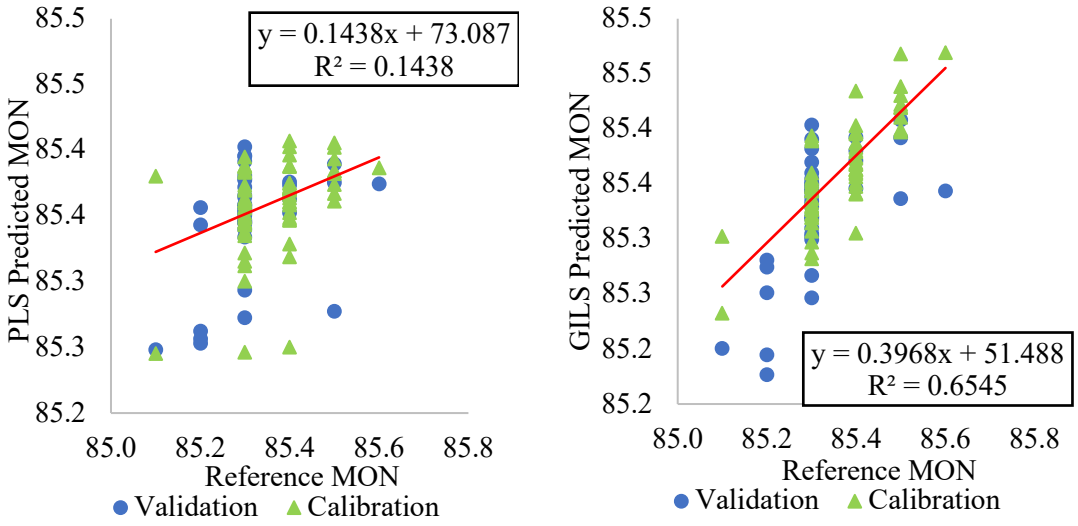
Table 22: Paired t-test Results for RON

	Reference-PLS Comparison	**Reference-GILS Comparison**	**PLS-GILS Comparison**
Observations	40		
df	39		
t Stat	-1.771	-1.049	**2.385**
p-value two-tail	0.084	0.301	**0.022**
t Critical two-tail	**2.023**		

According to Table 22, it can be said that, with 95% confidence, there is no statistically significant difference between reference measurements values and both model predictions. However, with the same confidence, there is a statistical difference between PLS and GILS predictions.

5.4.11. MON - Motor Octane Number

Reference values obtained from Cooperative Fuel Research (CFR) engines vs model predicted values (PLS and GILS) of FTIR spectra of gasoline sample are given in Figure 47 for MON parameter.



(a) (b)

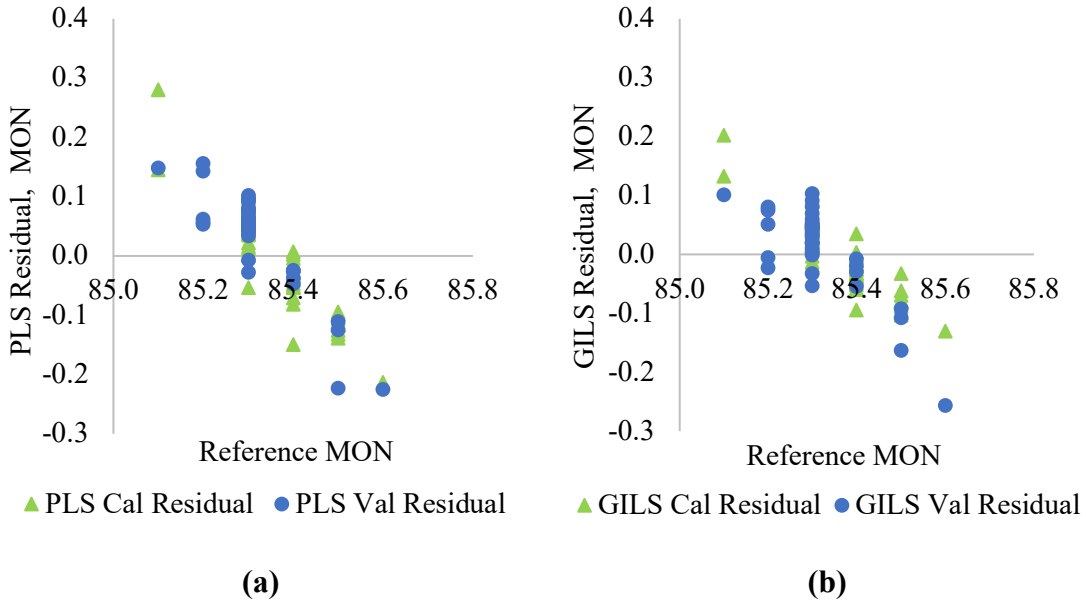
Figure 47: Reference MON Value vs Predicted MON Value; a) PLS, B)GILS

For PLS model, SECV and SEP values are found to be 0.0815 (v/v%) and 0.0900 (v/v%), respectively. The R^2 value for calibration set predictions are calculated as 0.1438, and the R^2 value for the validation set is 0.1733. For GILS model, SECV and SEP values are found to be 0.0589 (v/v%) and 0.0724 (v/v%) respectively. The R^2 value for calibration set predictions are calculated as 0.6545, and the R^2 value for the validation set is 0.3989. Prediction performances of the developed model for MON parameter are given in Table 23.

Table 23: Calibration Model Performance for MON

PLS Calibration Model Results		GILS Calibration Model Results		Data Range	
SECV	0.0815	SECV	0.0589	Max	85.6
SEP	0.0900	SEP	0.0724	Min	85.1
R^2 calibration	0.1438	R^2 calibration	0.6545	Interval	0.5
R^2 validation	0.1733	R^2 validation	0.3989	-	
Number of LVs	3	-		-	

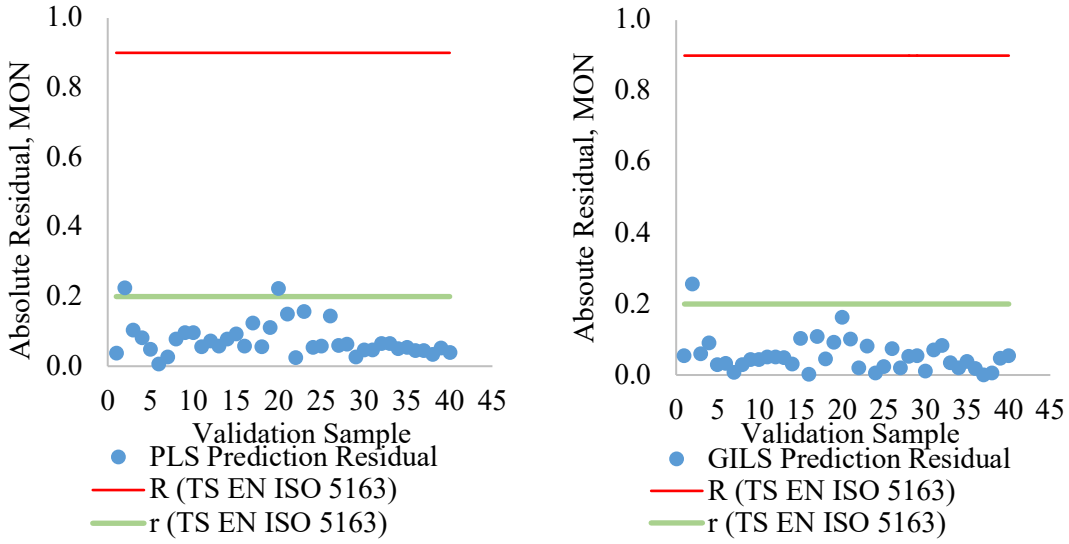
In order to determine the error range and possible residual trends, the residuals for both calibration and validation samples are given in Figure 48. When the residuals are examined, there seems to be a systematic error. Since the octan number is one of the most important parameters of gasoline, it is not a very variable parameter from product to product. It is seen in Table 23 that, the reference value range is very narrow. The narrow reference value range in the calibration set causes the R^2 value to be very low and a systematic error to appear, although the residual values are proportionally very small.



(a) **(b)**

Figure 48: Reference MON vs. Corresponding Model Prediction MON Residuals; a) PLS, b)GILS

The Reproducibility, R, for MONs given in TS EN ISO 5163 is described as 0.9, and repeatability, r, value is described as 0.2. The R-chart with absolute differences between GILS and PLS model predictions and reference analysis is given in Figure 49.



(a) **(b)**

Figure 49: MON Residual R-Chart; a)PLS, b)GILS

As seen in Figure 49, there is no value outside the reproducibility limit from both model predictions. One of the GILS predictions and two of the PLS predictions are outside the repeatability limit.

Paired t-test was performed to evaluate whether there was a significant difference between both the PLS and GILS model predicted values and between the reference measurements of each. The results are given in Table 24.

Table 24: Paired t-test Results for MON

	Reference-PLS Comparison	**Reference-GILS Comparison**	**PLS-GILS Comparison**
Observations	40		
df	39		
t Stat	**-2.433**	-1.025	**4.173**
p-value two-tail	**0.020**	0.312	**0.0002**
t Critical two-tail	**2.023**		

According to Table 24, it can be said that, with 95% confidence, there is no statistically significant difference between reference measurements values and GILS model predictions. However, with the same confidence, there is a statistical difference between PLS and GILS predictions and also between PLS and reference values.

CHAPTER 6

CONCLUSION

In this study, a new, fast and simple analytical method has been developed to determine crucial parameters of gasoline blends. For this purpose, molecular spectroscopic techniques, namely FTIR spectroscopy was used along with two different chemometric multivariate calibration methods which are GILS and PLS. All gasoline samples were collected at Tupras Izmır Refinery. The gasoline samples were analyzed at the Tupras Izmir Refinery Quality Control Laboratory by the reference test methods. GC, density meter, distillation instrument were used for reference analysis.

When the results were examined, it was observed that although there were significant differences in some parameters, in general, two multivariate calibration methods (PLS, GILS) gave close results, these results can be reached by looking at their SECV and SEP with R^2 values. Since working with commercial gasoline samples, sample specs are in a very narrow range. It is obvious that this situation has a negative effect on the calibration model performances. Although the R^2 value is found to be very small in parameters with a very small data range such as RON and MON, the SECV and SEP values are proportionally very small and meet the relevant international standard requirements. It has been observed that the GILS results are better in these parameters where the data range is narrow. Except for Density and E150 parameters, all parameters have met the reproducibility limits specified in the relevant international standards.

Although the data range used for calibration is very narrow, it can be accepted that the calibration models created are significantly successful, especially considering the reproducibility limits in the relevant reference measurement methods of the parameters. Using an experimental design while creating a calibration data set, or collecting a large number of samples for a long time to expand the data range can improve model performance.

REFERENCES

A. Savitzky, M. G. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures.

Agilent. (2018). Agilent 7890B GC Supplies. USA. Retrieved from https://www.agilent.com/cs/library/sales/public/59908916EN_7890B_Supplies_QRG.pdf

Akova, İ. (2019). Enerji ve Alternatif Enerji Kaynakları. İstanbul.

Ankara University. (2021). Infrared Spektroskopisi. Ankara Üniversitesi Açık Ders Malzemeleri: https://acikders.ankara.edu.tr/pluginfile.php/51386/mod_resource/content/0/IR%20spektroskopisi.pdf

Anton Paar. (2021). DMA 4500 M Density meter. https://www.anton-paar.com/kr-en/products/details/benchtop-density-meter-dma-4500-m/

Anton-Paar. (2009). Retrieved from https://wiki.anton-paar.com/tr-tr/atmosferik-distilasyon/

Aydoğmuş, Y. (2016, 12). Routing And Resource Allocation For Software Defined Mobile Networks.

Beşergil, B. (2009). Rafineri Prosesleri. İzmir: Ege Üniversitesi Basımevi.

Biancolillo, A., & Marini, F. (2018, 11 21). Chemometric Methods for Spectroscopy-Based Pharmaceutical Analysis. Rome, Italy. doi:https://doi.org/10.3389/fchem.2018.00576

Bohács, G., Ovádi, Z., & Salgó, A. (1998). Prediction of gasoline properties with near infrared spectroscopy. Journal of Near Infrared Spectroscopy. doi:10.1255/jnirs.155

Breitkreitz, M. C., Ivo M. Raimundo, J., Rohwedder, J. J., Pasquini, C., Filho, H. A., Joséb, G. E., & Araújo, M. C. (2003). Determination of total sulfur in diesel fuel employing NIR spectroscopy and multivariate calibration. The ANALYST, 1204-1207.

Chevron Corporation. (2009). Motor Gasolines Technical Review. USA.

Chung, H. (2007). Applications of Near-Infrared Spectroscopy in Refineries and Important Issues to Address. doi:https://doi.org/10.1080/05704920701293778

Colthub, N. (2001). Encyclopedia of Physical Science and Technology. R. Meyers. Connecticut.

Concawe. (2002, 10 1). The Focus of Aromatics in Automotive Fuels Specifications. https://www.concawe.eu/wp-content/uploads/2017/01/cr112-aromatics-2003-01897-01-e.pdf

Corfin Lubrication. (2020, 12 14). Corfin Lubrication. Corfin Lubrication: https://www.corfin.com.tr/petrol-ve-petrol-rafinasyonu/

Dinç, E. (2007). Kemometri Çok Değişkenli Kalibrasyon Yöntemleri. Hacettepe Üniversitesi, Eczacılık Fakültesi Dergisi, 61-92.

Fellah, F. (2016). Petrol ve Petrol Kimyası. Bursa. https://sayfam.btu.edu.tr/upload/dosyalar/1461049852KMS216-3.pdf

Fodor, G., & Hutzler, S. (1997). ESTIMATION OF MIDDLE DISTILLATE FUEL PROPERTIES BY FT-IR AND CHEMOMETRICS Part I. CALIBRATIONS AND VALIDATIONS. Michigan: U.S. Army TARDEC Mobility Technology Center.

Hönig, V., Orsak, M., Pexa, M., & Linhart, Z. (2015). The distillation characteristics of automotive gasoline containing. Agronomy Research, 558-567.

International Standard Organisation. (2007, 07 03). TS EN ISO 12185. Crude petroleum and petroleum products - Determination of density - Oscillating U- tube method.

International Standard Organisation. (2017, 12 18). TS EN 228:2012. Automotive fuels - Unleaded petrol - Requirements and test methods.

International Standards Organisation;. (2007). TS EN ISO 12185. Crude petroleum and petroleum products –.

International Standards Organisation. (2014). TS EN ISO 5163. Petroleum products - Determination of knock characteristics of motor and aviation fuels - Motor method.

International Standards Organisation. (2014). TS EN ISO 5164. Petroleum products - Determination of knock characteristics of motor fuels - Research method.

ISO. (2017, 12 18). TS EN ISO/IEC 17025. General requirements for the competence of testing and calibration laboratories.

Kalivas, J. (1997). Two data sets of near infrared spectra. Chemometrics and Intelligent Laboratory Systems, 255-259.

National Centre for Catalysis Research. (2006). An Introduction to Energy Sources. Chennai.

Özdemir, D. (2005). Determination of Octane Number of Gasoline Using Near Infrared Spectroscopy and Genetic Multivariate Calibration Methods. Petroleum Science and Technology, 1139-1152. doi:http://dx.doi.org/10.1081/LFT-200035547

Özdemir, D. (2008). Near Infrared Spectroscopic Determination of Diesel Fuel Parameters Using Genetic Multivariate Calibration. Petroleum Science and Technology, 101-113. doi:10.1080/10916460600705824

Özdemir, D., & Öztürk, B. (2003, 10 8). Genetic Multivariate Calibration Methods for Near Infrared (NIR) Spectroscopic Determination of Complex Mixtures. İzmir.

Özdemir, D., & Öztürk, B. (2006, 09 22). Near Infrared Spectroscopic Determination of Olive Oil Adulteration with Sunflower and Corn Oil. İzmir.

PAC. (2021). OptiDist: Atmospheric Distillation: https://www.paclp.com/lab-instruments/brand/herzog/product/7/optidist-atmospheric-distillation

R. J. BARNES, M. S. (1989). Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra. UK.

Rand, S. J. (2010). Significance of Tests for Petroleum Products. USA: ASTM International.

Reboucas, M., Santos, J., Pimentel, M. F., & Teixeira, L. (2011). A novel approach for development of a multivariate calibration model using a Doehlert experimental design: Application for prediction of key gasoline properties by Near-infrared Spectroscopy. Chemometrics and Intelligent Laboratory Systems, 185-193.

Reguant, R. (2021, 3 22). Roulette Wheel Selection in Python. https://rocreguant.com/roulette-wheel-selection-python/2019/

Schneider Electric. (2021). AIT Analect: http://www.aitanalyzers.com/ftir-ANALECT-lab.php

Sivasankar, B. (2008). Engineering Chemistry. New Delphi: Tata McGraw-Hill Publishing.

Temiz, H. T. (2019). Utilization Of Spectroscopic Methods For Determining Food Adulteration With Chemometric Approaches. Ankara.

Thermo Fisher Scientific Inc. (2013). Introduction to Fourier Transform Infrared Spectroscopy. https://tools.thermofisher.com/content/sfs/brochures/BR50555_E_0513M_H_1.pdf

Thomson, W. T. (2007). Theory of Vibrations with Applications. Pearson.

Turkey Ministry of Environment and Urbanization. (2019). Petrol Rafinerisi Sektörel Uygulama Klavuzu. https://webdosya.csb.gov.tr/db/sanayihavarehberi/icerikler/raf-ner--tes-sler-_r-20200110113315.pdf

Waukesha CFR. (2021). Combination Research & Motor Method Octane Rating. USA. https://www.corelab.com/refinery/cms/docs/F1F2_Product_Brochure_GEA_30653.pdf