Multivariate analysis and modeling of crude oil composition and fluid properties relevant for Multiphase Flow

Andreas Linge Tomren

Dissertation for the degree philosophiae doctor (PhD)

at the University of Bergen

2014

Dissertation date: 5. september 2014

Year: 2014

Title: Multivariate analysis and modeling of crude oil composition and fluid properties relevant for Multiphase Flow Metering

Author: Andreas Linge Tomren

Print: AIT OSLO AS / University of Bergen

Preface

The work presented in this PhD thesis was carried out at the Department of Chemistry at the University of Bergen in the period October 2008-October 2012. The thesis consists of a research report, four papers and a summary of the obtained results as well as the techniques and methods applied in the work. The work has been financed by the Norwegian Research Council and the industry partner Roxar through the Michelsen Centre for Industrial Measurement Science and Technology at Christian Michelsen Research in a project called “Epsilon”. The project is interdisciplinary, combining physical chemistry, petroleum chemistry, physics and chemometrics. During the three years of research I have participated at two international conferences and two Norwegian conferences.

Roxar is a leading provider of advanced technology to the oil and gas industry, and some of their products are the Multiphase Flow Metering systems. These systems can provide measurements of the water cut and flow rates of oil, gas and water in pipelines. Data from the systems can help in optimising petroleum production, increasing the oil recovery and lowering the investments and operational costs. This project seeks to improve the performance of the Multiphase Flow Metering systems by improving the understanding of how the different compounds in crude oil influence the physical properties of crude oil, in particular the properties that are involved in Multiphase Flow Metering.

Acknowledgements

I would like to thank my supervisors Tanja Barth, Kjetil Folgerø and Johan E. Carlson. A special thanks to my main supervisor Tanja Barth for her guidance and expertise in petroleum chemistry. She has acted as an inexhaustible resource of wisdom to me; numerous times have I entered her office with a question, and walked out as a wiser man, often with a book or reference for further exploration. Kjetil Folgerø should also be thanked for introducing me to the world of permittivity and dielectric spectroscopy.

Thanks to everyone in Tanja's group, as well as other friends and colleagues who made the Department of Chemistry a good working environment, Ina, Djurdjica, James, Espen and Dagfinn. A special thank to my office mate, Ina, for sharing laughs, discussions and coffee during my PhD period.

Thanks to my family for always encouraging me to do my best and not give up. A special thank to my father, Kai, for advising me to keep all options open by choosing science subjects like chemistry during high school.

Finally, thanks to my girlfriend and soul mate Susanne, our daughter Ellinor and son Magnus, for reminding me that there is a life outside the office at the Department of Chemistry, and for supporting and pushing me to do my best.

Abstract

Crude oil composition and properties have been subject for research for several decades, as well as the correlation between the chemical composition and physical properties of crude oil. The physical properties of crude oil are strongly dependent upon the chemical composition of the oil, and both physical properties and composition provide important information for the oil producers as different challenges might occur when handling certain types of crude oil; corrosion, deposition of solids and plugging of pipelines are just some examples. As a mean of detecting possible threats and monitoring the flow through the pipelines, a number of metering devices can be applied, for example a Multiphase Flow Meter, which can provide measurements of the water cut and the flow rates of oil, gas and water in pipelines, as well as pressure and density.

The aim of this work is to improve the general understanding of how the distribution of the different compounds in crude oil influence the physical properties of crude oil, in particular the properties that are involved in Multiphase Flow Metering, where permittivity is of special relevance. By improving this understanding, the project seeks to improve the knowledge base on the interactions between the chemistry of the petroleum fluids and the parameters that provide the basis for the performance of the Multiphase Flow Metering systems.

In this work, calibration models of several physical properties and compositional data has been built, based on GC and IR data respectively. Several of the models are obtained with good predictive quality, for instance the models for static and high frequency permittivity, density and velocity of sound. Also, a clear biodegradation effect has been identified in the regression coefficients for these properties, indicating that biodegradation has a significant effect on their variation.

A multivariate calculation tool for estimation of static permittivity of crude oils based on PVT data has been developed during the thesis work, and this tool has been

applied by Roxar ASA. The tool gives a better initial calibration of the Multiphase Flow Metering systems, as the static permittivity is more accurately determined compared to the previously used methods. The tool should also be able to identify significant changes in oil composition and properties for example during the production lifetime of an oil field (given representative PVT data), and recalibrate values in Multiphase Flow Metering systems accordingly.

List of publications

Research report: Tomren, A.L.; Barth, T.; Folgerø, K. Introductory chemometric analysis of crude oil composition and fluid properties.

Paper I: Tomren, A.L.; Barth, T.; Folgerø, K. 2012. Multivariate analysis of crude oil composition and fluid properties used in Multiphase Flow Metering. *Energy Fuels,* 56, 5679-5688.

Paper II: Tomren, A.L.; Barth, T. 2014. Comparison of Partial Least Squares calibration models of viscosity, acid number and asphaltene content in petroleum, based on GC and IR data. *Fuel,* 120, 8-21.

Paper III: Carlson, J.E.; Tomren, A.L.; Folgerø, K.; Barth, T. 2014. Estimation of dielectric properties of crude oils based on IR spectroscopy. *Submitted to* Chemom. Intell. Lab. Sys., *second revision*.

Paper IV: Folgerø, K.; Tomren, A. L.; Frøyen, S. 2012. Permittivity calculator. Method and tool for calculating the permittivity of oils from PVT data. *Conference proceedings, 30th International North Sea Flow Measurement Workshop, St. Andrews, 23rd – 26th October.*

Reprints were made with permission from Energy Fuels and Fuel.

Abbreviations

ATR	Attenuated Total Reflection
CP	Conical Plate
GC	Gas Chromatography
FID	Flame Ionization Detector
FT	Fourier Transform
LV	Latent Variable
k, M, G	kilo (10^3), Mega(10^6), Giga(10^9)
MFM	Multiphase Flow Meter
mPas, MPas	milli(10^{-3}) Pascal · seconds, Mega(10^6) Pascal · seconds
μm, μL	micro(10^{-6}) meter, micro Liter
NIGOGA	Norwegian Industry Guide to Organic Geochemical Analyses
IR	InfraRed
PC	Principal Component
PCA	Principal Component Analysis
PLS	Partial Least Squares
PLSR	Partial Least Squares Regression
TAN	Total Acid Number
WOGC	Whole Oil Gas Chromatography

Symbols

α	Alpha, empirical factor (permittivity)
c_{fluid}	Velocity of sound (in fluid), m/s
ε_∞	High frequency permittivity
ε_r	Relative permittivity
ε_s	Static permittivity
η	Shear viscosity, mPas
ρ	Density, g/cm^3
σ	Sigma, conductivity, Siemens/meter (permittivity)
τ	Tau, relaxation time, seconds (permittivity)

Contents

Preface iii

Acknowledgements iv

Abstract v

List of Publications vii

Abbreviations viii

Symbols ix

1 Introduction and outline 1

2 Petroleum/crude oil 3

2.1 Petroleum formation 3

2.2 Chemical composition and properties of petroleum 4

2.3 Petroleum types; condensates, biodegraded oil and nondegraded oil 5

2.4 Production and transport of petroleum 8

2.5 Flow assurance; Multiphase Flow Metering (MFM) 9

2.6 Representative oil sampling 10

3 Infra Red Spectroscopy 11

3.1 Vibrational spectroscopy 11

3.2 Absorption of infrared energy in Attenuated Total Reflection FT-IR 15

3.3 IR Spectrophotometer: ATR- FT- IR 17

3.4 Measuring procedure 18

3.5 FTIR sample spectra 19

4 Whole Oil Gas Chromatography (WOGC) 23

4.1 Measuring principle 23

4.2 Measuring procedure 25

4.3 WOGC sample chromatograms 26

5 Analysis of chemical composition and physical properties of crude oil 29

5.1 Density 29

5.1.1 Measurement principle 30

5.1.2 Instrument and setup 31

5.1.3 Thermostat control and temperature measurement 32

5.1.4 Calibration and control of the density meter 33

5.1.5 Measuring procedure 33

5.2 Velocity of sound 34

5.2.1 Theory 35

5.2.2 Measuring principle/procedure 36

5.3 Viscosity 37

5.3.1 Theory 37

5.3.2 Measuring principle/procedure 39

5.4 Total Acid Number 41

5.4.1 Acids in petroleum 41

5.4.2 Measuring principle 42

5.4.3 Measuring procedure 43

5.5 Asphaltene content 47

5.5.1 Theory 48

5.5.2 Measuring principle/procedure 52

5.6 Permittivity 52

5.6.1 The Cole- Cole model 56

5.6.2 Measuring procedure 58

5.7 Other investigated methods, variables and approaches 59

5.7.1 Magnetic susceptibility 59

5.7.2 Headspace gas GC 65

5.7.3 Model oils 66

5.7.4 Biodegradation 68

6 Multivariate analysis/chemometrics 71

6.1 Object space and variable space 71

6.2 Pretreatment 73

6.2.1 Centring 73

6.2.2 Block normalization 74

6.2.3 Logarithmic transformation 74

6.3 Principal Component Analysis 75

6.3.1 Latent variables 75

6.3.2 PCA 75

6.3.3 Scores and loadings 76

6.4 Multivariate calibration 78

6.4.1 Partial Least Squares Regression (PLSR) 78

6.4.2 External validation 79

7 Results from Papers **83**

7.1 Research report 83

7.2 Paper I 84

7.3 Paper II 86

7.4 Paper III 87

7.5 Paper IV 88

8 Concluding remarks and further work **89**

Bibliography **91**

Appendix **97**

Appendix A FT-IR spectra 97

Appendix B. WOGC chromatograms 108

Appendix C Model Characteristics 120

Papers **193**

Paper I

Paper II

Paper III

Paper IV

Chapter 1

Introduction and outline

The aim of this work is to improve the general understanding of how the distribution of the different compounds influence the physical properties of crude oil, in particular the properties that are involved in Multiphase Flow Metering, with permittivity as the most important. In crude oil production monitoring, Multiphase Flow Meters are used for online monitoring of the oil, water and gas flow in pipelines, where the output of the measurement is the volume or mass of each phase passing the flowmeter in a given time span. Data from multiphase meters help in optimizing petroleum production, increasing the oil recovery and lowering the investments and operational costs (Falcone e. al., 2011; Thorn et. al., 1999). Permittivity measurements provide input for the determination of flow rates and relative distributions of the fluids in several well established measurement technologies (Falcone et. al., 2002). Such systems are, however, calibrated to the initial oil composition, and may lose accuracy over time due to changes in the fluid compositions that result in changes in the actual permittivities relative to the incorporated calibration values. Monitoring the permittivity regularly is one approach to quality assurance of the meter readings, but the necessary instrumentation is often not easily available. Quality control based on monitoring the fluid composition using standard crude oil analytical data provided by generally available analytical instrumentation, in combination with multivariate modelling of the required parameters is thus an attractive alternative.

By improving the understanding of how the different compounds in crude oil influence the physical properties of crude oil, the project seeks to improve the

performance of the Multiphase Flow Metering systems, both for initial calibration of the systems and for recalibrations during the lifetime of the systems.

In this context, Partial Least Squares calibration models of the composition and the different properties of crude oil can be used as a tool to identify significant changes in oil composition and properties, for example during the production lifetime of an oil field, and hence highlight the need for updating calibration values in Multiphase Flow Meters.

This dissertation is divided in two main sections; one section gives the theoretical background of the measurement principles and analytical techniques used in this work as well as a summary of the results obtained in the papers is given, the other section is the papers, which present the main results of this work.

In chapter 2 some theory is given on the complex mixture which is the subject of analysis in this work; petroleum, or crude oil as it also is called. Then in chapter 3 and 4, the two main contributors of data to the multivariate analysis, Infra Red Spectroscopy and Whole Oil Gas Chromatography, are introduced.

In chapter 5, the principles behind the analysis regarding the chemical composition and physical and electrical properties of petroleum are given, while in chapter 6 a short presentation of theory regarding multivariate analysis and calibration modelling is given.

Chapter 7 gives a brief summary of the results presented in each paper, and in chapter 8 some concluding remarks and suggestions for further work are given.

The last section is the papers, which present the main results of this work.

Chapter 2

Petroleum/crude oil

Petroleum, or crude oil, is a naturally occurring complex mixture of hydrocarbons in a gaseous, liquid or solid state. Petroleum is typically trapped in porous rock deep under ground, both on land and at sea in locations throughout the world (Speight, 1980; Speight, 1998). Petroleum is generally considered a cheap resource for hydrocarbons, and the usage areas for the refined products are wide; petrol for vehicles, diesel fuel, jet fuel, fuels for ships, factories and central heating, paraffin and gas for heating and cooking, bitumen for roads and roofing, lubrication oils, waxes, polishes, and various chemicals are some general usage areas. As petroleum is generated over a long period of time, it is not regarded as a renewable energy source. This means that at some point in time, the petroleum reserves will be depleted.

2.1 Petroleum formation

Petroleum is formed from buried marine sediments, which has been experiencing high temperature and pressure for millions of years. A quick oil generation extends over 5-10 million years, whereas a slow generation may cover over 100 million years (Tissot and Welte, 1984). The organic material that is the origin of the petroleum can be bacteria, plankton, animals, fish and vegetation. Over a long period of time, this material gets buried by the geological activity of the earth. The organic material then undergoes a series of concurrent and consecutive chemical reactions collectively called *diagenesis*. These chemical reactions occur under the influence of the

temperature and pressure conditions, as well as the and long reaction times, at the reservoir locations. The conditions at each reservoir location may vary greatly, and the proportions of the different organic material differ from location to location around the world. Because of this, the petroleum from different petroleum reservoirs has different chemical composition and properties (Speight, 1998).

2.2 Chemical composition and properties of petroleum

Petroleum is a very complex mixture of paraffinic, naphthenic and aromatic hydrocarbons as well as compounds containing heteroatoms like nitrogen, oxygen and sulphur. Traces of a variety of compounds containing metals may also be present.

Although the conditions and base material of which petroleum is formed varies greatly, the elemental composition of petroleum has a rather narrow spread:

Carbon	83.0-87.0%
Hydrogen	10.0-14.0%
Nitrogen	0.1-2.0%
Oxygen	0.05-1.5%
Sulphur	0.05-6.0%

However, the properties of petroleum are not determined by the total elemental composition, but rather the ratios of the different hydrocarbon compounds and polar heteroatom compounds in the petroleum (Speight, 1998). Given the fact that the number of isomers (molecules having the same atomic formula) increase rapidly with increasing carbon atoms, the variance in the properties of petroleum is easier to understand:

Carbon atoms per hydrocarbon	Number of isomers
4	2
8	18
12	355
18	60523

In theory, the heavier fractions of petroleum can contain almost unlimited numbers of molecules, since the molecular weights of the molecules found in petroleum can range from that of methane (molecular weight=16) to several thousand.

The spread in elemental composition in petroleum is narrow, in particular for carbon, so the cause of the major differences between petroleum from different reservoirs is the contents of hydrogen and heteroatoms.

The specific gravity of petroleum varies from 0.75 to 1.00, with most of the crude oils falling in the range between 0.80 and 0.95. The viscosity of petroleum can vary greatly as well, as it can vary from free flowing to having difficulty in being mobile at room temperature. The viscosity can range from below 1 mPa s (milliPascal seconds) for the lightest crude oils to 10000 mPa s for the heaviest crude oils (Attanasi et al., 2010).

2.3 Petroleum types; condensates, biodegraded oil and nondegraded oils

In this project a distinction is made between *condensates*, *biodegraded oils* and *nondegraded oils*. This distinction is made as these three types of petroleum have different chemical composition.

Throughout the world, there are a wide variety of natural gas occurrences which may vary greatly in composition and modes of formation. In natural petroleum gas reservoirs, methane (CH_4) is the major constituent of the gas. Other gas constituents are the heavier hydrocarbons (ethane C_2H_6, propane C_3H_8, butane C_4H_{10}), carbon dioxide CO_2, hydrogen sulphide H_2S, nitrogen, hydrogen, argon and helium. In addition, liquid hydrocarbons are dissolved in the gas. These hydrocarbons condense as the gas mixture reaches the surface during production, and the petroleum is then called a *condensate* (Tissot and Welte, 1984). This type of petroleum consists mainly of lighter hydrocarbons, as they have lower boiling points and hence are more likely to dissolve in the gas phase in the petroleum reservoir.

In some areas where surface-derived, meteoric formation water is present, microbial alteration of crude oil, i.e. *biodegradation* of crude oil, can occur. Biodegradation of crude oil is a selective utilization of certain types of hydrocarbons by microorganisms. The sequence in which the selective removal of hydrocarbons occurs is as follows: *n*-alkanes (below *n*25), isoprenoid alkanes, low- ring cycloalkanes and aromatics. The degree of alteration depends on the different intensities and duration of the biodegradation; the more biodegraded a crude oil is, the higher the content of Nitrogen, Sulphur and Oxygen compounds. The aromatic character of the oil also increases with higher degree of biodegradation (Tissot and Welte, 1984). The different effects of biodegradation upon crude oils are well documented in the literature (Evans et al., 1971; Bailey et al., 1973, Deroo et al., 1974, Connan et al., 1975). Petroleum biodegradation is primarily a hydrocarbon oxidation process, producing CO_2 and partially oxidized species like organic acids (Peters et al., 2005). As organic acids are produced by the biodegradation process, a method for identifying biodegraded oils can be to measure the acid content of the oil. However, oils that are not biodegraded can also contain acids, although not in the same amounts as for biodegraded oils.

In figure 2.1 (from Head et. al.) a schematic diagram of physical and chemical changes occurring during crude oil and natural gas biodegradation is shown.

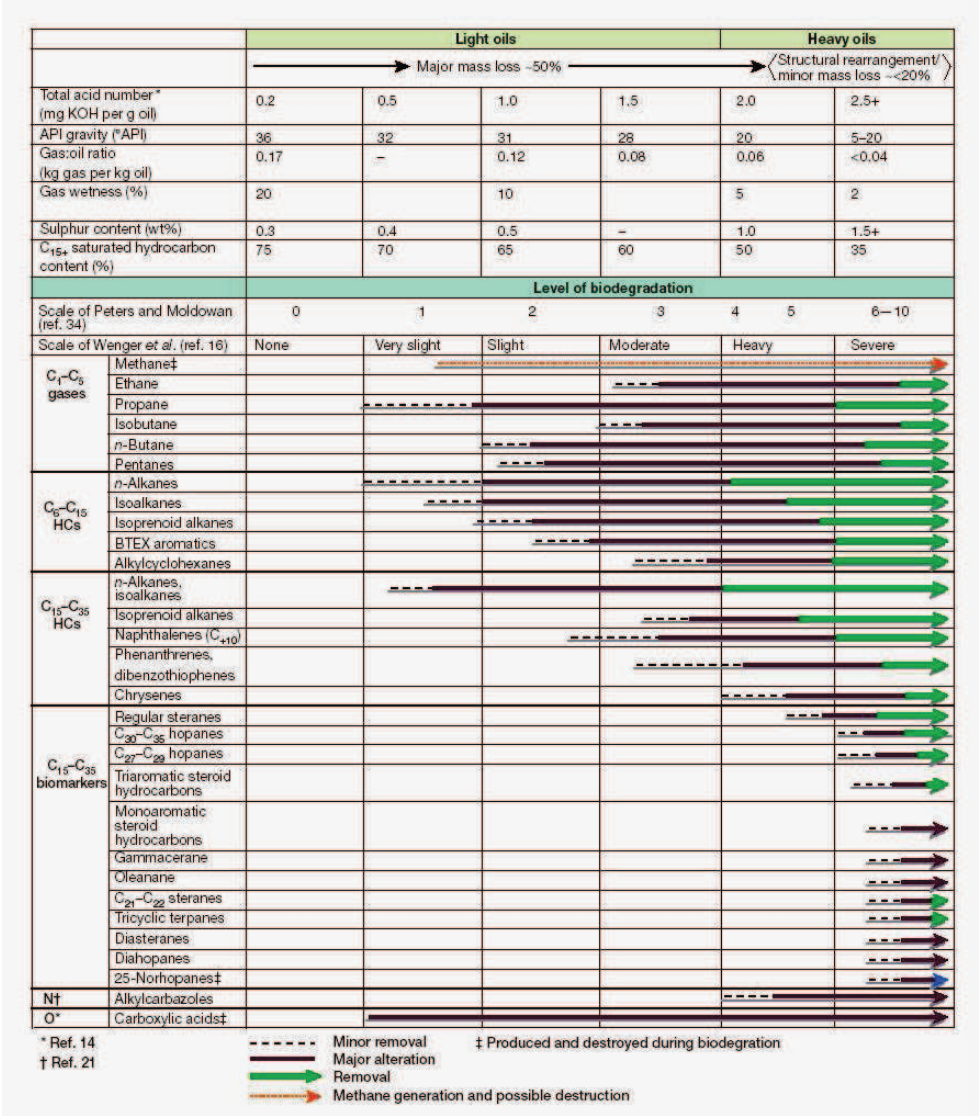


Figure 2.1: Schematic diagram of physical and chemical changes occurring during crude oil and natural gas biodegradation, from Head et.al. References 14, 16, 21 and 34 are, respectively, Peters et. al. (2002), Hunt (1979), Thorn et. al. (1998) and Pepper et. al. (2001).

Although the chemical composition of condensates, biodegraded and nondegraded oils are different, the variation in some of the physical properties of biodegraded and

nondegraded crude oils are somewhat similar and cover a continuum rather than comprising completely separate classes. Both biodegraded and nondegraded oils can vary from light to heavy crude oils; light oils have low densities (compared to heavy oils) and low viscosities, while heavy oils have high densities and high viscosities. No clear density limits are absolutely defined regarding separating light and heavy crude oil, but as a general reference we can say that heavy crude oils have densities higher than 0.932 g/cm^3, while light crude oils have densities lower than 0.853 g/cm^3 (Baboian, 2005). As for the viscosity, light crude oils have values in the order of around10^1 mPas, while heavy crude oils have values in the order of around 10^3 mPas (Speight, 1998). In general, lighter crude oils are more valuable than heavier oils, as they are easier to produce and require less treatment in order to give marketable products.

2.4 Production and transport of petroleum

Petroleum is produced by drilling wells down to the reservoir location, and then extracting the oil, gas and water present. There are many ways in which to produce the petroleum, some methods and means are: natural flow, applied flow, mechanical lift and water injection. For more information regarding these methods see e.g. Speight (1980).

Petroleum reservoirs can be located both offshore and onshore, where typically production of the offshore reservoirs is more challenging than the onshore ones. The remote locations and harsh environments are the main cause of this, and the depths that are involved can cause additional challenges, as the sea depth can be several hundred meters.

Once the petroleum is extracted from the well, it needs to be transported further on to refineries, in order to be separated into marketable products. The transport is typically done by tankers or pipelines.

2.5 Flow assurance; Multiphase Flow Metering (MFM)

The flow through the transport pipelines can be a mixture of water, oil and gas, as well as solids, either dissolved in the fluid or flowing freely as separate phases. The flow is therefore determined as *multiphase*. As the flow moves from reservoir to the surface, the temperature and pressure conditions experienced by the flow may change significantly. This can cause the mixture to destabilize; solid particles like gas hydrates, asphaltene, wax, scale and naphtenates (Gao, 2008) may separate from the mixture and deposit on the pipeline. This is especially an issue in deep water and cold weather conditions, as the temperature in the pipelines will be significantly lower than in the reservoir, increasing the possibility of destabilizing the petroleum. The deposition of solids can in worst case lead to plugging of the pipelines, which is an expensive situation since the production must be stopped while resolving the issue. As such, *flow assurance* is an important field in the petroleum industry. A variety of additives and inhibitors can be utilized in order to prevent unwanted issues as mentioned above. Some inhibitors are against corrosion, scale, gas hydrates, clay swelling and emulsions (Fink, 2012).

As a mean of detecting possible threats to the flow assurance and in order to monitor the flow in petroleum pipelines, multiphase flow metering (MFM) systems can be installed. These systems are placed directly at the pipelines, be it subsea at the sea floor, topside at platforms or at other locations at the petroleum transport route to the refineries. They give measurements of the flow rates of petroleum, gas and water in the pipelines, and can help improving the reservoir management, optimize petroleum production, increase oil recovery and lower investments and operational costs (Falcone et al., 2011; Thorn et al., 1999). MFM systems can also be important in detecting changes in oil composition over the production lifetime of a well.

One of the techniques that are used for measuring multiphase flow is *permittivity*, which is explained in more detail in chapter 5.6.

2.6 Representative oil sampling

The petroleum samples used in this project are stored in glass containers (250ml- 2L) in a dark, cool place when not in use, and with a topping of nitrogen gas in order to prevent evaporation and removal of gas upon opening of the containers. Some oils contain waxes, and these may precipitate and stick to the glass walls of the containers during storage. In order to dissolve these waxes the oil is placed in an oven at 60 °C for 4 hours, the oil is then mixed thoroughly in order to homogenize the oils. This procedure is done prior to all oil measurements when the oil has been in storage for more than a couple of weeks.

In this project we have been able to provide a total of 20 crude oil samples through the industry partner Roxar and Christian Michelsen Research, as well as some samples that are in fact residuals (unaltered and/or unused) from other PhD projects at the University of Bergen. This is a recurrent challenge in petroleum chemistry; even with industrial partners in the project group large datasets of crude oils are difficult to achieve. However, datasets with around 20 samples is considered acceptable in most journals. For example, 22 oils is used in Satya (2007), 17 oils is used in Talita (2011) and 30 oils is used in Abbasa (2011).

Chapter 3

Infra Red (IR) Spectroscopy

Infra Red (IR) Spectroscopy is an analytical technique in which IR light is used in order to identify functional groups and specific bond types in a sample. It can be used both qualitatively and quantitatively.

3.1 Vibrational spectroscopy

IR light is a part of the electromagnetic spectrum, which is a continuum of all electromagnetic waves sorted by frequency and wavelength. IR light consists of waves with wavelengths between 2.5 and 25 micrometers (μm) as shown in figure 3.1 (Pavia et al., 2001).

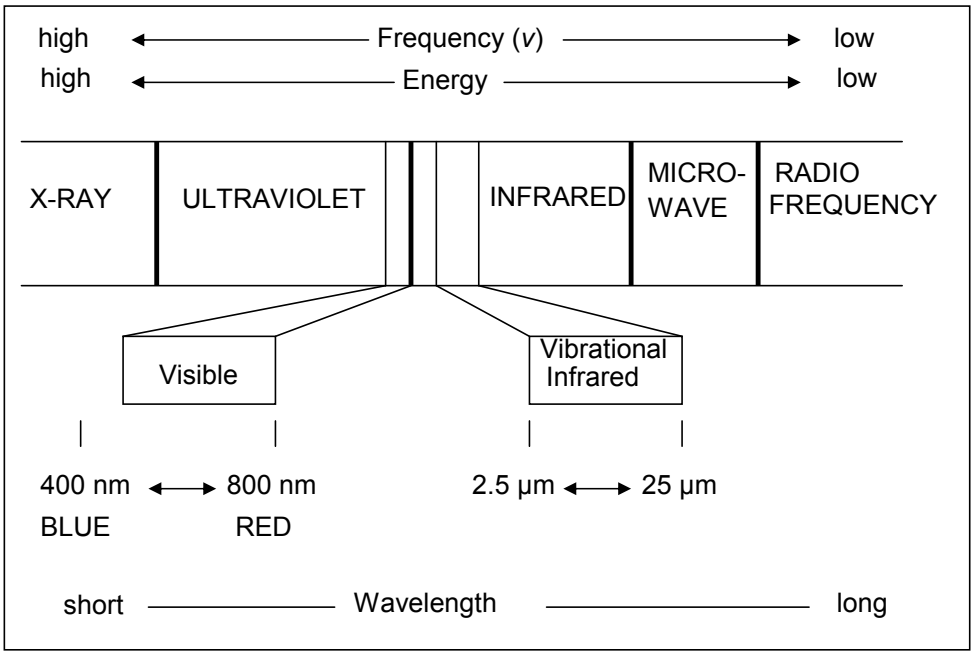


Figure 3.1: A portion of the electromagnetic spectrum, showing the relationship of the vibrational infrared to other types of radiation.

The energy in the electromagnetic spectrum is given by the Planck relation (Planck 1901), where the connection between the wavelength and frequency of the radiation is explained as:

$$E = h\nu = \frac{hc}{\lambda} \quad , \tag{3.1}$$

where h=Planck constant=6.63×10^{-34}Js, c=velocity of light, λ=wavelength, and ν=frequency.

The bonds that hold molecules together are never completely still, but vibrate continuously. When irradiating a molecule with light, the bonds will absorb energy, or light, with specific wavelengths. In the infrared area of the electromagnetic

spectrum, the absorption occurs as vibrations in kind of stretching and bending of the bonds.

The energy of molecules can be divided as follows:

$$E_{molecule} = E_{translation} + E_{rotation} + E_{vibration} + E_{electronic} \tag{3.2}$$

In vibrational spectroscopy $E_{rotation}$ and $E_{vibration}$ is studied, and each of these can be divided into several vibration levels. In these levels the molecule has increased energy, but the electrons are not excited. These vibrational levels, E_0=ground state, and E_x=increased energy, can further be divided into several rotational levels.

Transitions between E_0 and E_x happens when energy equivalent to $\Delta E=E_x-E_0=h\nu$, is either absorbed or released from the molecule. If the molecule is irradiated with light of a specific wavelength $\lambda=c/\nu$, the molecule will absorb the energy and its inner energy will increase from ground state to the increased energy level E_x. The same amount of energy will be emitted from the molecule when the irradiation stops.

If the frequency between two vibrating nucleus coincide with the electric field from submitted electromagnetic radiation and the vibration results in a net change in the molecular dipole moment, infrared radiation will be absorbed. This means that homonucleus molecules like H_2 and N_2, which have symmetrical charge distribution, does not absorb radiation in the infrared part of the electromagnetic spectrum. Further, a molecule with strong dipole moment absorbs more infrared radiation then a molecule with a weak dipole.

The infrared waves do not contain enough energy to excite electrons in chemical compounds, but can make the atoms vibrate. A molecule that absorbs electromagnetic radiation will increase its inner energy, which needs to be released in order to return to the stabile ground state. This energy is released by bending and stretching the bonds of the molecule.

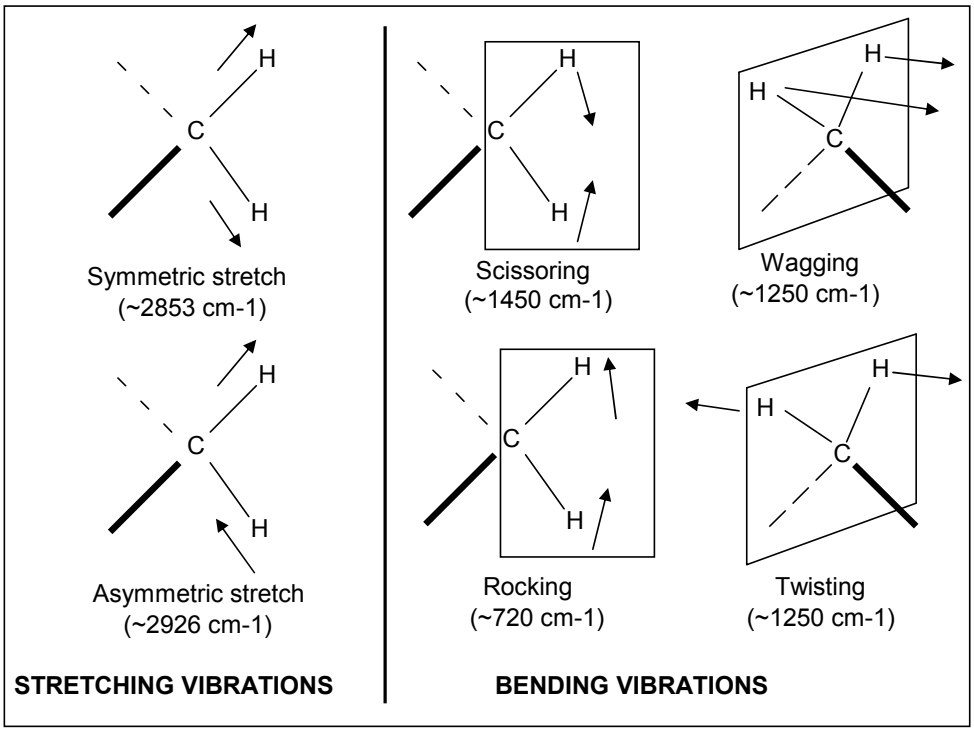


Figure 3.2: Different types of vibration in molecules.

Table 3.1 shows an IR correlation chart, where some general absorption peaks for common types of atomic bonds and functional groups are listed. The ability of IR spectroscopy to detect certain functional groups and atomic bonds is an important tool for identification of unknown samples, but is often not enough by itself for a definitive structure allocation.

Table 3.1: Simplified IR correlation chart, showing absorption areas for some functional groups and bonds. The intensity of the peaks is given as strong (s), medium (m) or weak (w) (Pavia et al., 2001).

Functional group/bond		Type of vibration	Frequency (cm -1)	Intensity
C-H	Alkanes	Stretch	3000-2850	s
	-CH3	Stretch	2962 and 2872	s
	-CH2	Stretch	2926 and 2853	s
	-CH3	Bend	1450 and 1375	m
	-CH2	Bend	1465	m
	Alkenes	Stretch	3100-3000	m
		Out-of-plane bend	1000-650	s
	Aromatics	Stretch	3150-3050	s
		Out-of-plane bend	900-690	s
C=C	Alkanes		1680-1600	m-w
	Aromatic		1600 and 1475	m-w
C=O	Aldehyde		1740-1720	s
	Ketone		1725-1705	s
	Carboxylic acid		1725-1700	s
	Ester		1750-1730	s
	Amide		1680-1630	s
O-H	Alchohols, phenols			
	Free		3650-3600	m
	H-bonded		3400-3200	m
	Carboxylic acids		3400-2400	m
N-H	Amines and amides			
		Stretch	3500-3100	m
		Bend	1640-1550	m-s

3.2 Absorption of infrared energy in Attenuated Total Reflection (ATR) FT-IR

In ATR FT-IR, a beam of infrared light is emitted through a crystal or diamond and into the sample. The sample will attenuate the beam in the areas of the infrared spectrum where the sample absorbs energy, and the changes in the beam are measured by the detector. Figure 3.3 shows an ATR reflection.

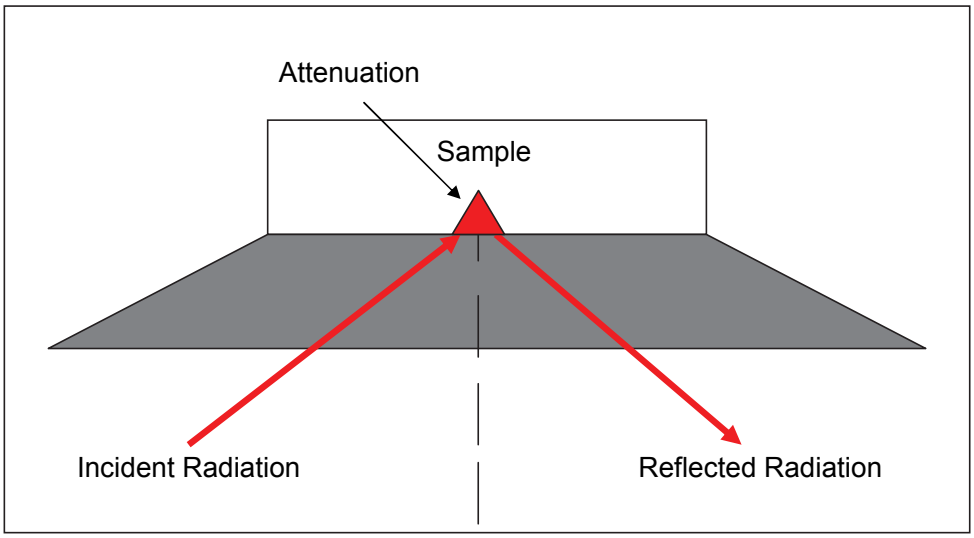


Figure 3.3: ATR reflection.

The amount of radiation that is reflected from a sample is given by:

$$R = \frac{I}{I_0}, \tag{3.3}$$

where R= reflectance, I= intensity of the light that hits the detector after reflecting of the sample and I_0= intensity of the light that hits the detector without being hindered by the sample.

The amount of absorbed radiation by a sample is given by:

$$A = \log\left(\frac{1}{R}\right) = -\log(R) = \log\frac{I_0}{I} \tag{3.4}$$

By inspecting Beer- Lamberts law (Svanberg, 2004):

$$A = \varepsilon \times C \times l \ , \tag{3.5}$$

where A= absorbance, C= concentration (mol/l) and ε= molar absorptivity ($\frac{1}{M \times cm}$), we see that the response in a spectrum is dependent on the concentration of analyte in the sample, and that quantitative analysis is possible using IR spectroscopy.

IR Spectroscopy can be defined as the study of absorption or emission of electromagnetic radiation in the infrared region of the electromagnetic spectrum.

IR spectroscopy is widely used in qualitative and quantitative analysis (FitzPatrick et al., 2012; Profeta et al., 2011; Parisotto et al., 2010; Peinder et al., 2009; Bak and Larsen, 1994; Karstang et al., 1991).

3.3 The IR Spectrophotometer: ATR- FT- IR

The most important units of an IR spectrophotometer are the light source, the wavelength separator, the detector and the computer. The light source, for example a Helium- Neon laser emitting monochromatic light, emits a beam which then enters the wavelength separator, for example an interferometer. In the interferometer the light is divided in two separate beams, where one of the beams is reflected of a mirror that is fixed in place while the other beam reflects of a mirror that is in motion. The beams now have different phase and intensity, and are recombined, resulting in a signal which is called an *interferogram* as the two beams are "interfering" with each other, cancelling some waves and enhancing other waves. The beam then enters the sample compartment, for example an ATR measuring cell, where some of the energy in the beam is absorbed by the sample. The changes in the beam are then measured by the detector, for example a thermoelectric Deuterated TriGlycerine Sulfate (DTGS) detector. The measured signal is then sent to the computer and the Fourier Transformation (FT) takes place, producing the final infrared spectrum (Bracewell, 1978; Application note, 2001).

3.4 Measuring procedure

A Nicolet Protegè 460 FTIR Spectrometer with an ATR measuring cell (shown in Figure 3.4), equipped with a diamond crystal, has been used for obtaining FT-IR spectra of the oils in this project. One drop of oil is placed on the crystal, and 32 scans are taken, giving an averaged spectrum. 5 drops has been measured for each oil, and the average of the resulting 5 spectra has been used in the modelling. The data is given as absorbance: A=log (1/R) where R is the percentage reflectance divided by 100.

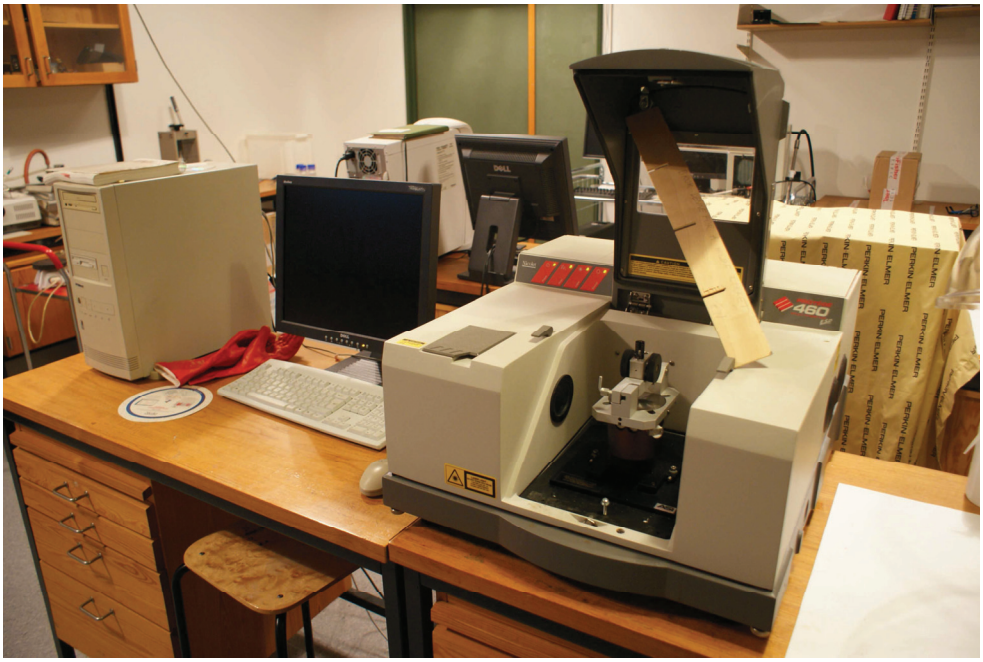


Figure 3.4: Nicolet Protegé 460 FT-IR spectrophotometer with an ATR measuring cell.

3.5 FTIR sample spectra

In this chapter spectra of some of the crude oils in the dataset are shown. Figure 3.5 shows the spectrum of a nondegraded crude oil (labelled 3S in the dataset), Figure 3.6 shows the spectrum of a biodegraded crude oil (labelled 6B in the dataset), and Figure 3.7 shows the spectrum of a condensate (labelled 8C in the dataset). FTIR spectra of all the crude oils are shown in Appendix A.

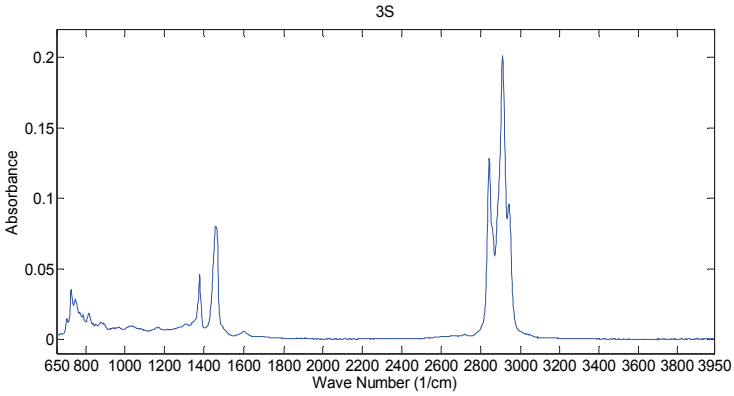


Figure 3.5: FTIR spectrum of the crude oil labelled 3S in the crude oil dataset.

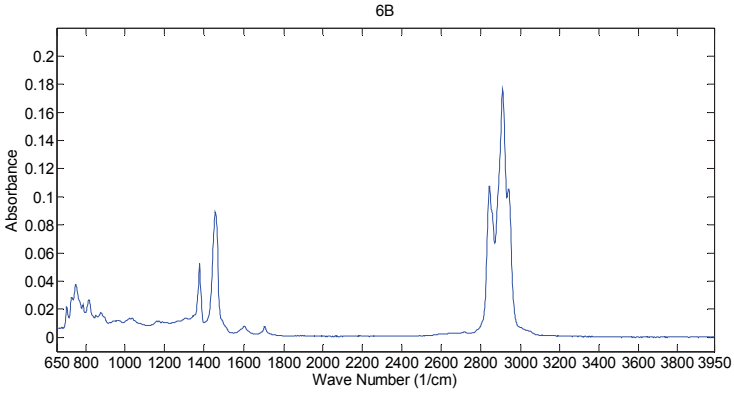


Figure 3.6: FTIR spectrum of the crude oil labelled 6B in the crude oil dataset.

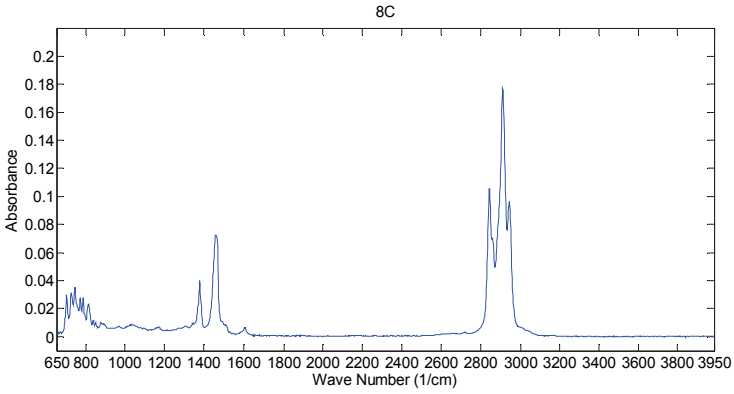


Figure 3.7: FTIR spectrum of the crude oil labelled 8C in the crude oil dataset.

As Figure 3.5- 3.7 shows, the spectra seem to be very similar. By looking closer at the regions 1200-1800 cm^{-1} and 2750- 3050 cm^{-1} we see that there are some variation between the spectra. This variation might be enough to work as a base for multivariate calibration modelling. The variations in these frequency ranges are expanded and shown in Figure 3.8 and 3.9.

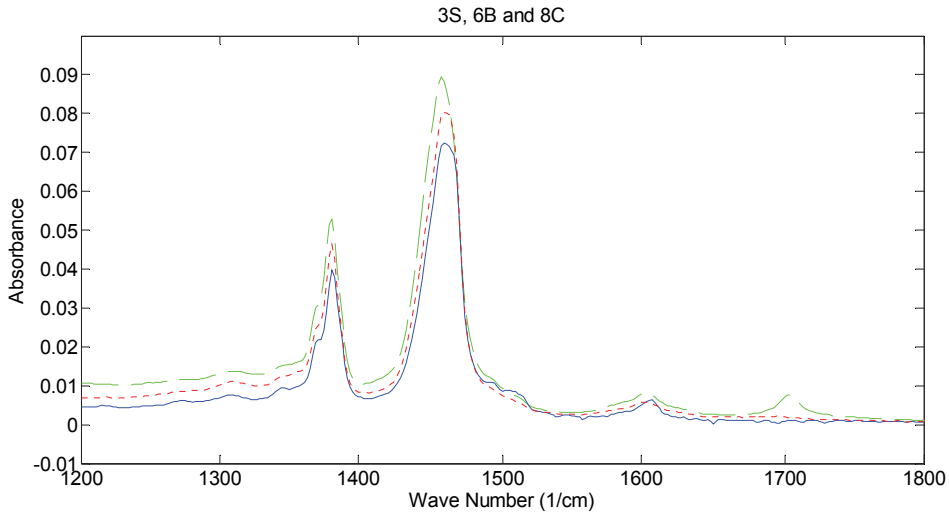


Figure 3.8: Comparison of a biodegraded oil (6B), a nondegraded oil (3S) and a condensate (8C) in the 1200-1800 cm^{-1}- region of the FTIR spectrum. The condensate (8C) is marked as the blue, solid line, the biodegraded oil (6B) is marked as the green, dashed line and the nondegraded oil (3S) is marked as the red, dotted line.

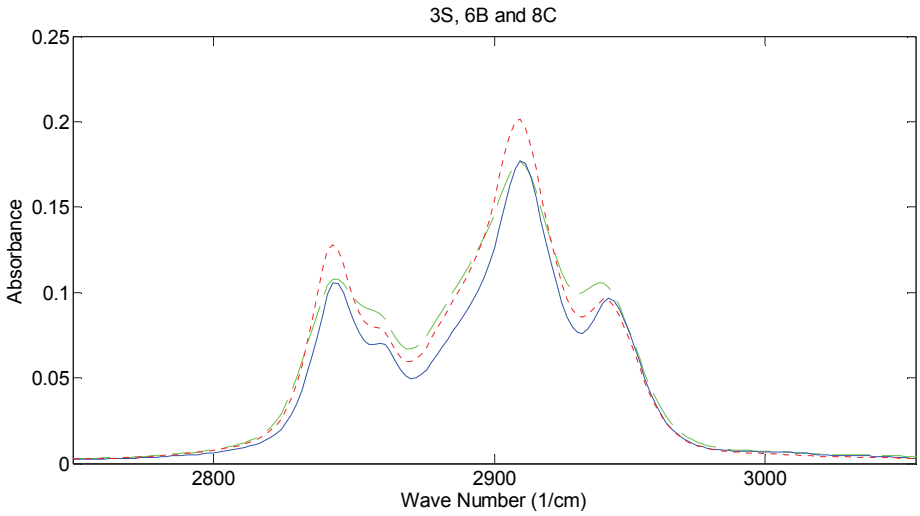


Figure 3.9: Comparison of a biodegraded oil (6B), a nondegraded oil (3S) and a condensate (8C) in the 2750-3050 cm^{-1}- region of the FTIR spectrum. The condensate (8C) is marked as the blue, solid line, the biodegraded oil (6B) is marked as the green, dashed line and the nondegraded oil (3S) is marked as the red, dotted line.

Chapter 4

Whole Oil Gas Chromatography (WOGC)

Chromatography is a laboratory technique for separation of mixtures. The sample mixture is dissolved in a fluid called the *mobile phase*, which carries it through a structure holding another material called the *stationary phase*. The various constituents of the mixture travel through the stationary phase at different speeds depending on the nature of the stationary phase, causing them to separate.

Gas Chromatography (GC) is a common type of chromatography used in analytical chemistry for separating and analyzing compounds in a mixture that can be vaporized without decomposition.

Whole Oil Gas Chromatography (WOGC) is a type of Gas Chromatography where a drop of the oil is analyzed, without solvents. WOGC is typically performed using 30-60 m fused silica columns with bonded apolar stationary phases and analyzes over the range of C4 to C40 (Grob and Barry, 2004).

4.1 Measuring principle

A GC instrument consists generally of three major components; the injection chamber, the column, and the detector.

The sample, typically from 0.1 to 2 μL in volume, is injected into the injection chamber. The temperature in the injection chamber is kept high, around 350 °C, and

the sample is quickly vaporized. The sample is then pushed into (and through) the column by the carrier gas, which is the mobile phase in GC, and is most commonly helium. If the analytes of interest constitute more than 0.1% of the sample, split injection is preferred; otherwise the amount of material can be too much for the column, leading to poor quality chromatograms. In split injection, typically a maximum of 2% of the total vaporized sample is injected into the column. The proportion of the sample that does not reach the column is called the split ratio, and ranges typically from 50:1 to 600:1, meaning that for each portion of sample reaching the column, 50 (or 600) portions does not (Harris, 2003).

The vaporized sample then enters the column, which is a long, thin tube, typically from 15 to 100 m long and with an inner diameter of typically 0.10 to 0.53 mm. The inside of the tube is coated with the stationary phase, typically a film ranging from 0.1 to 5 μm, which is selected based on the polarity of the analyte following the "like dissolves like" rule; Polar columns are best for polar analytes, and nonpolar columns are best for nonpolar analytes. The column is located in an oven, in which the temperature can be controlled and programmed. As the analytes are pushed through the column by the mobile phase (the carrier gas), the analytes are retained by the stationary phase, some more than others. Nonpolar columns cause the separation of analytes to be based primarily by the volatility of the analytes, so the compounds are eluted from the column nearly in order of increasing boiling point (Harris, 2003). To ensure that all the desired compounds elute from the column, a temperature program can be applied. A temperature program makes it possible to keep the temperature constant for a given amount of time, or increasing/decreasing the temperature at a given rate until a given temperature. A mix of both can also be applied.

The eluate exiting the column needs to be detected in order to be quantified and identified. This can be done by a Flame Ionization Detector (FID), in which the eluate is burned in a mixture of H_2 and air. Carbon atoms (except carbonyl and carboxyl carbons) produce CH radicals which are thought to produce CHO^+ ions in the flame:

$$CH + O \rightarrow CHO^+ + e^- \quad (4.1)$$

Electrons flow from the anode to the cathode, where they neutralize the CHO^+ ions in the flame. This current is the current signal. The signal then goes to the computer, where the resulting chromatogram is produced. A chromatogram is a graph showing the detector response as a function of elution time. Different compounds elute at different times called retention times, causing each component to have its own peak in the chromatogram. The area of a peak is proportional to the quantity of that component in the mixture, meaning that quantitative analysis is possible. Integration of each peak gives the relative abundance of each compound in the mixture, by using an internal standard with known concentration and comparing the integrated peak areas, the concentration of the compounds in the sample can be found.

4.2 Measuring procedure

1 μL of the crude oil is injected into the GC instrument, which is a ThermoFinnigan Trace GC equipped with a FID (shown in figure 4.1). The stationary phase is a HP-PONA dimethylpolysiloxane column (50 m × 0.20 mm × 0.5 μm) from Agilent technologies. The mobile phase is helium. The temperature programme is as follows; 30 ºC for 15 min, 1.5 ºC/min up to 60 ºC, 4 ºC/min up to 320 ºC, and in the end 320 ºC for 35 min. The injector temperature is 300 ºC while the FID is kept at 350 ºC, split injection is used with split ratio of 30:1. The program Chromeleon™, version 6.60 (Chromeleon, 2004), is used to obtain and treat the chromatograms.

Each oil is measured two times, and the averaged peak areas for each compound are used in the multivariate analysis and modelling.

In the first chromatogram in a series, the peaks are manually assigned and quantified according to the Norwegian Standard Oil (NSO-1, Weiss et al., 2000). The peaks in chromatograms acquired after the manual assignment will mostly be assigned automatically, but manual inspection of each chromatogram is necessary in order to validate the assignment and quantifications.

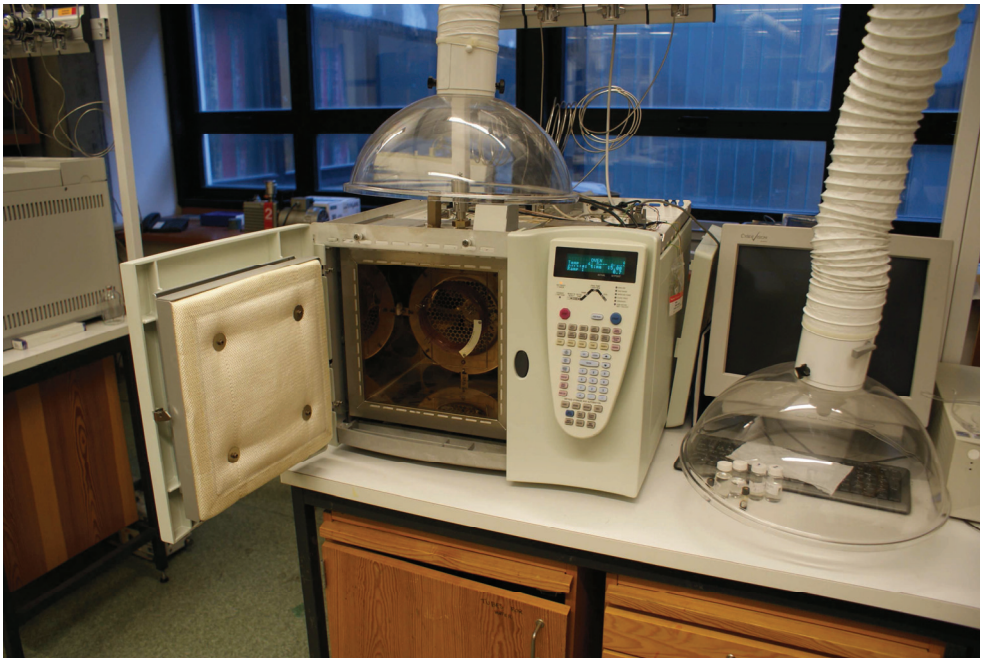


Figure 4.1: ThermoFinnigan Trace GC instrument.

4.3 WOGC sample chromatograms

In this chapter chromatograms of some of the crude oils in the dataset are shown. Figure 4.2 shows the spectrum of a nondegraded crude oil (labelled 3S in the dataset), Figure 4.3 shows the spectrum of a biodegraded crude oil (labelled 6B in the dataset), and Figure 4.4 shows the spectrum of a condensate (labelled 8C in the dataset). The y-axis maximum (voltage detected by the FID detector) varies between the three types of crude oil, where biodegraded crude oils typically have low maximum voltage and condensates typically have high maximum voltage. Chromatograms of all the crude oils are shown in Appendix B.

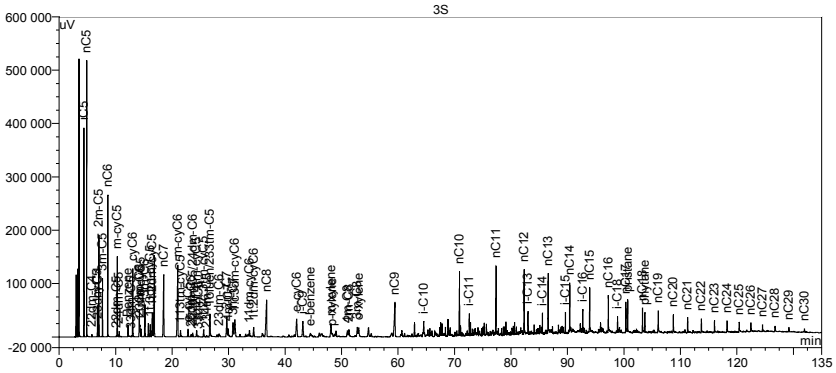


Figure 4.2: WOGC chromatogram of the crude oil labelled 3S in the crude oil dataset.

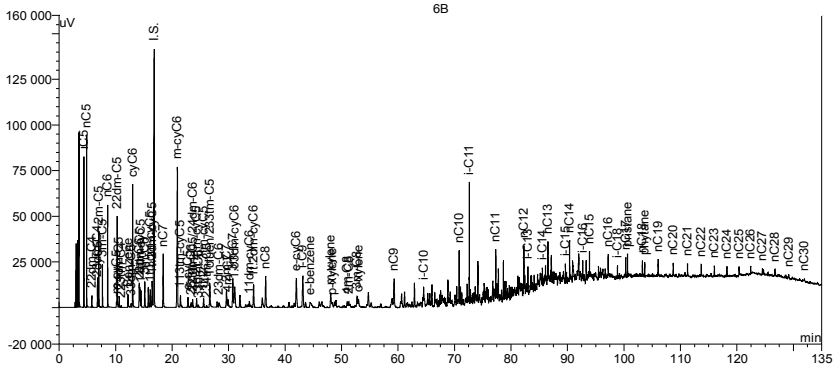


Figure 4.3: WOGC chromatogram of the crude oil labelled 6B in the crude oil dataset.

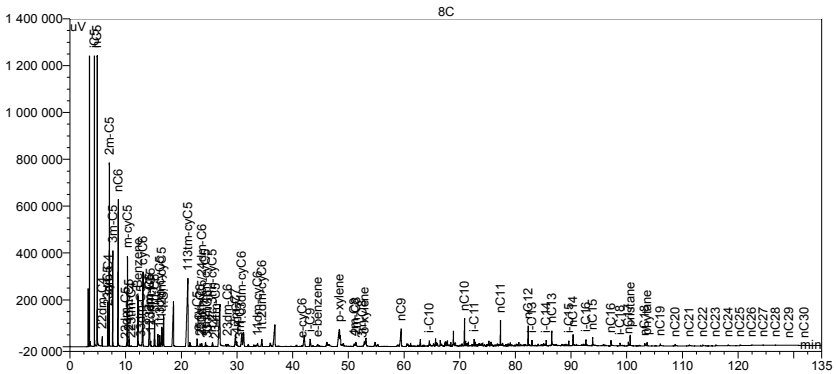


Figure 4.4: WOGC chromatogram of the crude oil labelled 8C in the crude oil dataset.

Chapter 5

Analysis of chemical composition and physical properties

In this chapter a brief introduction is given regarding the different analysis of chemical composition and physical properties of crude oil which has been performed in this project.

5.1 Density

Density ρ is defined as mass divided by volume:

$$\rho = \frac{m}{V} \tag{5.1}$$

ρ=Density, m=mass, V=volume. The unit for density is kg/m3 or g/cm3.

The density of gas and liquids is highly dependent on the temperature; hence density measurement requires an accurate temperature measurement. Density typically decreases with increasing temperature as the individual molecules require more space due to their thermal motion.

5.1.1 Measurement principle

The density of crude oils was measured with a density meter that uses a hollow U tube made of glass. This tube is electronically excited to oscillate at its characteristic frequency, which changes depending on the density of the sample inserted into the U tube (Picker et al., 1974). By measuring the period of the oscillation, the density of the fluid in the U tube can be calculated.

If the oscillator has mass m and volume V, then the frequency is given by equation 5.2 (Operating instructions, Anton Paar):

$$f = \frac{1}{2\pi}\sqrt{\frac{k}{m+\rho V}} \tag{5.2}$$

where ρ is the density of the fluid which fills the oscillator, and k is a power constant. The period, T, then equals:

$$T = 2\pi\sqrt{\frac{m+\rho V}{k}} \tag{5.3}$$

Squaring and simplification of equation 5.3:

$$B = \frac{1}{A} = \frac{4\pi^2 V}{k} \tag{5.4}$$

$$C = \frac{4\pi^2 m}{k} \tag{5.5}$$

The period is then given by:

$$T^2 = B\rho + C \tag{5.6}$$

By combining equations 5.2-5.6, we get the relationship between frequency and density. The following expression gives us information of the difference in density between two samples:

$$\rho - \rho^* = \frac{1}{A}\left(T^2 - T^{*2}\right) \tag{5.7}$$

Where T= the period of the fluid (s), A=instrument constant (g/scm3), ρ= density of the fluid (g/cm3). ρ* and T* denotes a pure solvent.

5.1.2 Instrument and setup

The density of the crude oils was determined using a DMA 60 density meter with DMA 602 measuring cell, produced by Anton Paar K.G. Anton Paar's method for measuring the density of fluids is based on the law of harmonic oscillations (Ozerov and Vorobyev, 2007). Because of its precision and versatility, it has become the standard method for density measurements worldwide.

The oscillator, or measuring cell, which holds about 1ml of fluid, is directly connected to a frequency counter. The reference frequency, 106Hz, is compared to the input frequency of the medium, which can range from 100 to 1000Hz. The measuring cell is placed inside a tube of stainless steel, and is isolated from the rest of the instrument with polystyrene. A thermostat cap regulates the water intake from the water bath to the area around the measuring cell. In order to achieve the best temperature equalization, the inner part of the oscillator is filled with gas with large thermal conductivity.

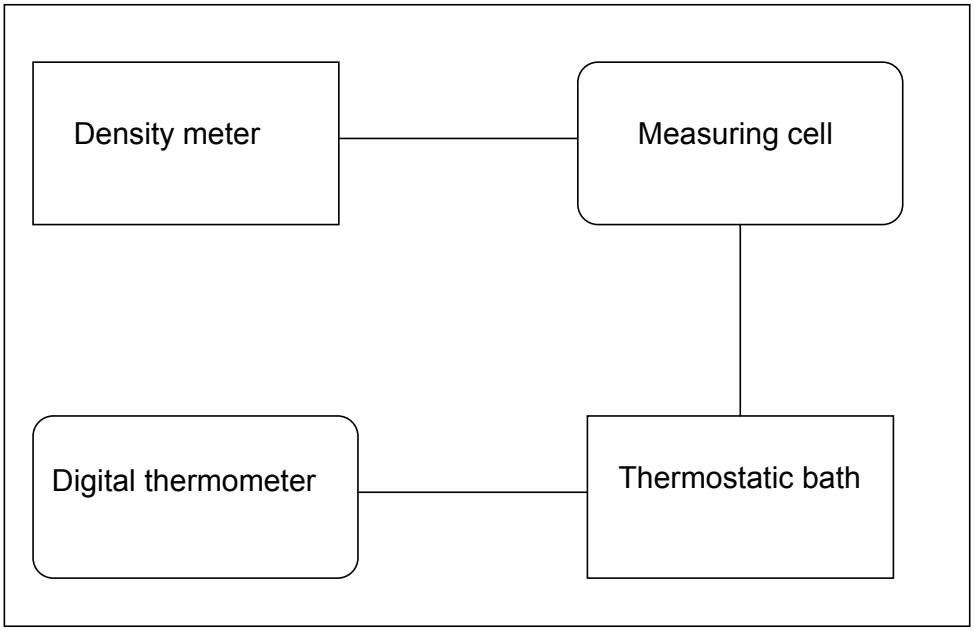


Figure 5.1: Simplified setup for density measurements.

5.1.3 Thermostat control and temperature measurement

The thermostat of the measuring cell was controlled by using a closed system, where the water circulates from the water bath to the area around the measuring cell. The walls of the measuring cell are thin enough to monitor temperature variations with good accuracy. The temperature in the water bath was regulated by a Hetofrig Birkerød thermostat, and a Fluke 2180 digital thermometer, with uncertainty of ± 0,0001°C, was used to measure the temperature in the water bath. The water circulated in the bath, and had a temperature of about 20°C when measuring the temperature. The variation in temperature in the water bath during the measurements was lower than ±0.01°C, and we can assume that the temperature in the measuring cell follows the same pattern.

5.1.4 Calibration and control of the density meter

The instrument constant, A, can be determined by measuring the period, T, of to systems, water and air, where the density of one of the systems is already known. The density of water at 298,15K is 0.997048 g/cm^3 (Del Grosso and Mader, 1972)**.**

The density of air, ρair, expressed as a function of temperature, T, in Kelvin, relative air humidity, B, in %, and air pressure, p, in mmHg, is shown in the equation below (Alagic, 2005):

$$\rho_{air} = 0.46464 \times \frac{p - 0.08987 \times B}{T} \times 10^{-3} \tag{5.8}$$

The instrument constant is then calculated by rearranging equation 5.7.

5.1.5 Measuring procedure

A 3ml sterile plastic syringe is used to inject the sample into the oscillator. The oscillator is washed with the sample that is to be measured, prior to every measurement. In order to avoid air bubbles in the measuring cell, the injections are done very carefully. A light source behind the measuring cell, and a window in front, makes it possible to control the sample inside. After about 8 minutes, the temperature inside the measuring cell is stabilized. The temperature in the measuring cell is expected to be constant when successive readings of the period do not vary more than ±1 by period numbers at 1,0x10^{-6}. The measured period at constant temperature is then inserted into equation 5.7 and the density of the sample is calculated.

Each oil is measured three times, in order to eliminate any effect from inhomogeneous sampling. Between each measurement the oscillator is rinsed with dichloromethane, ethanol, distilled water and acetone. Then it is dried with air from an air compressor.

The period of water and air was measured after each measurement in order to see that it is constant, indicating that the instrument constant (A, equation 5.7) is unaltered.

The measurement system is shown in figure 5.2:

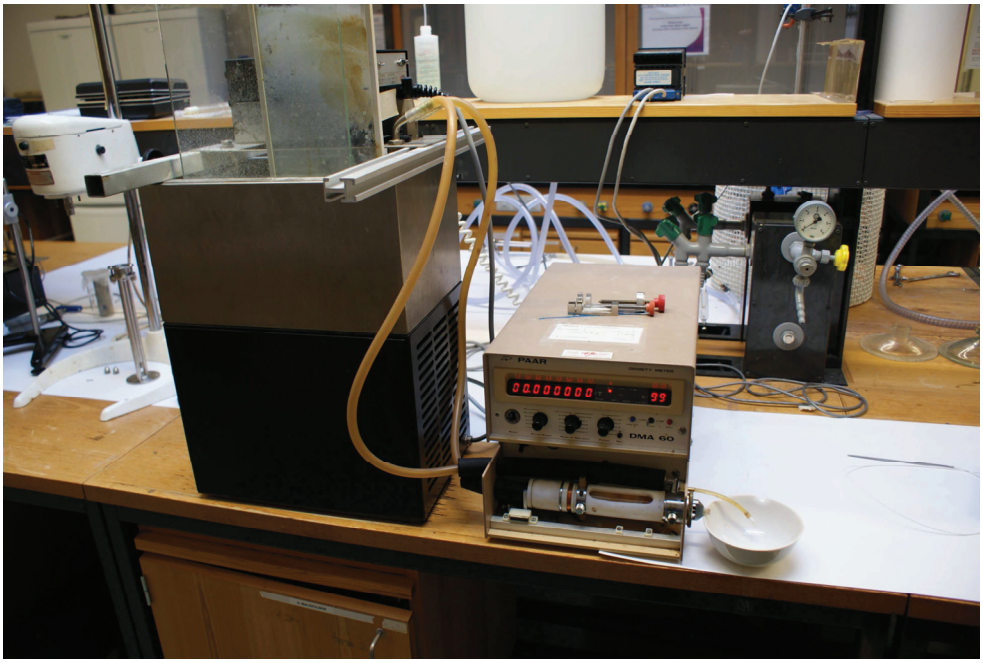


Figure 5.2: Measurement system for density measurements.

5.2 Velocity of sound

Sound travels through materials at different velocities depending on the state (gas, liquid or solid) and the chemical nature of the material. As such, sound has many usage areas for characterization of materials, for example in SONAR systems, medical acoustics as well as engineering acoustics (Rossing, 2007).

5.2.1 Theory

Sound is a longitudinal wave (a wave in which the disturbance occurs parallel to the line of travel of the wave) of pressure that propagates through compressible media such as air and water. The disturbance may for example be created by a vibrating object, and can only be created and transmitted in a medium, such as a gas, liquid or a solid. As the vibrating object moves into the medium, the space directly in front of the object is compressed, causing the pressure to rise slightly. This region of increased pressure is called a *condensation*, and travels away from the object at the speed of sound. As the vibrating object moves back inward, a region where the pressure is slightly less than normal is formed. This region is called a *rarefaction*, and also travels away from the vibrating object at the speed of sound. As the condensations and rarefactions are travelling outward, the individual molecules in the medium are not carried along with the wave. Instead, the molecules collide with each other and pass the condensations and rarefactions forward (Cutnell et al., 2004).

The sound waves can be reflected, refracted or attenuated by the medium during propagation, which can distort the propagating sound. The velocity of sound depends on the density and the bulk modulus K in the fluid (Benenson et al., 2002).

The velocity of sound in fluids is given by:

$$c_{fluid} = \sqrt{\frac{K}{\rho}} \tag{5.9}$$

where c_{fluid}=speed of sound in given fluid (m/s), K=bulk modulus of given fluid (N/m^2) , and ρ=density of the given fluid (kg/m^3) (Halliday et al., 1997;).

The bulk modulus measures a substance's resistance to uniform compression, and varies with the nature of the substance (Cutnell et al., 2004).

Velocity of sound in liquids is in the range of about 1000 m/s to about 2000 m/s, some sound velocities at 20°C are given in table 5.1.

Table 5.1: Velocity of sound in some liquids at 20°C (Cutnell et al., 2004; Benenson et al., 2002).

Liquid	Velocity of sound (m/s)
Chloroform	1004
Ethanol	1162
Methanol	1156
Naphta	1295
Water	1480
Seawater	1522
Benzene	1330
Mercury	1450
Transformer oil	1425

5.2.2 Measuring principle/procedure

The velocity of sound of the oils in this work was measured with a technique used for density measurement in liquids developed by Erlend Bjørndal. This is based on measurement of the liquid acoustic impedance in order to determine the density. A detailed version of the technique can be found in Bjørndal (2007). The measurements were done by technicians at Christian Michelsen Research AS (www.cmr.no).

5.3 Viscosity

Viscosity is a fundamental characteristic property of all fluids. The molecules in a flowing fluid have relative motion between each other, a process that is combined with internal frictional forces. Viscosity is a measure of the resistance to flow or shear (Metzger, 2006; Viswanath et al., 2007).

5.3.1 Theory

In this project, the shear viscosity of the crude oils is measured. Shear viscosity is defined as:

$$\eta = \frac{\tau}{\dot{\gamma}} \tag{5.10}$$

where η= shear viscosity(Pas), τ=shear stress (Pa), and $\dot{\gamma}$=shear rate (1/s) (Metzger, 2006).

Shear stress is defined as:

$$\tau = \frac{F}{A}\ , \tag{5.11}$$

where F= shear force (N) and A= shear area (m^2). Shear stress, τ, has the unit Pa (Pascal). 1 Pa =1 N/m^2

The Two-Plates- Model, shown in figure 5.3, shows how the upper plate with the shear area A is set in motion by the shear force F and the resulting velocity v is measured. The lower plate is stationary. The distance h is the distance between the plates, and the liquid sample is sheared in this gap.

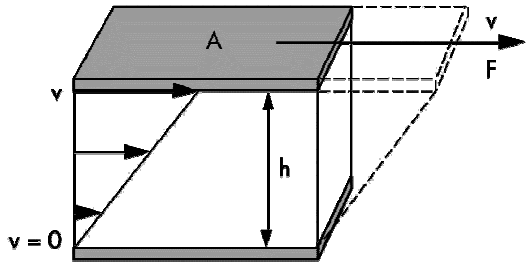


Figure 5.3: Two-Plates-Model (Metzger, 2006).

Shear rate is defined as:

$$\dot{\gamma} = \frac{v}{h} \quad , \tag{5.12}$$

where v= velocity (m/s) and h= distance (m) between the plates (Figure 5.3). The unit of shear rate is 1/s.

For low- viscosity fluids the unit mPas (1000 Pas) is normally used, the unit used earlier was cP("centipoise"), after the doctor and physicist Jean L.M. Poiseuille. 1 cP= 1mPas. Shear viscosity is a material constant, and Table 5.2 shows shear viscosity for some materials.

Table 5.2: Shear viscosity for some materials (Metzger, 2006).

Materials	**Shear viscosity η**
Gases / air	0,01 to 0,02 mPas
Water	1,00 mPas
Milk	2 to 10 mPas
Olive oil	approx. 100 mPas
Gear oil	300 to 800 mPas
Glycerine	1480 mPas (10^{-3} Pas)
Bitumen	0,5 Mpas (10^{6} Pas)

5.3.2 Measuring principle/procedure

The viscosity measurements in this work were done on an Anton Paar MCR 300 Rheometer with a Conical Plate (CP) measuring system at 20°C. The system was calibrated before each measuring series, by measuring air and distilled water at 20°C. If the viscosity of air is measured to be approximately 0.01 mPas and the viscosity of distilled water was measured to be approximately 1.00 (±0.03), the system is calibrated.

The measurements were done on 3ml of oil sample. The shear rate was increasing linearly from 1/s to 1000/s, then back from 1000/s to 1/s, with 32 measuring points each way. This is done in order to detect hysteresis or other effects that may influence the measured viscosity. The measured viscosity at the shear rates from 100/s to 900/s

(both from measuring from 1/s to 1000/s, and from 1000/s to 1/s) has been averaged in order to give the viscosity of the measurement. The measuring points from shear rates between 1/s and 100/s, and between 900/s and 1000/s have been omitted from the averaging, as these measuring points usually are noisy. After each measurement, the measuring cell is rinsed first with dichloromethane (or toluene), then ethanol, then distilled water and in the end acetone. The measuring cell is then dried with air from an air compressor.

In order to eliminate any effects produced by inhomogeneous sampling, two or more measurements is done for each oil, producing an average value and a standard deviation.

The measuring system is shown in figure 5.4:

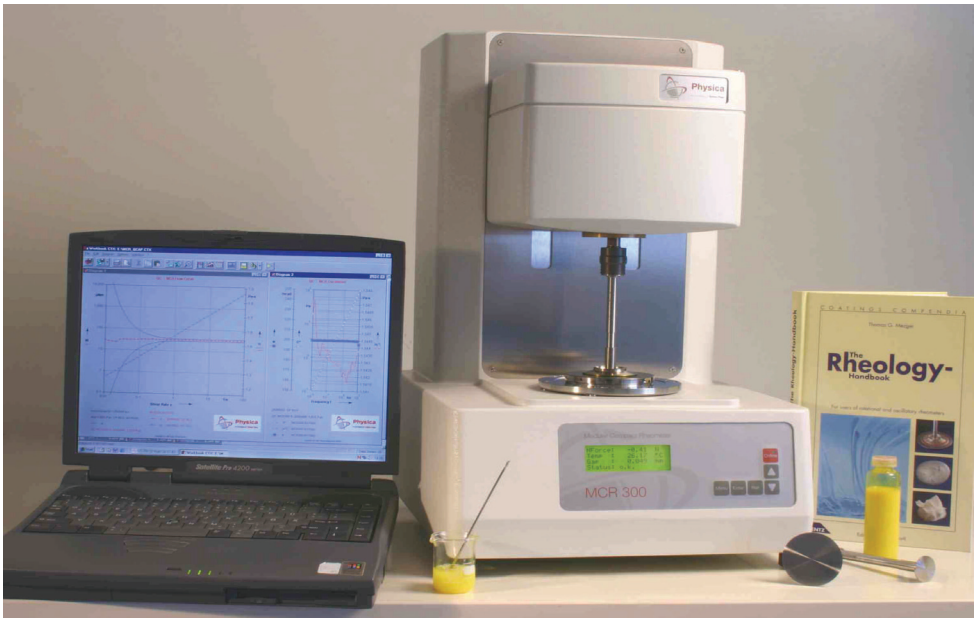


Figure 5.4: Measurement system for viscosity measurements; Anton Paar MCR 300 Rheometer with a CP measuring system.

5.4 Total Acid Number (TAN)

The Total Acid Number (TAN) is a measure of the acidity of a crude oil, and is defined as the mass of a base (Potassium hydroxide, KOH) in milligrams that is required to neutralize one gram of oil. Acid numbers greater than 1.0 is generally considered high, most crude oils have TAN lower than 5 mg KOH/g oil (Sheng, 2011).

5.4.1 Acids in petroleum

One source of acids in petroleum is the sediment in which the organic material was buried in, which can contain both acidic and basic minerals as well as metal containing ores (Speight, 1998). Another source of acids in petroleum is the lipids of many organisms, in which carboxylic acids are the major component. As such they are widespread chemical fossils, occurring in sediments and petroleum oils (Lochte and Litmann, 1955; Albrecht and Ourisson, 1971).

A third source for acids in petroleum is the microbial degradation of the petroleum occurring in the reservoir, which has been associated with increased acidity of the remaining oil phase (Behar and Albrecht, 1984; Wenger et al., 2002; Meredith et al., 2000).

The main cause of acidity in petroleum is the oxygen containing compounds. Some of the acids that are identified are the saturated fatty acids from C_1 to C_{20}, isoprenoid acids, naphtenic acids, cyclopentane and cyclohexane carboxylic acids from C_6 to C_{20}, and the cyclopentyl acetic acids from C_8 to C_{10} (Tissot and Welte, 1984) The most abundant carboxylic acids are the naphthenic and naphthenoaromatic types, followed by polyaromatic and heterocyclic types (Seifert and Teeter, 1970).

Other compounds that contribute to the acidity in petroleum are phenols like the cresols, which are abundant in the acidic fraction of crude oils (Seifert and Howells, 1969).

Figure 5.5 shows molecular structures for some of the acidic molecular types in petroleum.

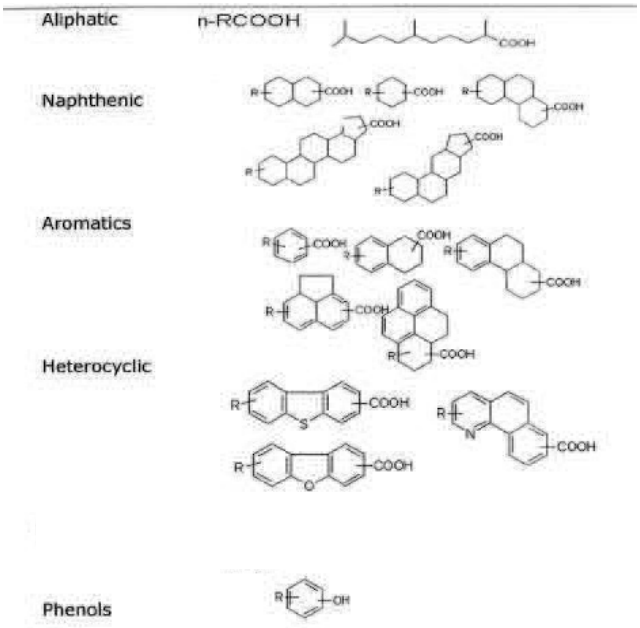


Figure 5.5: Molecular structures for some of the acidic molecular types in petroleum (Moen, 1996).

5.4.2 Measuring principle

TAN is determined by the standard method ASTM- D664, a non- aqueous potentiometric titration technique performed on an autotitrator. Potentiometric titration is a technique where the potential difference between a pair of electrodes immersed in the solution is measured. One of the electrodes displays a constant

potential and is the reference electrode, while the other electrode is the indicating electrode, which responds to the changes in concentration of one of the ions involved in the reaction. The electrodes are connected to a potentiometer, which measures the difference between them. The titrant is added in small increments with continuous stirring, and the potentiometer reading in millivolts is recorded after each addition. The endpoint of the titration is the steepest point of the S shaped plot obtained when plotting the millivolt readings against volume of added titrant (Cullum, 1994). An autotitrator automatically adds titrant, records potential difference, and determines the endpoint of the titration, all based on the settings the user has applied.

5.4.3 Measuring procedure

The following chemicals are needed in order to perform the TAN measurement:

-KOH, p.a.	(approx. 0.05M, 3.0g/L.Mm=56.106g/mol)
-Isopropanol (99.8%), p.a.	(dry)
-$CaCl_2$	(for drying tube)
-Toluene, p.a.	
-Distilled, deionised water	
-Potassium hydrogen phthalate (PHP)	(4 g/200 ml, Mm=204.23g/mol))
-Stearic acid, p.a.	(approximately 0.05 M)
-KCl, p.a.	(saturated solution, 36 g/100 ml. 3M=)
-Acetic acid 99-100%, p.a.	

These chemicals are used in the following solutions:

-Titrant solution

-Electrode solution

-Titration solution (solvent for the oil)

-PHP standard solution

The titrant solution is prepared by weighing 3g KOH and adding 1L dry isopropanol, then the solution is gently boiled with total reflux until all of the salt is dissolved. The solution is then put away to rest for two days, before it is bubbled with nitrogen for 30 minutes in order to remove traces of water, and vacuum filtrated in order to avoid carbonate precipitation. The titrant solution should then be a solution with approximately 0.05M KOH in isopropanol. When the titration solution is connected to the autotitrator, a drying tube filled with $CaCl_2$ is fitted in order to avoid CO_2 dissolving and forming carbonate.

The electrode solution is prepared by dissolving 36 g KCl in 100 ml distilled water, enough to make a saturated solution. There should always be an undissolved amount of KCl in the bottom of the flask. The electrode is stored in this solution when not in use.

The titration solution, used as solvent for the oil, is prepared by mixing toluene, distilled deionised water and dry isopropanol in a 50/0.5/49.5 wt% ratio. If a large amount of samples are to be tested, 4 L of titration solution should be prepared. The amounts is then 2000 g toluene, 16 g distilled deionised water and 1984 g dry isopropanol.

The PHP standard solution is prepared by drying approximately 3 g PHP at 100°C for 3 hours, and cooled to room temperature in a desiccator. About the exact 0.5 g PHP is weighed and dissolved in 100 g acetid acid 99-100%. The molality (mmol/g) of the PHP solution is then calculated.

The stearic acid solution is prepared by weighing approximately 2.85 g stearic acid and dissolving it in 100 ml of titration solution. This solution is used in order to move the endpoint of titration for oils with low TAN value, as we do not want the endpoint of titration to occur below 1 ml of added titrant due to instrument noise in this region. 0.5 g of the stearic solution will move the endpoint of titration with approximately 1 ml.

All of the required solutions for the TAN measurements are now prepared, and the titrant solution (KOH solution) can be standardized. Titration of an exact weighed PHP solution will give an consumption of KOH equal to the amount of PHP + the amount of acid in the solvent: n(KOH)=n(PHP)+n(blank). When n(PHP) is known, both n(blank) and n(KOH) can be determined by two or more measurements by producing a standard curve. The standard curve is produced by doing 5 titrations of PHP with the KOH titrant with 5 different amounts of PHP, for example 1g, 2g, 3g, 4g, and 5g of PHP. Each titration is done by first weighing the exact amount of PHP, and then adding exactly 40 ml of the titration solution. The mixture is then titrated with the standard settings (see end of chapter), and the volume and potential (or pH value) for the endpoint, along with the weight of the PHP, is noted. The data is plotted in a spreadsheet, as volume KOH added at endpoint against mass of PHP weighed, resulting in a regression line explained by for example y=3.1457x+0.6489. n(KOH)=n(PHP)+n(blank) can be written as:

$$C(KOH)\bullet V(KOH,endpoint)=m(KHFT)\bullet M'(KHFT,mmol/g)+n(blank) \quad (5.13)$$

This can be plotted as a straight line, V(KOH) against m(PHP), and give both C(KOH) and n(blank) as results:

$$V(KOH) = [M'(KHFT)/C(KOH)] \bullet m(KHFT) + [1/C(KOH)] \bullet n(blank) \quad (5.14)$$

y = a • x + b

With the example y=3.1457x+0.6489, this gives:

$a=3.1457= [M'(KHFT)/C(KOH)] \rightarrow C(KOH) = M'(KHFT)/3.1457 =0.0158$ M

b = 0.6489 = n(blank)

The concentration of the titrant solution is now known, and the blank value as well. The concentration of the stearic acid solution is acquired by titratrion a known amount of stearic acid solution in 40 ml titratrion solution with the titrant solution, and the concentration is determined directly from the following equation:

$$C(KOH) \bullet V(KOH, \text{end point}) = C(\text{stearic acid}) \bullet V(\text{added stearic acid}) + n(blank) \quad (5.15)$$

The TAN values of petroleum can now be determined by weighing approximately 10 g oil and dissolving it in 40 ml titratrion solution. The mixture is then titrated using the standard settings, if the endpoint comes at less than 1 ml added titrant, stearic acid is added to the solution in order to move the endpoint. The TAN value is then calculated using the following formula:

$$TAN = TA * Mm(KOH)/m(oil), \quad (5.16)$$

where TA= amount of acid in the mixture. If stearic acid is used in the titration, $TA=C(KOH) \bullet V(KOH, \text{endpoint}) - C(\text{stearic acid}) \bullet V(\text{added stearic acid}) - n(blank)$, if no stearic acid is used, $TA= C(KOH) \bullet V(KOH, \text{endpoint}) - n(blank)$.

A minimum of 3 parallels are measured, in order to obtain an average value and a standard deviation.

The electrode is washed after each measurement, first with dichloromethane, then with titration solution and last with distilled water. The electrode is then put in a

storage solution (3M KCl, buffer pH 7) for 5 minutes. The electrode is then washed with distilled water, and titration solution before the next measurement can be started.

When the measurements are finished, the electrode is stored in the electrode solution.

The TAN measurements in this project were done on a Metrohm 798 MPT Titrino Autotitrator, shown in Figure 5.6:

Figure 5.6: Metrohm 798 MPT Titrino Autotitrator.

5.5 Asphaltene content

Asphaltenes are the heaviest constituent of crude oil, and are considered the least valuable component in crude oil. This is due to the fact that they mostly cause problems in transportation and production, as well as it is difficult to convert it to valuable products. However, at the field of highway surface, a derivative of asphaltenes is the material of choice, namely asphalt (Yen and Chilingarian, 2000).

Crude oils have varying amounts of asphaltenes; some oils contain nothing while others have large amounts.

5.5.1 Theory

Asphaltenes are brown to black amorphous materials that are separated from petroleum during petroleum refining. The preferred method for separating the asphaltenes from the petroleum is by addition of *n*-pentane in a ratio of more than 40 times the volume of the petroleum.

The composition of asphaltene is still subject to analysis, because asphaltene does not have one single molecular formula. Due to this fact, asphaltenes does not have a definite melting point, and asphaltenes are actually a solubility class (Speight, 1980). Asphaltenes are soluble in liquids having a surface tension of above 25 dyn/cm, some examples are pyridine, carbon disulfide, carbon tetrachloride, carbon dichloride and benzene. Asphaltenes are insoluble in liquids having a surface tension of below 25 dyn/cm, some examples are petroleum ether, pentane and hexane.

Asphaltenes isolated from petroleum by use of *n*- pentane as precipitation agent gives an elemental variation as follows:

82 ± 3% carbon

8.1 ± 0.7% hydrogen

0.3 to 4.9% oxygen

0.3 to 10.3% sulphur

0.6 to 3.3% nitrogen

The H/C ratio is in the range of 1.15 ± 0.05, giving a very high aromaticity level. As mentioned, asphaltene does not have a single molecular formula, but consist of various aromatic structures like condensed polycyclic aromatic ring systems bearing alkyl side chains. Some proposed structures and origins for asphaltene are shown in figures 5.7- 5.10.

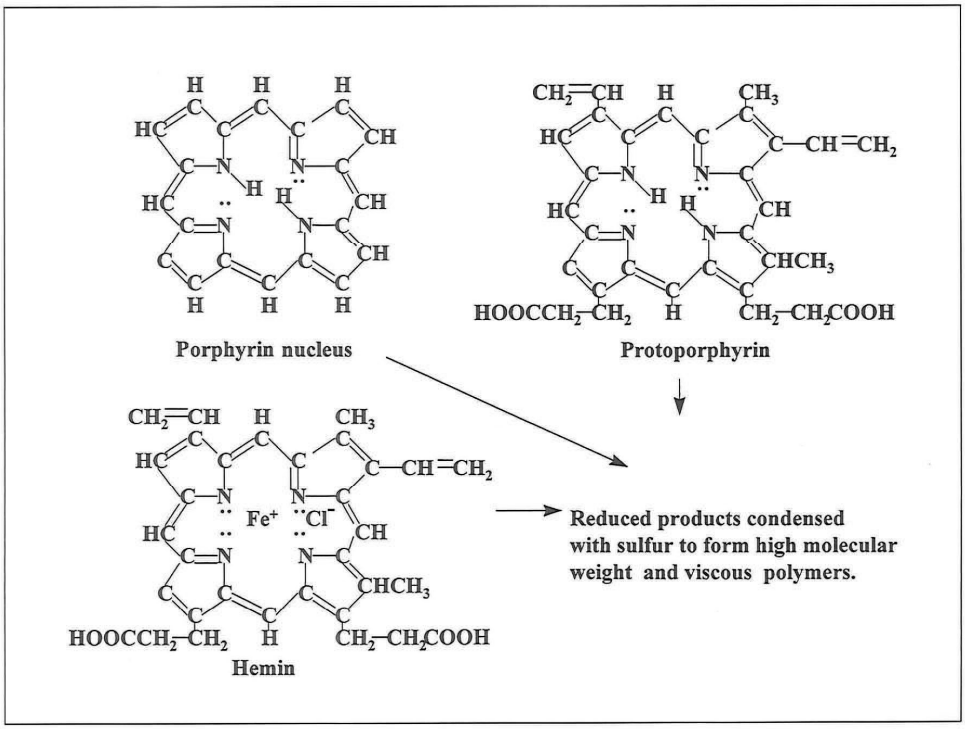


Figure 5.7: Possible origins for asphaltene (Becker, 1997).

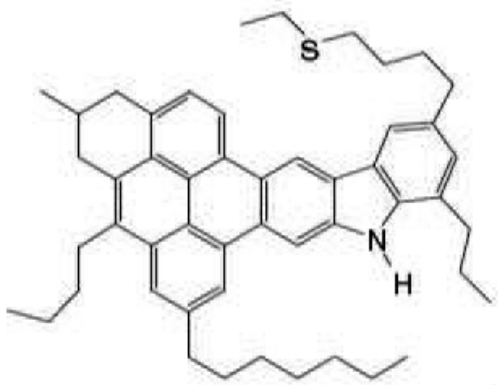


Figure 5.8: Proposed asphaltene structure (Mullins et al., 2007).

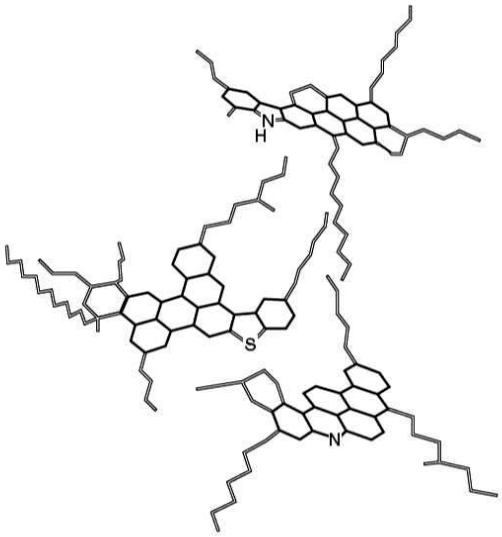
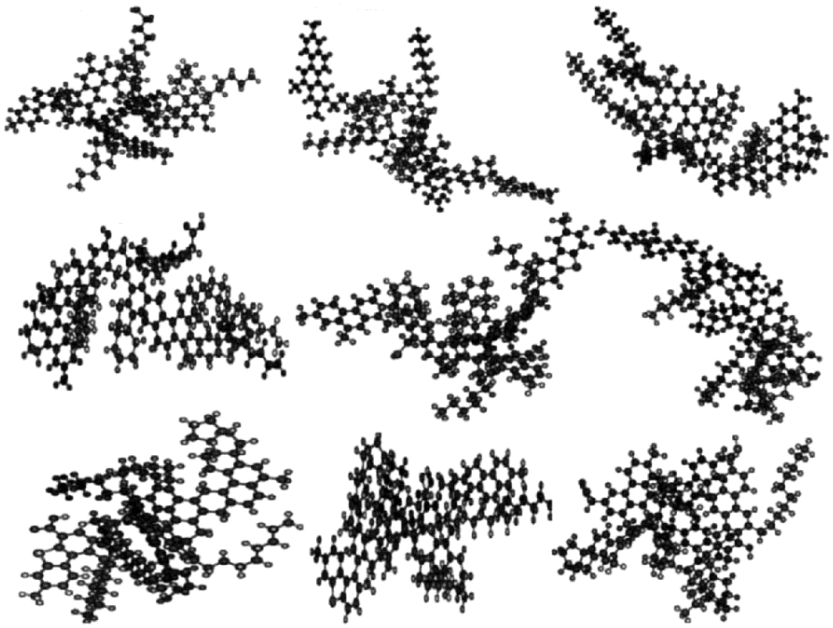


Figure 5.9: Proposed asphaltene structures (Groenzin and Mullins, 2000).



5.10: Three dimensional structures of some Arab model asphaltene structures (Yen and Chilingarian, 2000).

Upon recognition of the complexity of the proposed structures for asphaltene, it makes sense that the molecular weight of asphaltenes has been found to range from about 600 up to 300 000.

Asphaltenes are regarded as the most polar constituents in petroleum (Kontogeorgis and Folas, 2010); hence the relative amount of asphaltene in petroleum influences the physical properties of the petroleum, especially the properties that depend on polarity like permittivity.

5.5.2 Measuring principle/procedure

The asphaltene content of the crude oils in this work has been determined by precipitation with 40 volumes of pentane per volume petroleum by the method described in NIGOGA (Norwegian Industrial Guide to Organic Geochemistry Analyses, Weiss et al., 2000).

0.2 ml crude oil is added to a weighed glass centrifuge tube, weighed again, then 8 ml of *n*- pentane is added to the tube and the solution is mixed thoroughly. The tube is then put in an ultrasound bath for ten minutes, and then placed in a dark, cool place for 24 hours in order to let the asphaltenes precipitate. The tube is centrifuged at 3000 rpm for 5 minutes and most of the supernatant liquid is carefully removed with a Pasteur pipette, in order to avoid removal of precipitate. The precipitate is washed with 5 ml *n*- pentane, mixed thoroughly and centrifuged. The supernatant liquid is removed with a Pasteur pipette. This washing procedure is repeated until the supernatant liquid is colourless. The tube, containing precipitate and a low amount of supernatant liquid, is then dried for 2-3 days before it is weighed. The result is given as mg asphaltene/g crude oil.

5.6 Permittivity

Permittivity is a measure of how an electric field affects, and is affected by, a dielectric medium. A medium is classified as dielectric if it has the ability to store energy when an electric field is applied (Application note, Agilent Technologies Inc.). Permittivity is a fundamental material constant that is specific to a material under given conditions of temperature, frequency and moisture content (Laughton et al., 2003).

Measurement of dielectric properties is one technique that is often utilized in multiphase flow metering (Falcone et al., 2009).

On a molecular level, the main cause of variation in the permittivity of different materials is polarity, or more specific; the strength of the dipoles in the different mixtures or materials. As the different atoms in molecules have different abilities of holding on to their electrons (electronegativity), the charge distribution in the molecule can be uneven. This causes the molecule to have one negative side and one positive side. Some molecules are polar, like water (H_20), and is therefore a permanent dipole (Chang, 2003).

Other molecules are nonpolar, meaning that the charge distribution in the molecules are in balance, like propane (C_3H_8), and are *not* permanent dipoles. However, as the electrons in the molecules move around the molecule, the electrons are for short periods of time concentrated in one area of the molecule, creating a temporary dipole. The temporary dipole then induces neighbouring molecules to become temporary dipoles as well. This effect increases with increasing molecular weight; the temporary dipole moment in propane (C_3H_8) is therefore stronger than the temporary dipole moment in butane (C_4H_{10}) (Chang, 2003).

Permittivity is a complex quantity, given by:

$$\varepsilon_r = \varepsilon_r^{'} - j\varepsilon_r^{''}, \tag{5.17}$$

where ε_r=relative permittivity, ε_r'=real part of permittivity, ε_r''=imaginary part of petroleum. j is an indicator of a complex number (Application note, Agilent Technologies Inc.; Jin, 2010).

The real part of permittivity is a measure of how much energy is stored in a material when an external electric field is applied, while the imaginary part of permittivity is called the *loss factor* and is a measure of how dissipative or lossy a material is to an external applied field. In order to examine the complex dielectric properties of a

material, several isothermal scans as a function of frequency is performed, producing a complex permittivity spectrum as shown in figure 5.11.

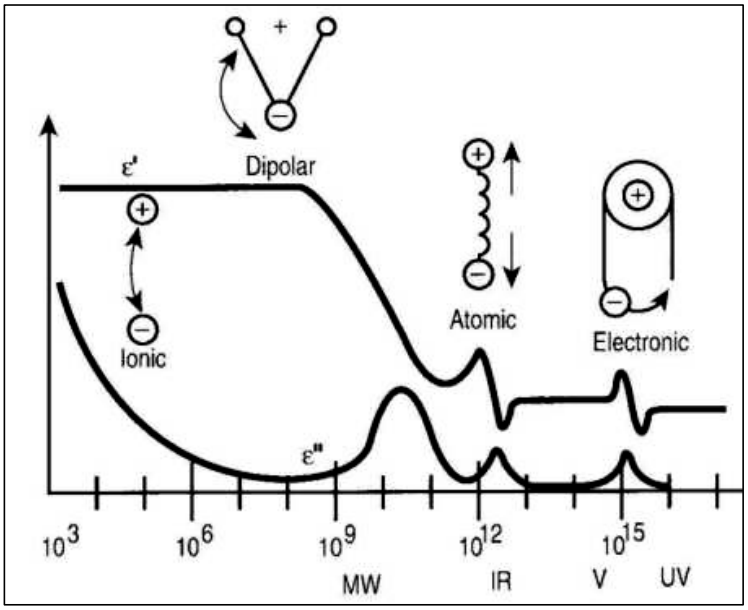


Figure 5.11: Complex permittivity spectrum, showing which polarization mechanisms contribute in the different frequency regions (Figure from http://en.wikipedia.org/wiki/Permittivity).

At low frequencies the response is caused by ionic polarization, which is composed of ionic conductivity and interfacial or space charge polarization. Space charge polarization occurs when more than one material component is present or when segregation occurs in a material containing incompatible chemical sequences and when translating charge carriers become trapped at the interfaces of these heterogeneous systems (Boiteux, 1987).

When increasing the frequency, the response caused by ionic polarization decreases and the response caused by permanent and temporary dipoles increases. In this region

the dipoles orient themselves with the applied electrical field, as shown in figure 5.12. At first, the dipoles manages to align themselves with the field, but as the frequency increases further the dipoles fail to align themselves, causing a drop in the real part of permittivity, and an increase in the imaginary part of permittivity (loss factor). This is called relaxation, and can be expressed by relaxation time, τ, which is a measure of the mobility of the molecules that exist in a material. The relaxation time is the time required for dipoles to become oriented in an electric field.

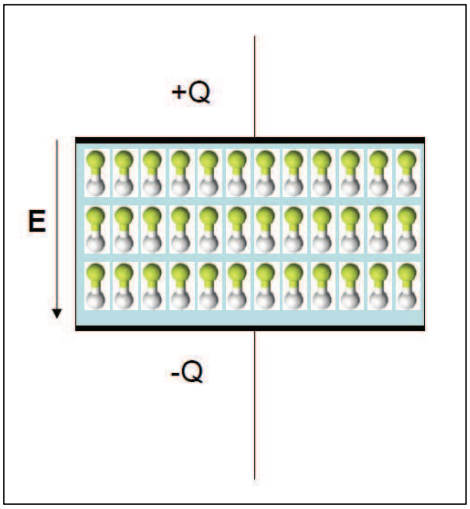


Figure 5.12: Alignment of dipoles in an electric field. White end= positive part of molecule, green end= negative part of the molecule.

Upon further increase of the frequency, the response is caused by atomic polarization. This occurs when neighbouring positive and negative ions "stretch" under an applied electric field.

Electronic polarization occurs in neutral atoms when an electric field displaces the nucleus with respect to the electrons that surround it. Both electronic and atomic

polarization creates induced dipole moments depending on the polarizability of the atoms or molecules (Boiteux, 1987).

5.6.1 The Cole- Cole model

The complex permittivity spectrum contains information regarding the physical properties and mechanisms in the material, and one method of extracting some of this information is by fitting the spectrum to a curve fit model like the Cole- Cole model (Cole et al., 1941), given in equation 5.18:

$$\varepsilon_r = \varepsilon_\infty + \frac{\varepsilon_s - \varepsilon_\infty}{1+(j\omega\tau)^{1-\alpha}} - j\frac{\sigma}{\omega\varepsilon_0} \quad , \qquad (5.18)$$

where ε_r=relative permittivity (dimensionless), ε_∞=high frequency permittivity (dimensionless), ε_s=static permittivity (dielectric constant at low frequencies, dimensionless), ω=angular frequency (Radians/seconds), τ=macroscopic relaxation time (Seconds), σ=finite conductivity (Siemens/meter), α=empirical factor (distribution factor, dimensionless), ε_0=permittivity in vacuum$\approx 8.85.\cdot 10^{-12}$ (Farads/meter).

The static permittivity for some molecules is given in table 5.3.

Table 5.3: Static permittivity (ε_s) of some organic molecules at 20°C (Marycott et al, 1951).

Molecule	**Static permittivity (20°C)**
C2H4O2 (acetic acid)	6.15
C3H6O2 (propionic acid)	3.30 (10°C), 3.44 (40°C)
***n*C5H12**	1.844
***n*C6H14**	1.89
Cyclohexane	2.024
Benzene	2.285
***n*C7H16**	1.924
3-ethylpentane	1.939
2,3-dimethylpentane	1.939
***n*C8H18**	1.948
C8H16O2 (octanoic acid)	2.45
***n*C9H20**	1.972
***n*C10H22**	1.991
***n*C11H24**	2.005
***n*C12H26**	2.014
C18H34O2 (linoleic acid)	2.71
H2O (water)	80.37

5.6.2 Measuring procedure

The dielectric spectra of the petroleum samples in this work were measured on a measurement system for complex permittivity measurements, based on a system developed by Christian Michelsen Research AS in 1996 (Folgerø, 1996), shown in figure 5.13.

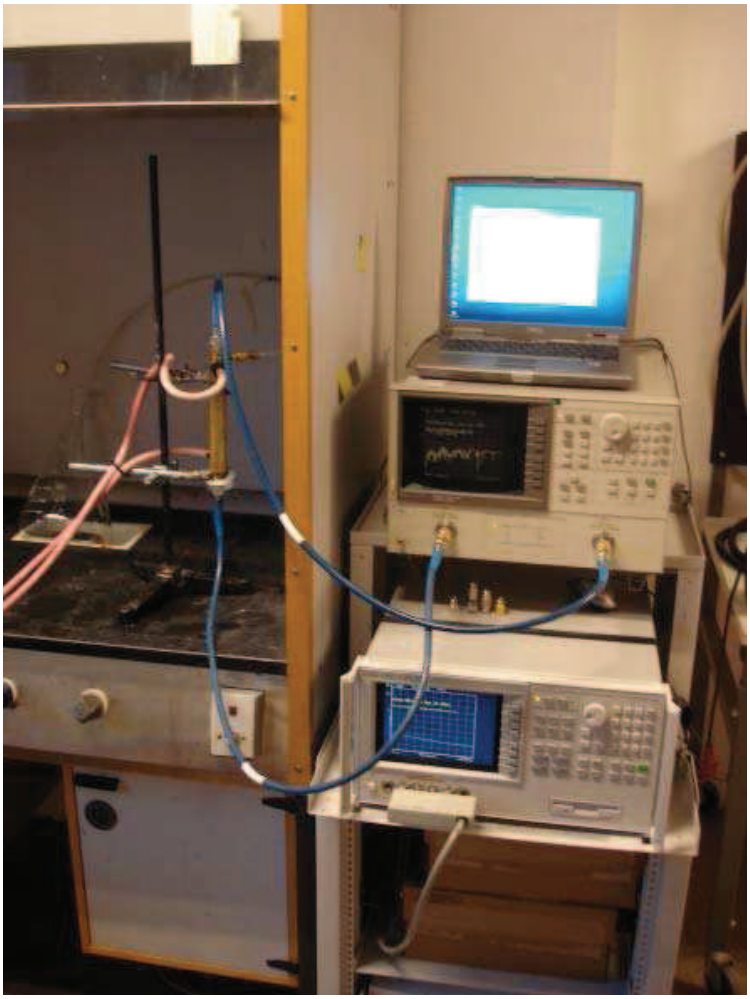


Figure 5.13: Measurement system for complex permittivity measurements.

For the frequency range from 1 kHertz to 20 MHertz, impedance analyzer measurements with a large reflection cell is performed, while the frequency range from 20 MHertz to 4 GHertz is covered by transmission coefficient measurements with a network analyzer.

5.7 Other investigated methods, variables, and approaches

In this chapter a brief summary of some of the methods, variables and approaches investigated during the thesis work is given. None of these methods, variables or approaches has been given any attention in the papers, as they did not give results that could support modelling of the target properties. However, several of them could be of use in other contexts or projects, given a different data set with more oils and with different properties to be investigated.

5.7.1 Magnetic Susceptibility

Magnetic Susceptibility is defined as a dimensionless proportionality constant that indicates the degree of magnetization of a material in response to an applied magnetic field:

$$\mathbf{M} = x_v \mathbf{H} \,, \tag{5.19}$$

where M=magnetization (magnetic dipole moment per unit volume) of the material (measured in amperes per meter), H=magnetic field strength, measured in amperes per meter, and X_v=volume magnetic susceptibility.

Magnetic susceptibility varies with the amount of the measured sample. The Mass Magnetic Susceptibility (MMS) is therefore calculated, including the mass in the

account. The measuring device assumes a sample volume of 10 mL, the corresponding MMS is calculated by the following formula:

$$MMS(m^3/kg) = (MS*10dm^3*10\text{^-}6(m^3/dm^3))/(m(g)*10\text{^-}3(kg/g)) \quad (5.20)$$

The magnetic susceptibility of petroleum fluids seems to be dependent on the heteroatom content of the crude oil (Ivakhnenko & Potter, 2004), and the magnetic susceptibility of petroleum fluid was therefore considered to be an interesting variable. The magnetic susceptibility of the sample set at hand was then measured; 12 crude oils and 4 model oils, in order to investigate the variation of the set. The model oils were Decane, Diesel, a Medical oil and a Transformer oil, and the purpose of including these samples in the dataset was to mimic different physical and electrical properties of crude oils.

The magnetic susceptibility measurements were done at the Geophysical department at the University of Bergen, since magnetic susceptibility is a measurement normally performed on rock and soil samples and rarely on liquids. Due to this, some experimenting was needed in order to determine the ideal filling grade of the sample glass and the ideal magnetic field strength. This was done by measuring the magnetic susceptibility of all of the crude oils in the sample set at room temperature at four different filling grades and with a magnetic field strength of 5, 50, 100, 200, 400 and 700 A/m. This was done in a total of 4 replicate measurement series, in order to eliminate any effects due to poor homogenization. This results in a total of 96 measurements for every oil. For each magnetic field strength and filling grade an average value is calculated, with corresponding standard deviation.

In Table 5.4 the typical average values of a crude oil, Crude 8 in this example, is shown.

Table 5.4: Average values for Crude 8, Mass Magnetic Susceptibility

Crude 8				
H(A/m)	**Volume 1**	**Volume 2**	**Volume 3**	**Volume 4**
5	-1,11E-08	-1,41E-08	-1,33E-08	-1,17E-08
50	-1,52E-08	-1,35E-08	-1,32E-08	-1,22E-08
100	-1,57E-08	-1,26E-08	-1,31E-08	-1,21E-08
200	-1,55E-08	-1,27E-08	-1,30E-08	-1,21E-08
400	-1,58E-08	-1,25E-08	-1,32E-08	-1,21E-08
700	-1,55E-08	-1,23E-08	-1,30E-08	-1,22E-08

No typical trend has been observed regarding field strength and MMS, it seems to vary a bit up and down between the strengths, and an average trend has been established. This was done by inspecting the trends for all volumes and field strengths, for all oils, and the average trend is exemplified in figure 5.14, with Crude oil 9, volume 1, which happened to have a similar trend as the average trend.

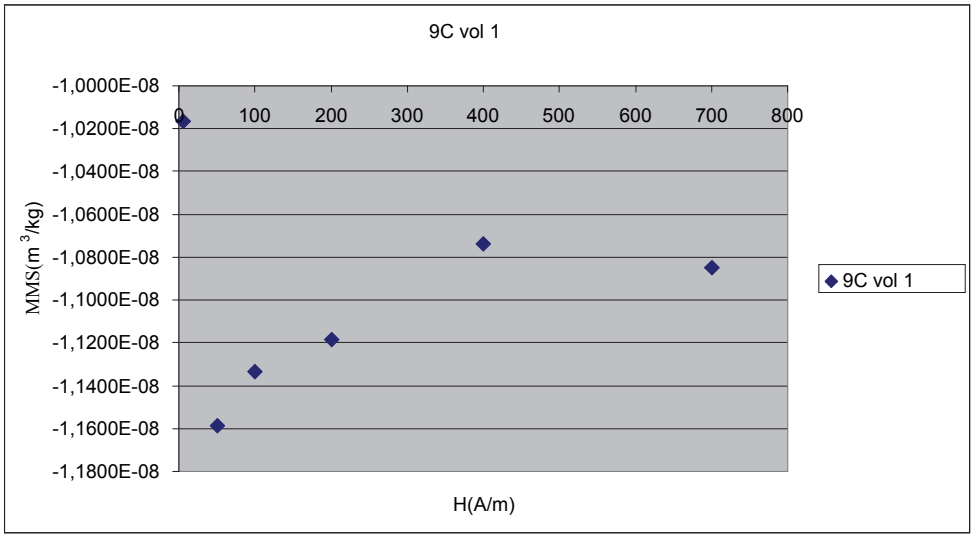


Figure 5.14: MMS vs Magnetic field strength (H) for the crude oil labelled 9C, vol 1.

For field strength 5A/m the measured value varies quite a lot, from being the highest to lowest measured value in each measuring series, and the situation in the figure does not give a 100% accurate description of the situation. Also, the measured values for 200 A/m is more similar to the measured values for 400A/m and 700A/m than what is the situation in the figure. The measured values for field strength 50 and 100 A/m are situated in the lowest region of the measured values. As such, neither 5, 50 nor 100 A/m seems to be good choices for measuring the magnetic susceptibility of crude oils, while 200, 400 and 700 A/m is quite similar in size and seems to be the field strengths of choice.

Regarding the standard deviation a trend is observed; the higher the field strength, the lower the standard deviation gets. In Table 5.5 the standard deviations for the different filling grades and magnetic field strengths for 8C are given. The standard deviations given here are typical results for the different filling grades and magnetic field strengths.

Table 5.5: Standard deviations for the crude oil labelled 8C.

8C				
H(A/m)	**Vol 1**	**Vol 2**	**Vol 3**	**Vol 4**
5	4,19E-09	3,17E-09	1,08E-09	7,70E-10
50	1,67E-09	1,22E-09	4,08E-10	2,36E-10
100	1,76E-09	1,26E-09	4,73E-10	1,27E-10
200	7,09E-10	7,27E-10	4,17E-10	1,85E-10
400	9,15E-10	5,04E-10	1,83E-10	4,70E-11
700	1,03E-09	6,59E-10	3,17E-10	4,04E-11

The standard deviations for Volume 1, the lowest filling grade, is typically lower than for the other filling grades, and is therefore considered to be a poor choice of measuring magnetic susceptibility. Furthermore, standard deviations for magnetic field strengths 5, 50 and 100 A/m is somewhat higher than for 200, 400 and 700 A/m, and is considered to be poor choices for measuring magnetic susceptibility.

In total, a filling grade of volume 2, 3 and 4, and with field strengths of 200, 400 and 700 A/m is considered to produce satisfactory measurements for the magnetic susceptibility of crude oil. However, the best standard deviations is achieved for volume 4, at magnetic fields strengths 400 and 700 A/m, but since a large number of measurements have been done, the average MMS have been calculated based on the measurements with filling grade 2,3 and 4 and with field strengths of 200, 400 and 700 A/m. A total of 36 measurements are then used in the calculation of every MMS value, resulting in figure 5.15.

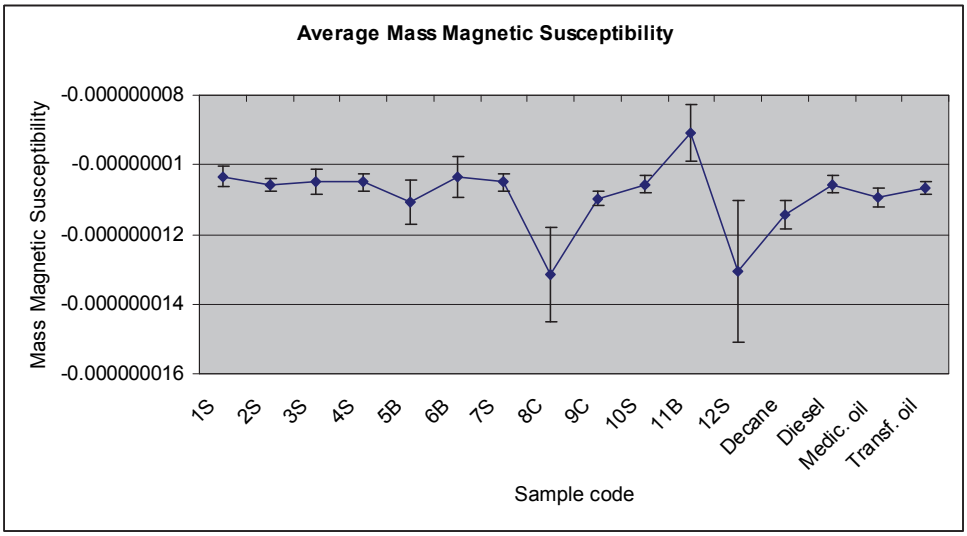


Figure 5.15: Average Mass Magnetic Susceptibility of the test set, with standard deviation ranges marked for every sample.

As shown in Figure 5.15, the variation in MMS is very low; only two or three oils have a significantly higher or lower value than the rest of the oils. No group trend is observed for biodegraded oils, condensates or nondegraded oils. The observed standard deviations for the oils with significantly higher or lower value are quite high as well; a spot sample of one of these oils could easily be measured to be within the range of the rest of the oils, indicating no variation at all.

Multivariate analysis and modelling of a sample set that has very low variance, and with just a few objects with significant variance, is very difficult. A quick test analysis confirmed that, as the R^2 value was only 0.55 with GC data as base. Accurate calibration models was not achieved, even distinguishing between high and low value was very difficult. Based on the low variation in MMS and high standard deviation in some of the samples, measuring of Magnetic Susceptibiliy of crude oils is considered not to give relevant information about the present data set.

Based on external information, we know that 11B has approximately 4 times as high sulphur content compared to 10S. In addition we know that 11B has the highest content of asphaltenes (which contains sulphur, oxygen and nitrogen) in the sample set. An observed trend is that the higher content of heterocompounds, sulphur in particular, the higher the MMS (Ivakhnenko & Potter, 2004). These two facts could explain the measured MMS for 11B. Given a data set with more oils that contain more heteroatoms like sulphur, the method of measuring MMS combined with analysis of the content of heteroatoms could be interesting to explore.

5.7.2 Headspace gas GC

With headspace gas it is referred to the gas volume above the crude oil. A sample of the crude oil is left in a gas tight container for 24 hours; it is then assumed that the liquid and gas in the container has reached a state of equilibrium. A gas sample is then extracted from the container and injected in a gas GC and the results are analysed. The analysed gases in the headspace gas sample of a crude oil are CO, CO_2, CH_4, C_2H_6, C_3H_8, iC_4H_{10}, nC_4H_{10}, iC_5H_{12}, and nC_5H_{12}.

During some of the early tests of this project it was suggested that the headspace gas could be an interesting variable to investigate, as the headspace gas composition should clearly be very dependent on the composition of the composition of the liquid it is in equilibrium with.

As such, gas GC analysis was performed on the data set at hand; 12 crude oils and 4 model oils, and the results were subjected to multivariate analysis and calibration modelling together with variables like permittivity, density and asphaltene content. But, as with the magnetic susceptibility, the variation of these samples was very low and the resulting calibration models had poor predictive quality; none was able to distinguish between high and low value.

As the gas GC gives results for 9 compounds, the information in the data is quite limited compared to the data given by whole oil GC (82+ compounds), and it is not likely that the information given by the gas GC would give more information than for the whole oil GC. At best it could say something about the volatility of the oil, and possibly distinguish between light crude oils (like condensates) and heavier crude oils (crude oils containing wax for example), and this information is already covered by the information given by whole oil GC.

The fact that the gas content is dependent on the sampling method and the use of oil/gas separation processes at the different oil fields, which we have no information of for any of the oil samples in the data set, further indicates that head space gas is a variable with high uncertainty.

As such, it was concluded that headspace gas was an uninteresting variable, at least for the purpose of this study.

5.7.3 Model oils

As mentioned in part 5.7.1, model oils (decane, diesel, Medical oil and transformer oil) were used in order to broaden the variation space for the different physical, magnetic and electric properties. This was done early in the project for several of the variables like magnetic susceptibility, gas GC, whole oil GC, density and permittivity. The reason for including these model oils was that they were used as calibration liquids for the permittivity measurement, in order to broaden the variation space. They were therefore used further for the other variables in order to broaden the variation space of those variables as well.

From the analytical chemistry perspective, this works perfectly well, but when performing multivariate analysis on these oils the problems starts. The main issue with these oils is that they have a chemical composition that is vastly different

compared to crude oils, so when creating a multivariate calibration model of the GC data for example, the model needs to stretch itself quite a lot in order to be able to model the information given from the model oils. Since the model oils are so different compared to the crude oils, they are weighted higher than the other oils and it seems like they are the most important objects in the data set, which they are not. The information given by the crude oils are deemed less important, and the model suffers from this as the ability of the model to predict variables of new unknown crude oils will be greatly reduced by introducing the model oils to the model. The model oils are classified as huge statistical outliers in all multivariate analysis and modelling performed, and should therefore not be included in the multivariate analysis or modelling. Diesel is the only model oil that is remotely similar to any of the crude oils in the data set (the condensates), but is still regarded as an outlier.

One other point that also eliminates the model oils from the data set is the fact that we want to investigate the variation of representative crude oil samples in the dataset. That is, crude oil samples that originates from real oil fields, with as little modifications, treatment or additions as possible. As model oils do not come close to falling in this category, model oils cannot be applied in the modelling stages.

It is for the same reasons that direct modification of crude oils in the sample set, like dilution with lighter hydrocarbons like pentane, hexane and so on, or mixing crude oils in order to make “new” crude oils is not an acceptable approach. We have no guaranty that the “new” oil we get is similar to the crude oil of any other oil field in the world, and as such we have no guaranty that the “new” oil is a representative crude oil. We considered modifying the crude oils in the sample set, but when the results from the initial multivariate analysis was ready, together with the fact that we would not have representative samples, the idea was rejected.

5.7.4 Biodegradation level

In the early stage of the project, we were interesting in knowing the level of biodegradation of the crude oils in the data set in order to group the oils into biodegraded oils, nondegraded oils and condensates. The biodegradation level was determined by inspecting the GC chromatograms of the crude oils (only the chromatograms) and comparing them to the Peters and Moldowan biodegradation scale (Figure 2.1). This resulted in a table of biodegradation levels, and as such made it possible to build calibration models for the biodegradation level based on GC or FTIR. The predictive quality of these models were quite good, as shown in table 5.6

Table 5.6: Predicted biodegradation levels for calibration model based on GC data. 1S and 11B are validation objects.

Object number	Predicted value	Measured value	Residuals	%Error
1S	0.25	0.10	-0.15	156.0
2S	0.09	0.10	0.00	3.4
3S	0.09	0.10	0.00	4.1
4S	0.08	0.10	0.01	13.1
5B	1.97	2.00	0.02	1.3
6B	3.01	3.00	-0.01	0.4
7S	0.16	0.10	-0.06	67.3
10S	0.11	0.10	-0.01	17.3
11B	1.69	2.00	0.30	15.2
12S	0.04	0.10	0.05	51.6

The results for the model based on IR is very similar, and shows that it is possible to predict the biodegradation level of crude oils based on GC or IR data. However, this requires that the biodegradation levels of the model objects are determined correctly,

and this is a big uncertainty with the method as the determination is done manually and based on the experience of the person who is performing the evaluation.

The usefulness of such models is quite limited as well, and considering the fact that the biodegradation info already is present within the GC and IR data (as will be shown in Paper I and II) we have no need for the models either, since we already have determined the levels manually. The models could be used by inexperienced users for a quick determination of the biodegradation level of a crude oil, given that the models are based on accurately determined biodegradation levels.

Based on these facts, it was determined that the biodegradation level was an uninteresting variable, at least for building calibration models.

Chapter 6

Multivariate data analysis/chemometrics

When handling large amounts of data, for example in spectroscopy where thousands of variables might be available, it is difficult to extract important information. Multivariate data analysis, or chemometrics as it is called when applied to chemical data, is a field where mathematical and statistical methods are applied in order to plan and optimize experiments and processes, recognise patterns and extract relevant information from data. Hence, application of multivariate data analysis to large datasets can help with the interpretation of the data (Nordtvedt et al., 1996; Martens and Næs, 1989).

6.1 Object space and variable space

Data can be represented as a matrix **X**, called the data matrix, which consists of N rows equivalent to a series of experiments or objects, and columns equivalent to M variables (temperature, wavelengths etc.), which in total constitutes a N x M matrix.

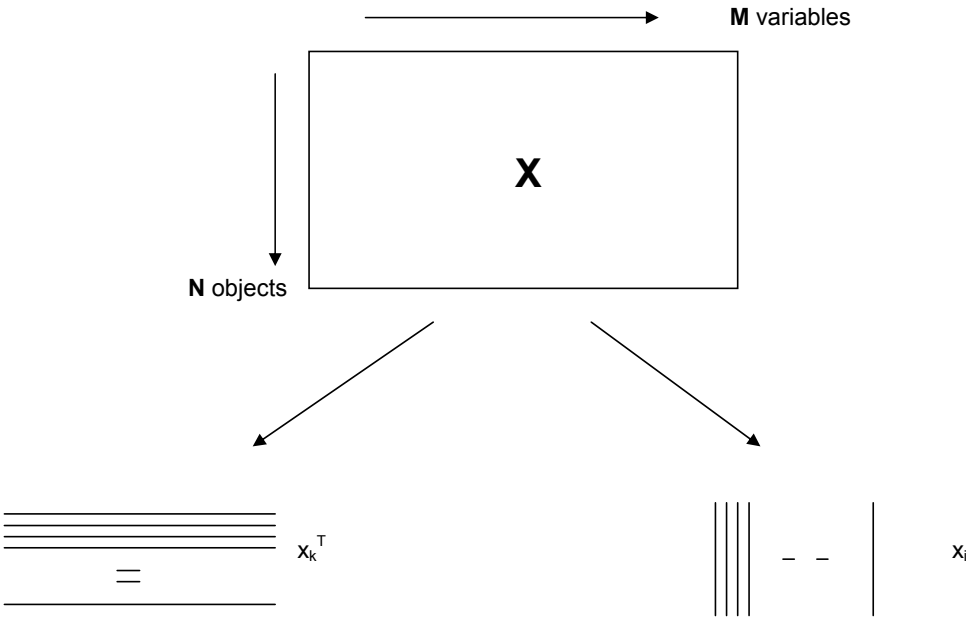


Figure 6.1: Data matrix X consists of N row vectors x_k^T and M column vectors x_i.

By plotting all of the object vectors in the M dimensional coordinate system where the axes are the M variables, you get a coordinate system that is called the *variable space.* Similarly you get the coordinate system called the *object space* by plotting the variable vectors in a N dimensional coordinate system where the axes are the N objects. Together, the variable and object space contains all the information in the dataset; the variable space contains all the information of the objects in the **X** matrix, for example correlations between different objects. Correlations between two objects *a* and *b* is given as:

$$\cos\varphi = \frac{x_a^T \times x_b}{\|x_a\| \times \|x_b\|} \tag{6.1}$$

If the angle φ is 0° the objects are perfectly positive correlated, if the angle is 180° the objects are perfectly negative correlated, and if the angle is 90° there is no correlation

between the objects. Similarities between objects can be found by looking at the angles and distances between them; the smaller the distance and angle between two objects, the more similar they are. The same principles can be used for the object space.

6.2 Pretreatment

Data collected directly from measuring instruments are termed *raw data*. Raw data can contain noise, baseline drift, scattering effects, dominant variables and other factors that may conceal the significant information in the dataset (Nordtvedt et al., 1996; Martens and Næs, 1989). Therefore, it might be necessary to pretreat the raw data, in order to remove any effects that do not represent chemical or physical properties in the dataset.

6.2.1 Centring

The objects positions in relation to each other is of higher interest than the objects positions in relation to the centre of the coordinate system, and in order to detect the variation between the objects, the raw data needs to be centred. This can be done by calculating the mean value for each variable and then subtracting the mean from the original values, as equation 6.2 shows:

$$\boldsymbol{X}_{\text{centred}}(i, j) = \boldsymbol{X}(i, j) \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{x}(i, j) , \tag{6.2}$$

where **X**=full matrix, i=column in matrix, j=row in matrix, N=total number of objects.
The centre of the coordinate system is then set to the centre of the dataset (Nordtvedt et al., 1996; Martens and Næs, 1989).

6.2.2 Block normalization

Some measuring procedures and instrumentation processes might include manual sampling, which might produce measurements with varying amounts of sample. In order to eliminate any effects from this, the variables can be normalized to constant sum. This is done by dividing the selected variables of each object with the sum of the objects to obtain the relative distribution of the variables in each object. This procedure is applied to GC data (Blomquist et al, 1979; Karrer et al, 1983).

6.2.3 Logarithmic transformation

Some variables might have a significant spread in variation causing the dataset to consist of, for example, a high amount of low value objects, and only a few medium and high value objects for a given variable. This can cause problems in the modelling, since there is such a large difference between the values. This effect can be minimized by a logarithmic transformation of the variable, causing the spread of variance to look smaller than it actually is (Aitchison, 1984). The transformation is done by doing the following calculation for the variables that are subject to transformation: New value=$\log_{10}$(1+old value). By adding a 1, negative values can also be transformed.

In this work, all datasets have been centred. Also, GC data has been block normalized in datasets where GC data is included. In addition, TAN and viscosity has been logarithmically transformed in datasets where they are included, as they contain a significant spread in variation.

6.3 Principal Component Analysis (PCA)

When handling large datasets with thousands of variables, for example from IR spectroscopy, the variable space will be very large and difficult to interpret. By gathering the relevant information in principal components (PCs), the amount of variables can be reduced to an amount that makes it easier to interpret (Nordtvedt et al., 1996; Martens and Næs, 1989; Wold et al., 1987).

6.3.1 Latent variables

The different variables in data sets might contain the same information – they are interdependent. This is called colinearity, and makes it possible to gather a smaller set of variables that explain all systematic variation in the dataset. These variables are linear combinations of the original variables, and are called Latent Variables. In PCA the LVs are called Principal Components (PCs).

6.3.2 PCA

PCA on a data matrix **X** can be described as:

$$\boldsymbol{X} = \sum_{a=1}^{A} \boldsymbol{t}_a \boldsymbol{p}_a + \boldsymbol{E} = \boldsymbol{T}\boldsymbol{P}^T + \boldsymbol{E}\,, \tag{6.3}$$

where A is the total amount of PCs, **X** is the start matrix with N x M dimension, **T** is a N x A matrix containing orthogonal score vectors t, **P** is a A x M matrix containing orthogonal loading vectors p. **E** is the residual matrix, has the same dimensions as **X**, and contains noise only.

The PCs can not be written neither as multiples nor as linear combinations of each other; they are linearly independent. The colinearity from the original data matrix is

removed by decomposing the dataset to PCs. In the multivariate space the PC now face in different directions and are perpendicular to each other.

The first PC is placed in the direction where the largest variation is situated. The next PC is placed orthogonal on the latter and in the direction that has the second highest variance. This continues until all the variance in the dataset is accounted for.

6.3.3 Scores and loadings

Every PC consists of a score vector and a loading vector. The loading vectors describes the directions of the PCs in the variable space, while the score vectors contains the different objects coordinates on the PCs. When performing a PCA, a new coordinate system is formed where the PCs makes up the axis. The core vector has dimension equal to the number of rows in the data matrix, while the length of the loading vector is equal to the number of columns in the data matrix.

The score of an object can be found by projecting the object down on a PC, the distance between the point to the origin of the coordinate system is then the score of the object. The loadings of different variables can be found using the same approach.

A score plot can be obtained by projecting the variable space down in the space spanned by the PCs. Similarities between different objects can be inspected by using the principles of distance and angles (chapter 6.1). Figure 6.2 shows an example:

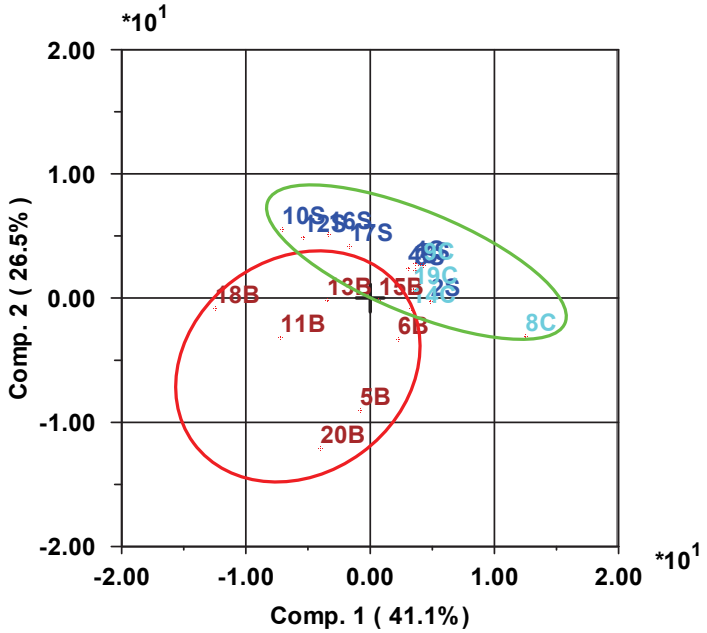


Figure 6.2: Score plot obtained from PCA analysis of GC data of the 20 oils in this project. Biodegraded oils=brown, nondegraded oils=blue, condensate=cyan.

Two oil samples being clearly separated in the score plot in figure 6.2 are 8C (a condensate) and 18B (a biodegraded oil), two oils that are chemically different as condensates generally contain lighter hydrocarbons, while in biodegraded oils the lighter hydrocarbons are removed by the degradation process. Groupings can be identified as well, as biodegraded oils are located in one part of the plot (red circle) while the nondegraded oils are located in the other part (green circle).

Similarly, loading plots can be obtained by projecting the object space down in the space spanned by the PCs. In a loading plot similarities between different variables can be inspected.

By plotting the scores and loadings in the same plot you obtain a biplot. This plot shows which objects are explained by certain variables, again by using the principles of distance and angle.

6.4 Multivariate calibration

By expressing one or more response variables *y* as a function of a given amount of *x* variables it is possible to predict the response values. What is actually done is to adapt the regression model (or calibration model) given as:

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{f} \ , \tag{6.4}$$

where **y** is the response variable subject to prediction, **X** contains the measured variables, **b** contains estimated regression coefficients and **f** contains residuals from **y**.

6.4.1 Partial Least Squares Regression (PLSR)

Partial Least Squares (PLS) (Hoskuldson, 1995) is a method for expressing a response variable **y** is a method to express a response variable **y** as a function of a given amount of **x** variables. The LVs obtained during a PCA explains as much as possible of the variance in **X**, while in PLS the LVs explains as much as possible of the covariance between **X** and **y**. Covariance is a measure of the linear independence between to varying sizes, in this case **X** and **y**.

One way to explain PLS is by imagining that you perform one PCA on **X** and one PCA on **y**, and then rotate two and two LVs (one from **X** and one from **y**) against each other until the correlation is at its highest; when the angle between the score vector in the **X** space and the score vector in the **y** space is 0. The context between the scores is then explained by:

$$\boldsymbol{u}_a = \boldsymbol{b}_a \boldsymbol{t}_a \,, \tag{6.5}$$

where **u** is the score vector in the **y** space, **t** is the score vector in the **X** space and **b** is the inner relation regression coefficient.

The decomposition that the PLS performs is done in one step, simultaneously on **X** and **y**. A vector **w**, called loading weights, is sought for each pair of LVs. These vectors are compromises between the LV in the **y** space and the LV in the **X** space. The PLS finds one vector at a time, the most significant first, and then subtracts the information explained from that vector from the **X** and **y** matrixes. Then the next vector is found.

In the end one ends up with a model where the regression vector is given by:

$$\boldsymbol{B} = \boldsymbol{W}\,(\boldsymbol{P}^T\boldsymbol{W})^{-}1\;\boldsymbol{Q}^T \,, \tag{6.6}$$

where **W** is the loading weights, **P** is loadings in the **X** space and **Q** is the loadings in the **y** space.

6.4.2 External validation

One way of inspecting the predictive quality of the PLS model is by using *external validation*. This is done by dividing the original data matrix into a training set and a validation set. The PLS model is build based on the training set, the validation set is then fitted to the model. By doing this the model is validated externally, as the validation set acts as unknown samples, and the prediction errors from the fitting is then a measure of the predictive quality of the model.

It is important that both training set and validation set originates from representative samples; that they are in the same concentration range and generally have the same conditions at sampling. A PLS calibration model is only valid in the range spanned by

the objects in the model; given a validation object with higher value than the object in the training set with the highest value, the prediction will most likely be uncertain.

Normally, the predictive quality of a PLS model increases with increasing number of samples in the training set. Similarly, increasing the amount of samples in the validation set will increase the total impression of the predictive quality of the PLS model. The number of samples in the training set and the validation set must then be balanced in such a way that the validation set does not decrease the predictive quality of the model, but still gives a reasonably good impression of the predictive quality of the model.

In this work external validation is used in order to inspect the predictive quality of the established models. The validation set, consisting of three objects, is chosen manually for each model; One object with low value for the modelled variable, one with medium value and one with high value. Typically, the object with the second or third lowest and highest value is chosen as the low and high value object, while one of the objects with about the mean value of the modelled variable is chosen as the medium value object. As the aim of the modelling is to achieve the best predictive quality, different validation sets are tested by exchanging one or more of the objects in the validation set, and after some trial and error the model with the best predictive quality is obtained. This is summarized on a qualitative scale ranging from “Poor”, via “OK” and “Good” to “Very good”. This is done by comparing the experimental uncertainty of the measurements with the prediction error of the validation objects; if the error is much lower than the experimental uncertainty, the predictive quality of the model is rated as “Very good”, if the error is similar to the experimental uncertainty, then the predictive quality of the model is rated as “Good”. If the difference between the experimental uncertainty and the prediction error is larger, the prediction errors are more closely examined. If the validation objects are predicted to be of high value, while the measured values are low, the model is rated as “Poor”. If the validation objects are predicted to be of high value, and the measured values are of high value, the model is rated as either “OK for distinguishing between high and low value”, or

"OK", depending on how close to the experimental uncertainty the prediction errors are.

Chapter 7

Results from papers

In this section a brief summary is given of the different papers and results obtained.

7.1 Research report: Introductory chemometric analysis of crude oil composition and fluid properties.

This report is based on a poster presented at the 25th International Meeting on Organic Geochemistry, held in Interlaken in September 2011. In this report a brief summary of the results from the initial chemometric analysis of the data set is presented. This report was an initial test to investigate whether the methods would work. The report is meant to be an introduction to the following papers.

Modelling results for the model for static permittivity based on GC, TAN and asphaltene content is shown. The prediction errors for the model lie in the same range as the experimental error for the permittivity measurements, indicating that the predictive quality of the model is good. The regression coefficients for the model are shown as well, with TAN as the most significant variable. In addition, the low to medium molecular weight straight chained alkanes have negative effect on the PLS model, and the high molecular weight straight chained alkanes have a positive effect on the PLS model. As the low to medium molecular weight straight chained alkanes are the first compounds to be altered during biodegradation, and since acids are

generated during biodegradation, a clear biodegradation effect is identified in the regression coefficient.

Also, modelling results for the model for viscosity based on GC, TAN and asphaltene content is shown. Reasonably good predictive quality is obtained for the high viscosity validation object, but the results are not so good for the low and medium value validation objects; precise predictions for low value objects is not possible. The model is good enough to distinguish between high and low viscosity objects, and give reasonably good predictions for high viscosity objects.

7.2 Paper I: Multivariate analysis of crude oil composition and fluid properties used in Multiphase Flow Metering. *ENERGY & FUELS, Volume: 26 Issue: 9 Pages: 5679-5688 Published: SEP 2012.*

This paper focuses on PLS modelling of parameters that are involved in multiphase flow metering systems, like density, permittivity and velocity of sound.

Score plots from PCA analysis of GC and FT-IR data, respectively, shows that three groupings occur; one grouping for biodegraded oils, one for nonbiodegraded oils and one for condensates. As these three types of petroleum oils are chemically different, this grouping pattern is expected.

PLS calibration models for static permittivity, high frequency permittivity, density and velocity of sound, based on GC and IR data respectively, is obtained with good predictive quality (compared to the experimental deviations for the measurements). The results are comparable, and somewhat better, to similar modelling work (Satya et al., 2007). For a quality assurance purpose, in order to detect changes in chemical composition in comparison with the last calibration of the MFM, these models are considered precise enough to be useful.

On the other side, the calibration models for permittivity variables tau and sigma, based on GC and IR data respectively, is obtained with poor predictive quality. This

indicates that estimating complex permittivity spectra by predicting the Cole- Cole parameters from IR or GC data will not be possible without significant error.

Upon inspection of the regression coefficients of the PLS calibration models, biodegradation seems to have the most significant effect on most of the models. For the models for static permittivity (shown in Paper 1), high frequency permittivity, density and velocity of sound based on GC, there is a positive effect from branched alkanes and higher molecular weight straight chained alkanes, while negative effect originates almost exclusively from low to medium molecular weight straight chained alkanes. This indicates that biodegradation of crude oil has an important effect on the variance of density in crude oils, since the smallest straight chained alkanes are the first to be removed or altered during biodegradation.

The same trend is observed for the models for static permittivity, high frequency permittivity, density and velocity of sound based on IR data; Biodegradation seems to have the most significant effect on the models, as the signals originating from CH3 stretch have positive effect on the model while the signals originating from CH2 stretch have negative effect on the model.

Although the models based on both IR and GC, respectively, show good predictive quality, IR spectroscopy is the best approach if MFM calibration is the major purpose. It is a much faster and easier technique, also; portable measuring devices for IR spectroscopy already exist. GC measurements take several hours, but are already in use for quality control in other contexts, and may therefore be a good choice in a combined flow assurance perspective.

7.3 Paper II: Comparison of Partial Least Squares calibration models of viscosity, acid number and asphaltene content in petroleum, based on GC and IR data. *FUEL, Volume: 120 Pages: 8-21 Published: MAR 2014.*

In this paper PLS models of some important crude oil parameters like viscosity, acid number and asphaltene content, based on GC and IR data respectively, have been compared based on predictive quality and chemical significance.

The PLS calibration models obtained in this paper have varying predictive quality, and the models based on IR data are generally better than the ones based on GC data, even when the GC models is supplied with additional information in form of TAN and/or asphaltene content. Although the models based on IR data are better than the ones based on GC data, they can not provide accurate predictions of viscosity, TAN and asphaltene content of an unknown sample of crude oil. However, if an OK estimate is required, the models should be able to give a satisfactory value. IR is an easy, rapid and generally available technique, while TAN, viscosity and asphaltene content measurements are more time consuming (30 minutes for viscosity and TAN, 2-3 days for asphaltene content) and require more preparations and equipment. Hence, an OK estimate produced by IR measurements would save both time and effort.

Regarding the chemical significance of the models, the regression coefficients for the models for TAN based on GC and IR data, respectively, show that biodegradation has a significant effect on the TAN value. As acids are generated during biodegradation, this is expected. For the models for viscosity based on GC and IR data, the regression coefficients show no clear indication of biodegradation. Biodegraded oils tend to have higher viscosities than nondegraded oils, and this is also observed in the dataset in this work. However, this is not reflected in the regression coefficients. The same is partly observed for the models for asphaltene content; the biodegraded oils in this dataset tend to have higher contents of asphaltenes than nondegraded oils, but this is only partly reflected in the regression coefficients for the model based on GC. The

model based on IR does not have a clear indication that biodegradation has a significant effect on the variation.

The fact that IR spectra give overall somewhat better models than GC can be understood as the spectroscopic "snapshot" of the oil composition contains more information – or more relevant information – than the individual hydrocarbon distribution in the GC data range, which can be considered to represent the "solvent" part of the crude oil in relation to the lower concentration, more polar compounds like asphaltenes and petroleum acids.

7.4 Paper III: Estimation of dielectric properties of crude oil based on IR spectroscopy. Submitted to *Chemom. Intell. Lab. Sys., second revision.*

In this paper the feasibility of estimating permittivity spectra from measured IR spectra using multivariate calibration techniques has been investigated. In addition, a physical model of permittivity is fitted to both the estimated spectra and the original measured spectra, and the resulting model parameters are compared in order to assess the usefulness of the proposed method.

The estimated permittivity spectra in this paper are obtained well within the experimental uncertainty of the dielectric spectroscopy used to obtain the original permittivity spectra, and hence show that it is possible to obtain information about dielectric properties based on IR data.

The fitting of the Cole- Cole model to the spectra are not that conclusive, as there are large deviations for some crude oils. The parameter estimation is sensitive to experimental noise, and there are some issues in the algorithm for the estimation that needs to be sorted in order to be robust.

IR spectrophotometers are already widely used, and the method proposed in this paper would add functionality to these, as additional information in form of permittivity spectra can be obtained by applying the method to IR spectra.

7.5 Paper IV: Permittivity calculator. Method and tool for calculating the permittivity of oils from PVT data. ***Conference proceedings, 30th International North Sea Flow Measurement Workshop, St. Andrews, 23rd – 26th October 2012 (http://www.nfogm.no/ikbViewer/Content/873131/NSFMW%202012%20-%20Technical%20Papers.pdf)*****.**

In this paper the method and tool for calculating the permittivity of crude oils from Pressure Volume Temperature (PVT) data is presented. The work presented here was initiated as it was realised that some of the results from paper II and III could be transferred to an area where it could be used directly. GC data and PVT data are in many ways similar, and as the calibration models from paper II and III showed good predictive quality for both density and static permittivity based on GC data investigations was started in order to see if the models could be modified to work on PVT data. The presented method in this paper show that the permittivity of a crude oil can be predicted within 4% at standard conditions, and considering the fact that the measuring uncertainty of permittivity measurements are estimated to 2% we can conclude that the permittivity can be predicted with good accuracy.

The method presented in this paper has been implemented as a software tool which applies an equation and a regression model given in the paper to calculate the static permittivity of a given crude oil based on PVT data from the given crude oil. The tool has been implemented by Roxar ASA for initial calibration of multiphase flow meters, and it could also be applied to accurate PVT data sampled during the production lifetime of an oil field in order to detect changes in permittivity and density, and to recalibrate the meter.

Chapter 8

Concluding remarks and suggestions for further work

The work presented in this thesis confirms established results like the fact that biodegradation has a significant effect on several crude oil properties. Also, this work has presented new results regarding the permittivity of crude oil as a function of composition. This subject has not been investigated much earlier, and hence the results presented here have contributed to new knowledge on that topic.

Estimation of permittivity spectra from IR data can be used as a routine quality control of calibration values for MFM, however; the algorithm must be improved in order to increase the robustness.

The project sought to improve the performance of the MFM systems. As a result of this thesis, a software tool has been developed for initial and recalibration purposes of MFMs. As such, some of the results obtained in this thesis have provided important information and knowledge to the MFM industry, as well as a useful tool for direct use.

In order to further investigate the effect of polarity on petroleum properties, measurements like Total Base Number, elemental composition and SARA (Saturates, Aromatics, Resins and Asphaltenes) should be done on the crude oils, and PLS

calibration models of these results based on IR data and GC data, respectively, should be built. In addition, these new variables should be included in the models for static and high frequency permittivity based on GC in order to investigate whether they have further information that can contribute to improved prediction quality of the calibration models.

Bibliography

Abbasa, O.; Rebufaa, C.; Dupuya, N.; Permanyerb, A.; Kistera, J. 2008. Assessing petroleum oils biodegradation by chemometric analysis of spectroscopic data, *Talanta*, 4, 75, 857-871.

Alagic, E. 2005. Ultralydsmålinger under høyt trykk – En undersøkelse av solubilisering av 1-heksanol i natrium dodecylsulfat som funksjon av trykk, *Master thesis in physical chemistry, University of Bergen*.

Albrecht, P.; Ourisson, G. 1971. Biogenic substances in sediments and fossils, *Angew. Chem. Int. Edn.* , 10, 209-225.

Application note; Basics of Measuring the Dielectric Properties of Materials, *Agilent Technologies Inc.* Available from World Wide Web: http://cp.literature.agilent.com/litweb/pdf/5989-2589EN.pdf, 07.04.14.

Application note. 2001. Introduction to Fourier Transform Infrared Spectroscopy, *Thermo Nicolet Corporation*. Available from World Wide Web: http://mmrc.caltech.edu/FTIR/FTIRintro.pdf, 07.04.14.

Attanasi, E.D.; Meyer, R.F. 2010. Survey of energy resources (22 ed.). *World Energy Council*, 51.

Baboian, R. 2005. Corrosion tests and standards: application and interpretation, *ASTM International*.

Bailey, N.J.L.; Krouse, H.H.; Evans, C.R.; Rogers, M.A. 1973. Alteration of crude oil by waters and bacteria – evidence from geochemical and isotope studies, *Am. Assoc. Pet. Geol. Bull.*, 57, 1276- 1290.

Bak, J.; Larsen, A. 1994. Quantitative analysis of the gases CO, CO2 and CH4 by Fourier Transform Infrared (FTIR) spectroscopy, *Dansk Kemi*, 75, 8, 16-21.

Becker, J.R. 1997. Crude oil. Waxes, Emulsions, and Asphaltenes, *PennWell Publishing Company*.

Behar, F.H., Albrecht, P. 1984. Correlation between carboxylic acids and hydrocarbons in several crude oils. Alteration by biodegradation, *Organic Geochemistry*, 6, 597–604.

Benenson, W.; Harris, J. W.; Stocker, H.; Lutz, H. 2002. Handbook of Physics, *Springer-Verlag New York Inc*.

Bjørndal, E. 2007. Acoustic measurement of liquid density with applications for mass measurement of oil, *University of Bergen*, PhD thesis.

Blomquist, G.; Johansson, E.; Söderström, B.; Wold, S. 1979. Reproducibility of pyrolysis-gas chromatographic analyses of the mould penicillium brevi-compactum, *Journal of Chromatography*, 173, 7-17.

Boiteux, G. 1987. Dielectric and Related Molecular Processes in Ion Containing Polymers, *Proceedings of the NATO Advanced Research Workshop on Structure and Properties of Ionomers*, 227-245.

Bracewell, R.N. 1978. The Fourier Transform and its Application, *McGraw-Hill Book Co.*, New York, 2nd edition.

Chang, R. 2003. General Chemistry: The essential concepts, *The McGraw- Hill Companies*, 3rd edition.

Christian Michelsen Research AS, www.cmr.no, 07.04.14.

Chromeleon™, Thermo Scientific Dionex. Version 6.6 available from March 2004, version 7.1 available from February 2011. Available from World Wide Web: http://www.dionex.com/en-us/products/chromatography-software/chromeleon7/lp-73027.html, 07.04.14.

Cole, K.S.; Cole, R.H. 1941. Dispersion and absorption in dielectrics I. Alternating current characteristics, *J. Chem. Phys.* , 9, 341-351.

Connan, J.; Le Tran, K.; van der Weide, B. 1975. Alteration of Petroleum in Reservoirs. Proc. 9th World Pet. Congr. Tokyo, *Applied Science Publ.*, 2, 171-178.

Cullum, D.C. 1994. Introduction to surface analysis, *Springer*.

Cutnell, J.D.; Johnson, K.W. 2004. Physics, *Wiley and Sons*, 6th edition.

Del Grosso, V.A.; Mader, C.W. 1972. Speed of sound in pure water, *J. Acoust. Soc. Am.*, 52, 1442–1446.

Deroo, G.; Tissot, B.; McCrossan, R.G.; Der, F. 1974. Geochemistry of the heavy oils of Alberta, *Oil Sands Fuels of the Future. Memoir 3, Can. Soc. Pet. Geol*, 148-167, 184-189.

Evans, C.R.; Rogers, M.A.; Bailey, N.J.L. 1971. Evolution and alteration of petroleum in Western Canada. *Chem. Geol.*, 8, 147-170.

Falcone, G.; Harrison, B. 2011. Forecast expects continued multiphase flowmeter growth. *Oil & Gas Journal*, Mar 7, 109.

Falcone, G.; Hewitt, G.F.; Alimonti, C. 2009. Multiphase Flow Metering, *Elsevier B.V.*, 1st edition.

Falcone, G.; Hewitt, G.F.; Alimonti, C.; Harrison, B. 2002. Multiphase flow metering: Current trends and future developments, *J. Pet. Technol.* 54, 77-84.

Fink, J.K. 2012. Petroleum Engineer's Guide to Oil Field Chemicals and Fluids, *Elsevier Inc.*, 2012.

FitzPatrick, M.; Champagne, P.; Cunningham, M. F. 2012. Quantitative determination of cellulose dissolved in 1-ethyl-3-methylimidazolium acetate using partial least squares regression on FTIR spectra, *CARBOHYDRATE POLYMERS*, 87, 2, 1124-1130.

Folgerø, K. 1996. Coaxial Sensors For Broad Band Complex Permittivity Measurements of Petroleum Fluids. *University of Bergen*, PhD thesis.

Gao, S. 2008. Investigation of Interactions between Gas Hydrates and Several Other Flow Assurance Elements, *Energy & Fuels*, 22 (5), 3150–3153.

Grob, R.L.; Barry, E.F. 2004. Modern Practice of Gas Chromatography, *Wiley & Sons*, 4th edition.

Groenzin, H.; Mullins, O.C. 2000. Molecular size and structure of asphaltenes from various sources, *Energy Fuels*, 14, 677-684.

Halliday, D.; Resnick, R.; Walker, J. 1997. Fundamentals of Physics, *Wiley and Sons*, 5th edition Extended.

Harris, D.C. 2003. Quantitative Chemical Analysis, *W.H. Freeman and Company*, 6th edition.

Head, Ian M., Martin Jones, D., Larter, Steve R. 2003: Biological activity in the deep subsurface and the origin of heavy oil, *Nature*, 426, 344-352

Hunt, J. M. 1979. Petroleum Geochemistry and Geology, *Freeman,* San Francisco.

Hoskuldsson, A. 1995. A combined theory for PCA and PLS. J. *Chemometrics*, 9, 91.

Ivakhnenko, O. P.; Potter, D. K. 2004. Magnetic susceptibility of petroleum reservoir fluid. *Physics and Chemistry of the Earth*, 29, 899-907.

Jin, J-M. 2010. Theory and Computation of Electromagnetic Fields, *Wiley & Sons Inc.*

Karrer, L. L.; Gordon, H. L.; Rothstein, S. M.; Miller, J. M.; Jones, T. R. B. 1983. Bayesian statistical methods for use in mass spectral assignment, *Analytical chemistry*, 55, 1723-1728.

Karstang, T.V.; Christy, A.A.; Dahl, B.; Kvalheim, O.M. 1991. Diffuse reflectance Fourier-Transformed infrared- spectroscopy in petroleum exploration: A multivariate approach to maturity determination. *Journal of Geochemical Exploration*, 41, 213-226.

Kontogeorgis, M.G.; Folas, G.K. 2010. Thermodynamic models for industrial applications. From classical and advanced mixing rules to association theories, *John Wiley & Sons Ltd*.

Laughton, M.A.; Warne, D.F. 2003. Electrical engineer's reference book, *Elsevier Science*, 16th edition.

Lochte, H.L.; Litmann, E.R. 1955. The petroleum Acids and Bases, *Chemical Publishing Company*, New York.

Martens, H.; Næs, T. 1989. Multivariate Calibration, *John Wiley & Sons*.

Marycott, A.A.; Smiths, E. R. 1951**.** Table of Dielectric Constants of Pure Liquids**,** *United States Department of Commerce*, National Bureau of Standards Circular 514.

Meredith, W.; Kelland, S.-J.; Jones, D.M. 2000. Influence of biodegradation on crude oil acidity and carboxylic acid composition. *Organic Geochemistry*, 31, 1059–1073.

Metzger, T.G. 2006. The Rheology Handbook: For users of rotational and oscillatory rheometers, *Vincentz Network*, 2nd revised edition.

Moen, L.K. 1996. Effekter av reaksjonsmiljø på dannelse og nedbrytning av organiske syrer i olje, *Master thesis in organic chemistry, University of Bergen*.

Mullins, O.C.; Sheu, E.Y.; Hammami, A.; Marshall, A.G. 2007. Asphaltenes, Heavy oils and Petroleomics, *Springer Science+Business Media*.

Nordtvedt, R.; Brakstad, F.; Kvalheim, O. M.; Lunstedt, T. 1996. Anvendelse av kjemometri innen forskning og industri, *Tidsforlaget Kjemi AS*.

Operating Instructions for Digital Precision Density Measurement System for Liquids and Gases, *Anton Paar*, www.antonpaar.com.

Ozerov, R.P.; Vorobyev, A.A. 2007. Physics for Chemists, *Elsevier B.V.*

Parisotto, G.; Ferrao, M.; Muller, A.L.H. ; Muller, E-I.; Santos, M.F.P.; Guimaraes, R.C.L.; Dias, J.C.M.; Flores, E.M.M. 2010. Total Acid Number Determination in Residues of Crude Oil Distillation Using ATR-FTIR and Variable Selection by Chemometric Methods, *Energy Fuels*, 24, 5474-5478.

Pavia, D.L.; Lampmann, G.M.; Kriz, G.S. 2001. Introduction to Spectroscopy, *Thomsom Learning*, Bellingham, Washington, 3rd edition, 13-83.

Peinder, P., Visser, T., Petrauskas, D. D., Salvatori, F., Soulimani, F., Weckhuysen, B. M. 2009. Partial least squares modeling of combined infrared, ^{1}H NMR and ^{13}C NMR spectra to predict long residue properties of crude oils. *Vibrational Spectroscopy* 51, 205– 212.

Pepper, A.; Santiago, C. 2001. Impact of biodegradation on petroleum exploration and production: Observations and outstanding problems. *Abstracts of Earth Systems Processes*. Geol. Soc. London, Geol. Soc. of America, Edinburgh, 24-28 June.

Peters, K. E.; Fowler, M. G. 2002. Applications of petroleum geochemistry to exploration and reservoir management. *Org. Geochem.*, 33, 5-36.

Peters, K.E.; Walters, C.C.; Moldowan, J.M. 2005. The Biomarker Guide Volume 2, *Cambridge University Press*, 2nd edition.

Picker, P.; Tremblay, E.; Jolicoeur, C. 1974 A high-precision digital readout flow densimeter for liquids, *J. Sol. Chem*., 3, 377-384,

Planck, M. 1901. On the Law of Energy in the Normal Spectrum, *Annalen der Physik*, 4, 553.

Profeta, L. T. M.; Sams R. L.; Johnson, T. J. 2011. Quantitative Infrared Intensity Studies of Vapor-Phase Glyoxal, Methylglyoxal, and 2,3-Butanedione (Diacetyl) with Vibrational Assignments, *Journal of Physical Chemistry A*, 115, 35, 9886-9900.

Rossing, T.D. 2007. Springer handbook of acoustics, *Springer Science+Business Media*.

Satya, S.; Roehner, R.M.; Milind, D.D.; Hanson, F.V. 2007. Estimation of Properties of Crude Oil Residual Fractions Using Chemometrics, *Energy & Fuels,* 2, 21, 998-1005.

Seifert, W.K.; Howells, W.G. 1969. Interfacially active acids in a California crude oil. Isolation of carboxylic acids and phenols, *Anal. Chem.*, 41, 554-562.

Seifert, W.K.; Teeter, R.M. 1970. Identification of polycyclic aromatic and heterocyclic crude oil carboxylic acids, *Anal. Chem.*, 42, 750-758.

Sheng, J.J. 2011. Modern Chemical Enhanced Oil Recovery - Theory and Practice, *Elsevier Inc.*

Speight, J.A. 1998. Petroleum Chemistry and Refining, *Taylor &Francis*.

Speight, J.A. 1980. The chemistry and technology of petroleum, *Marcel Dekker Inc*.

Svanberg, S. 2004. Atomic and Molecular Spectroscopy: Basic Aspects and Practical Applications, *Springer-Verlag*, 4th edition.

Talita Lammoglia, T.; de Souza Filho, C.R. 2011. Spectroscopic characterization of oils yielded from Brazilian offshore basins: Potential applications of remote sensing. *Remote Sensing of Environment*, 10, 115, 2525-2535.

Thorn, K. A.; Aiken, G. R. 1998. Biodegradation of crude oil into non- volatile acids in a contaminated aquifer near Bemidji, Minnesota. *Org. Geochem.* 29, 909-931.

Thorn, R.; Johansen, G.A.; Hammer, E. A. 1999. Threephase flow measurement in the offshore oil industry: Is there a place for process tomography. In 1st World Conference on Industry Process Tomography, Buxton, Greater Manchester, U.K.

Tissot, B.P.; Welte, D.H. 1984. Petroleum Formation and Occurrence, *Springer-Verlag*, 2nd revised and enlarged Edition.

Viswanath, D.S.; Ghosh, T.K.; Prasad, D.H.L.; Dutt, N.V.K.; Rani, K.Y. 2007. Viscosity of liquids, *Springer*.

Weiss, H. M.; Wilhelms, A.; Mills, N.; Scotchmer, J.; Hall, P. B.; Lind, K.; Brekke, T. 2000. NIGOGA - The Norwegian Industry Guide to Organic Geochemical Analyses [online]. *Edition 4.0 Published by Norsk Hydro, Statoil, Geolab Nor, SINTEF Petroleum Research and the Norwegian Petroleum Directorate*. Available from World Wide Web: http://www.npd.no/engelsk/nigoga/default.htm , 07.04.14.

Wenger, L.M.; Davis, C.L.; Isaksen, G.H.. 2002. Multiple controls on petroleum biodegradation and impact on oil quality. *SPE Reservoir Evaluation and Engineering*, 5, 375–383.

Wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Permittivity, 07.04.14

Wold, S.; Esbensen, K.; Geladi, P. 1987. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2, 37.

Yen, T.F.; Chilingarian, G.V. 2000. Asphaltenes and Asphalts, Volume 2, *Elsevier Science B.V*.

Appendix

Appendix A. FT-IR spectra

In this section the FT-IR spectra of all the crude oils in this thesis is presented. In the figures the x-axis is the wave number region (1/cm) at which the FT-IR spectra are measured, and the y-axis is the absorbance for each wave number.

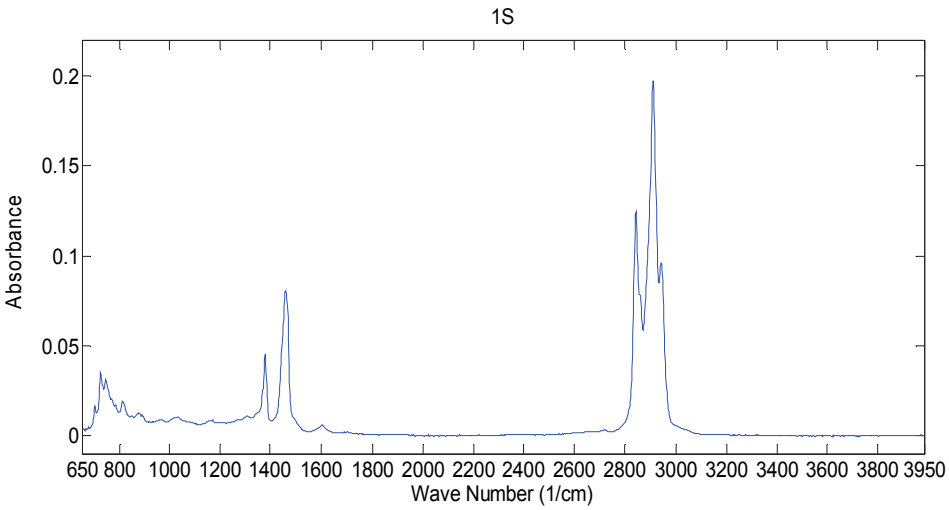


Figure A.1: FTIR spectrum of the crude oil labelled 1S in the crude oil dataset.

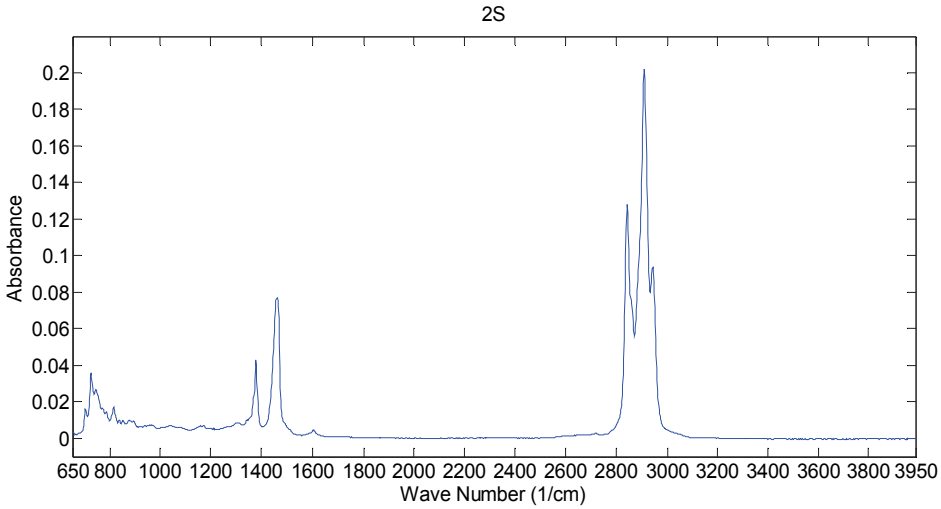


Figure A.2: FTIR spectrum of the crude oil labelled 2S in the crude oil dataset.

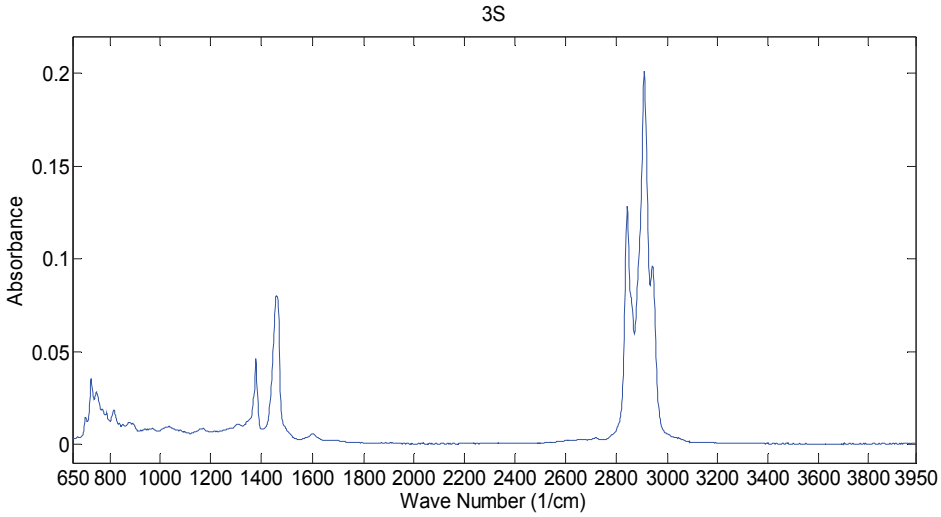


Figure A.3: FTIR spectrum of the crude oil labelled 3S in the crude oil dataset.

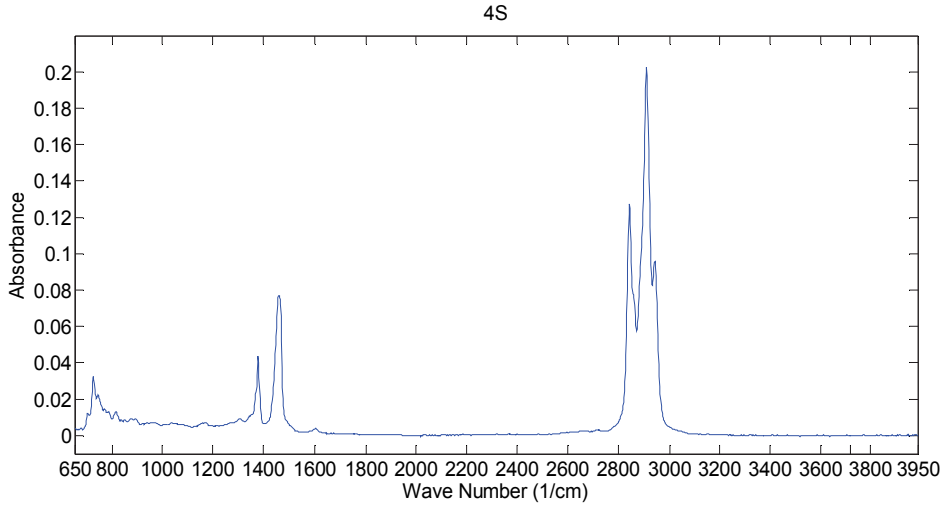


Figure A.4: FTIR spectrum of the crude oil labelled 4S in the crude oil dataset.

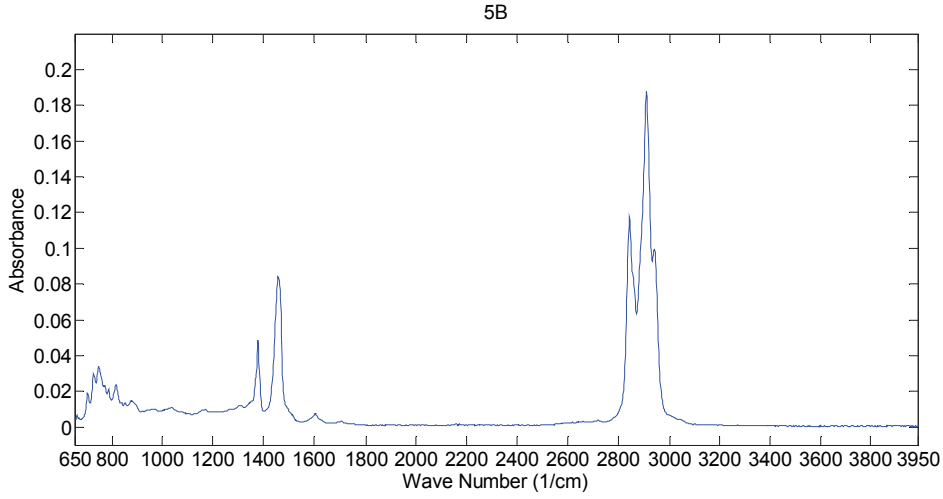


Figure A.5: FTIR spectrum of the crude oil labelled 5B in the crude oil dataset.

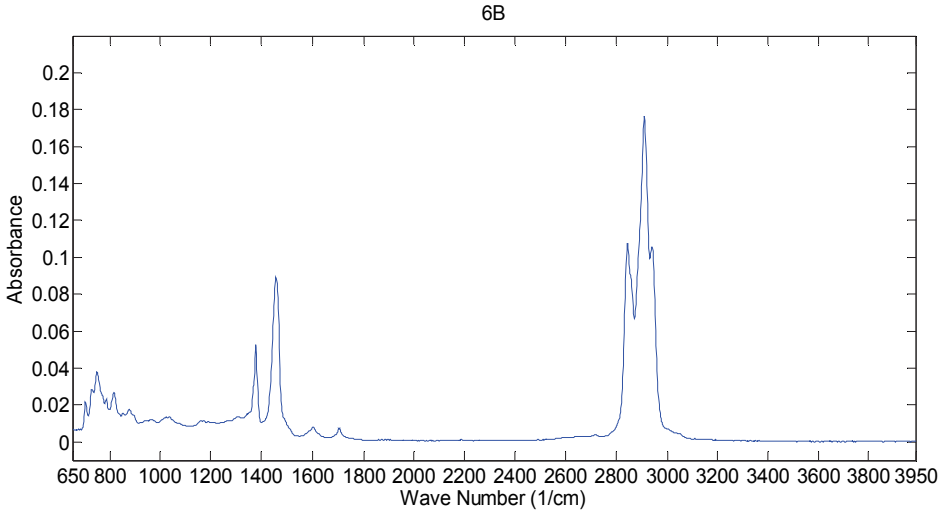


Figure A.6: FTIR spectrum of the crude oil labelled 6B in the crude oil dataset.

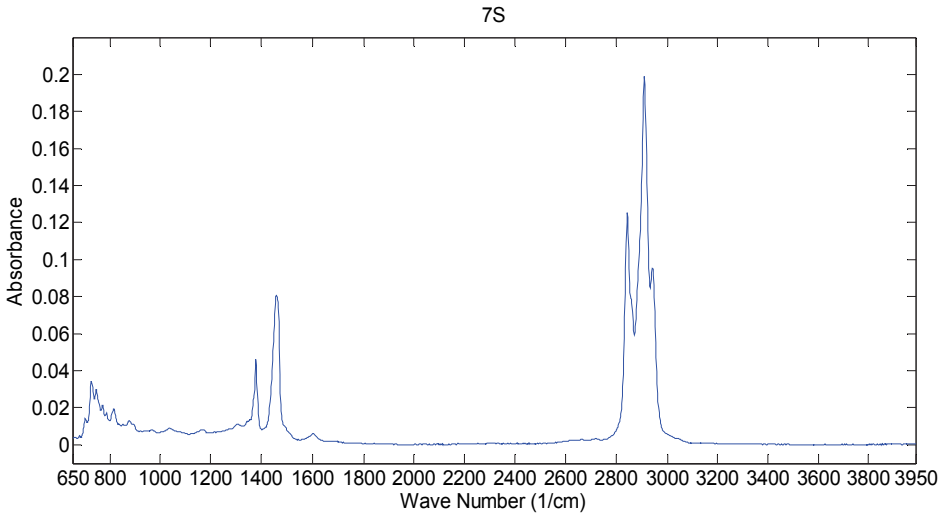


Figure A.7: FTIR spectrum of the crude oil labelled 7S in the crude oil dataset.

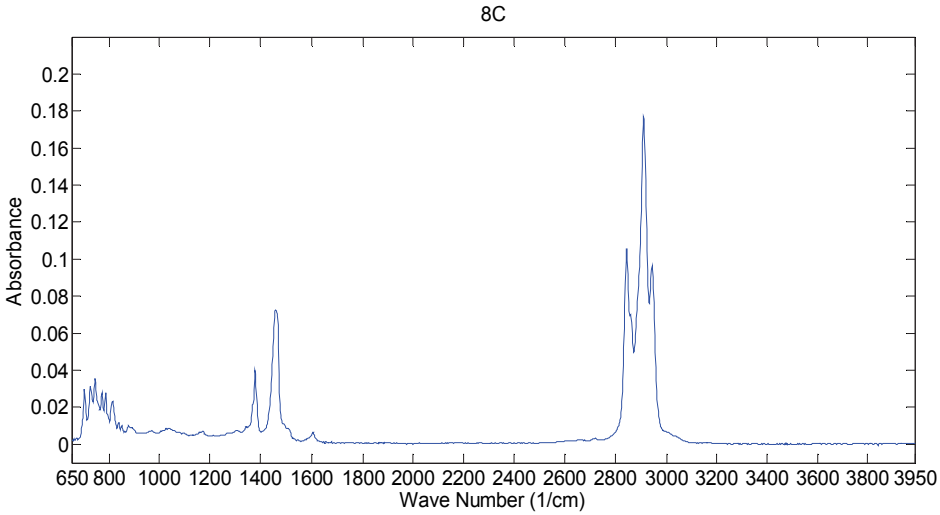


Figure A.8: FTIR spectrum of the crude oil labelled 8C in the crude oil dataset.

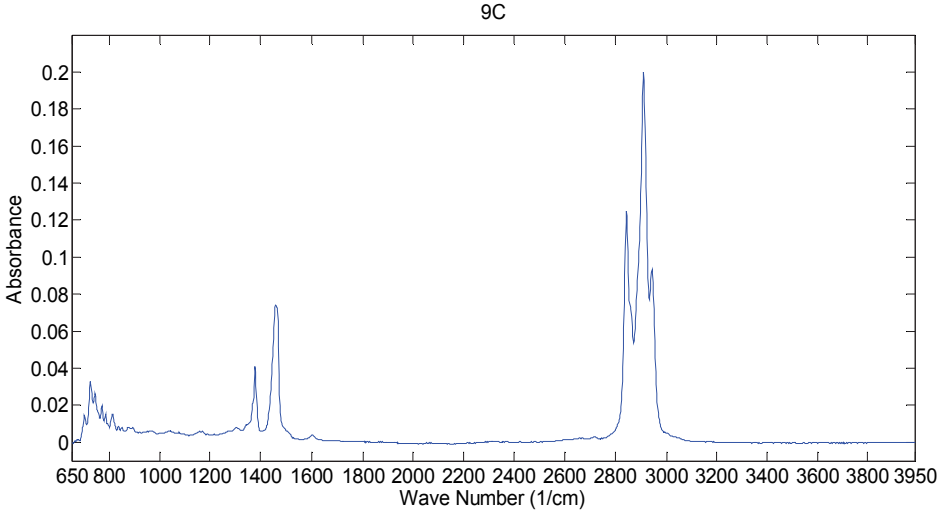


Figure A.9: FTIR spectrum of the crude oil labelled 9C in the crude oil dataset.

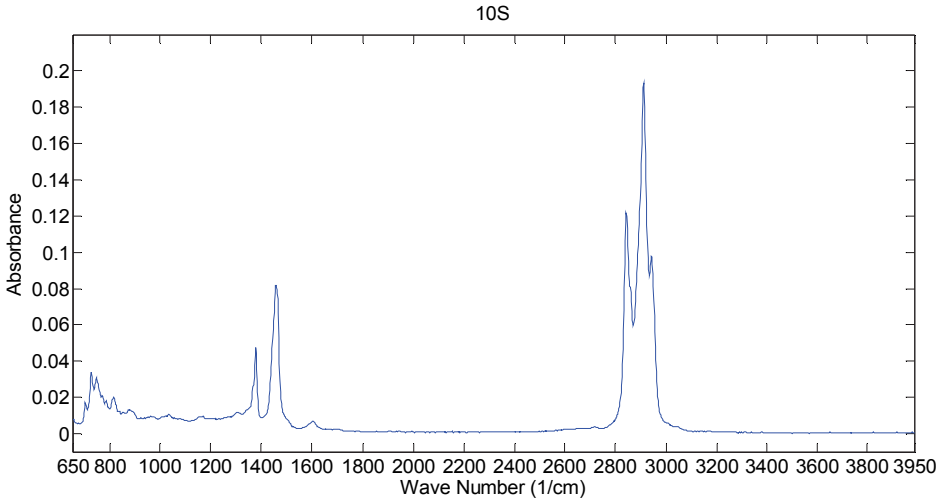


Figure A.10: FTIR spectrum of the crude oil labelled 10S in the crude oil dataset.

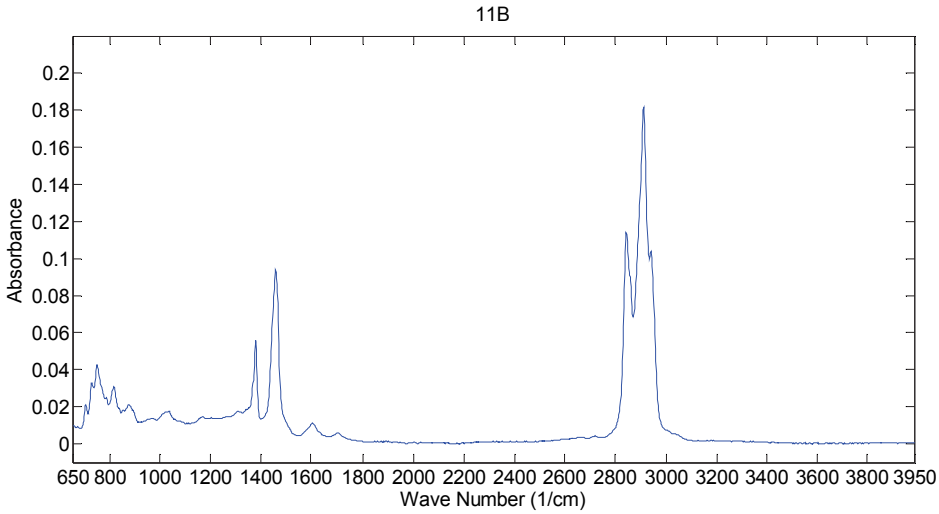


Figure A.11: FTIR spectrum of the crude oil labelled 11B in the crude oil dataset.

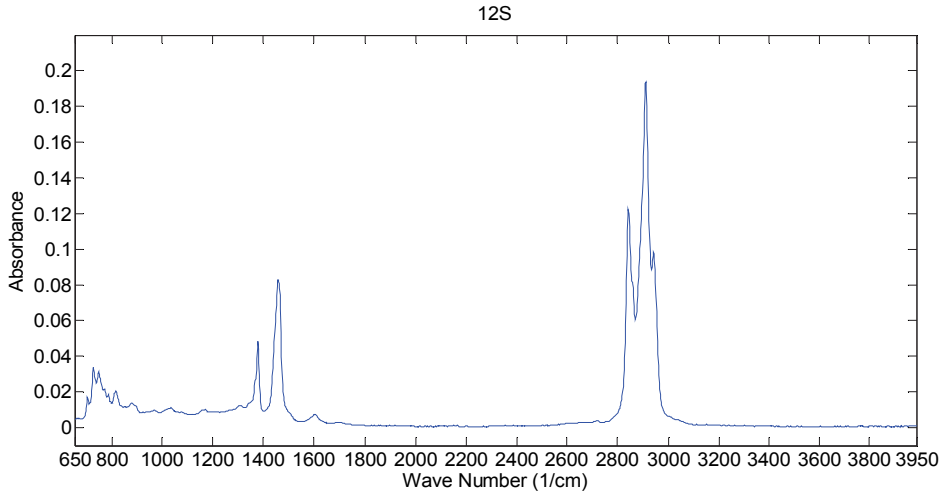


Figure A.12: FTIR spectrum of the crude oil labelled 12S in the crude oil dataset.

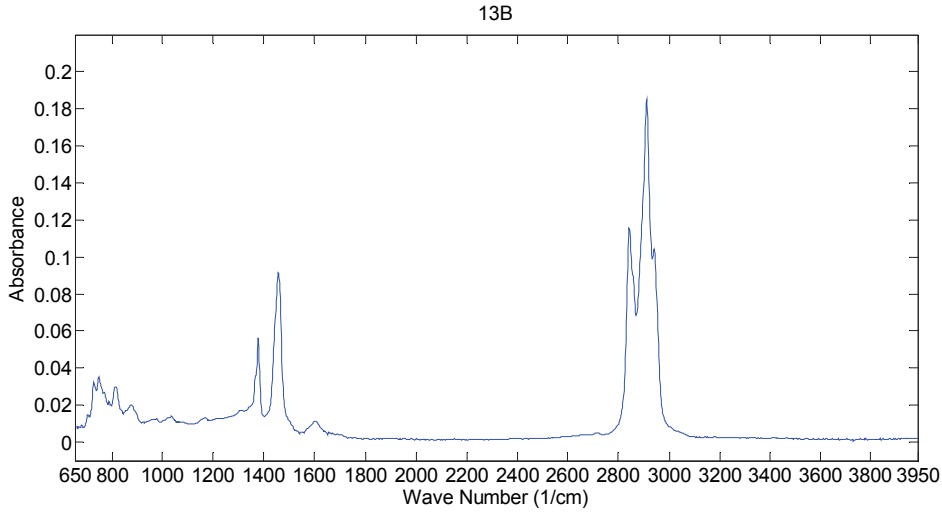


Figure A.13: FTIR spectrum of the crude oil labelled 13B in the crude oil dataset.

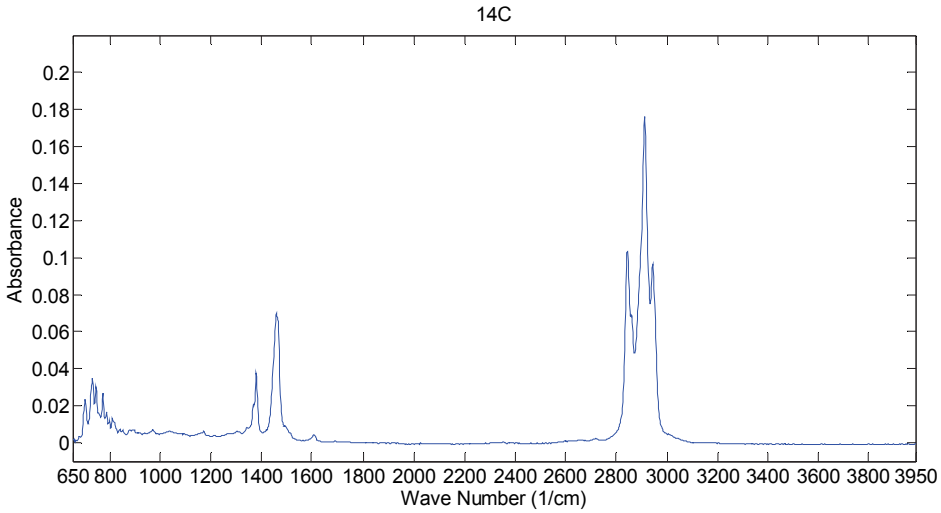


Figure A.14: FTIR spectrum of the crude oil labelled 14C in the crude oil dataset.

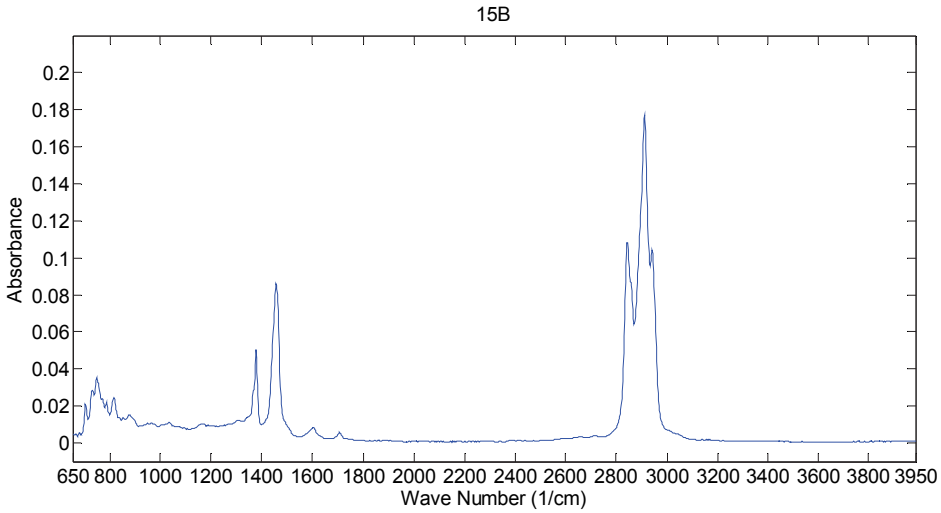


Figure A.15: FTIR spectrum of the crude oil labelled 15B in the crude oil dataset.

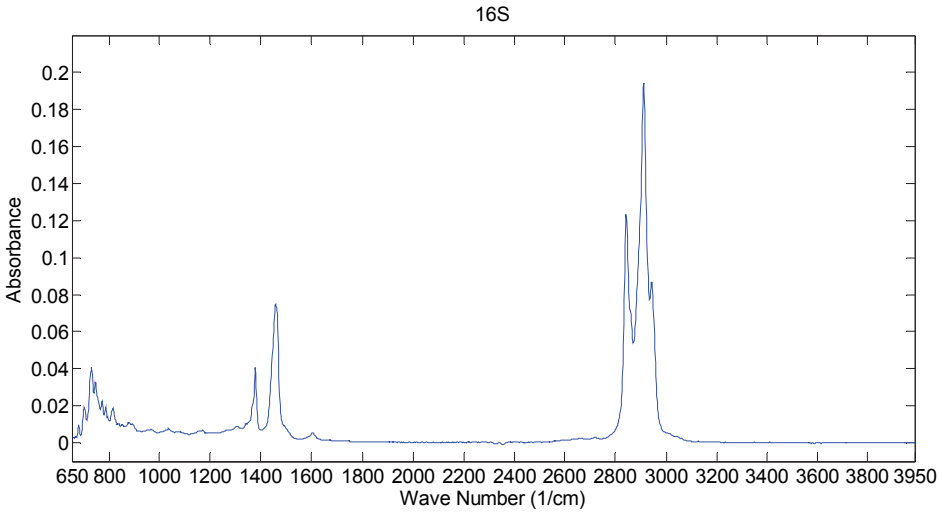


Figure A.16: FTIR spectrum of the crude oil labelled 16S in the crude oil dataset.

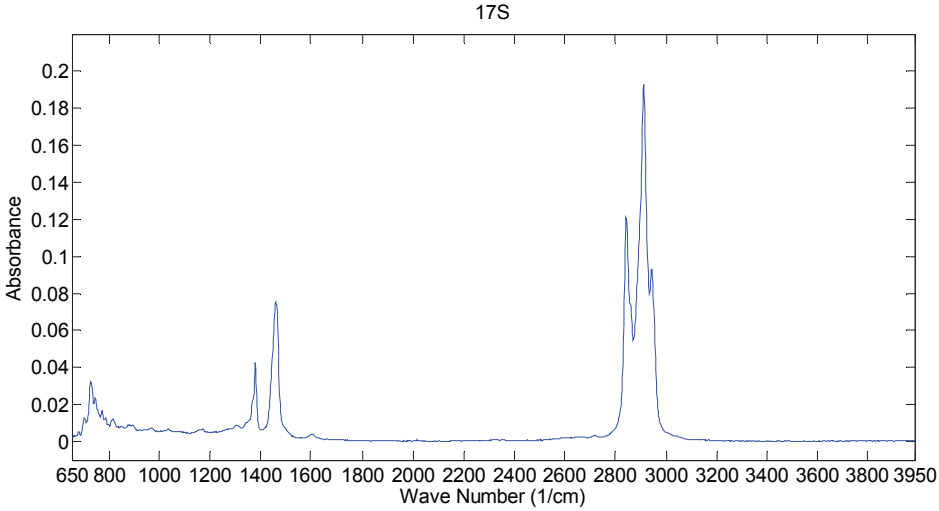


Figure A.17: FTIR spectrum of the crude oil labelled 17S in the crude oil dataset.

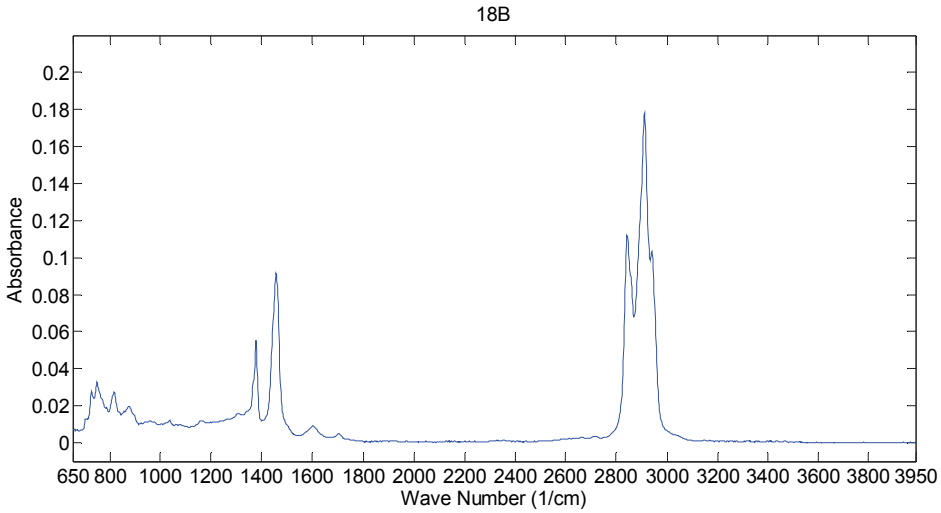


Figure A.18: FTIR spectrum of the crude oil labelled 18B in the crude oil dataset.

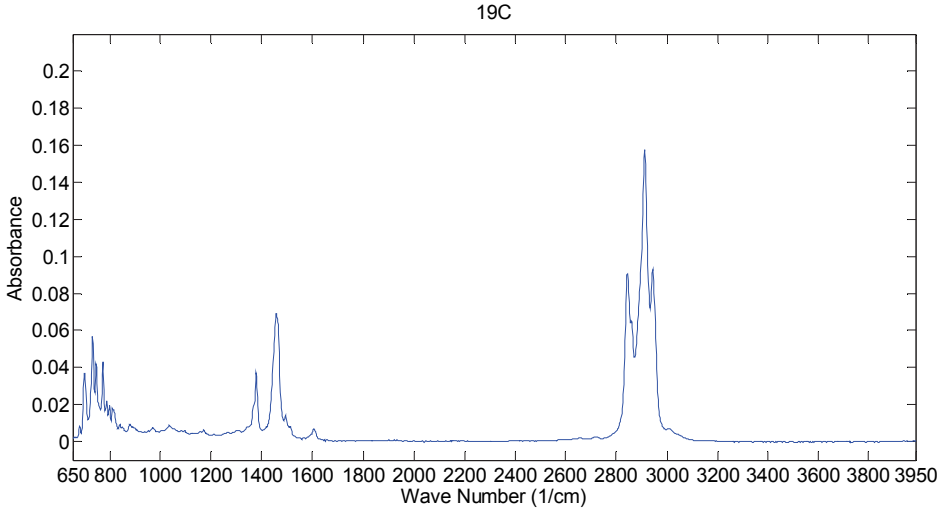


Figure A.19: FTIR spectrum of the crude oil labelled 19C in the crude oil dataset.

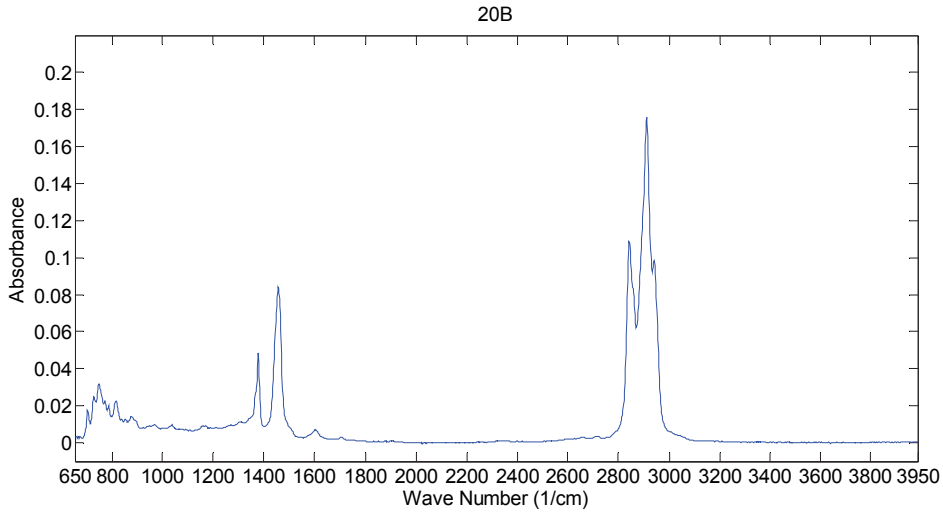


Figure A.20: FTIR spectrum of the crude oil labelled 20B in the crude oil dataset.

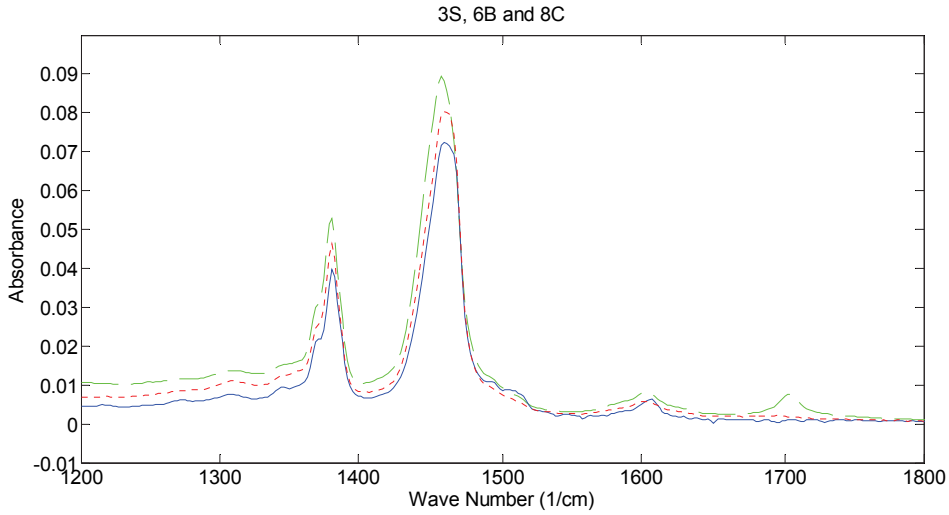


Figure A.21: Comparison of a biodegraded oil (6B), a nondegraded oil (3S) and a condensate (8C) in the 1200-1800 cm^{-1}- region of the FTIR spectrum. The condensate (8C) is marked as the blue, solid line, the biodegraded oil (6B) is marked as the green, dashed line and the nondegraded oil (3S) is marked as the red, dotted line.

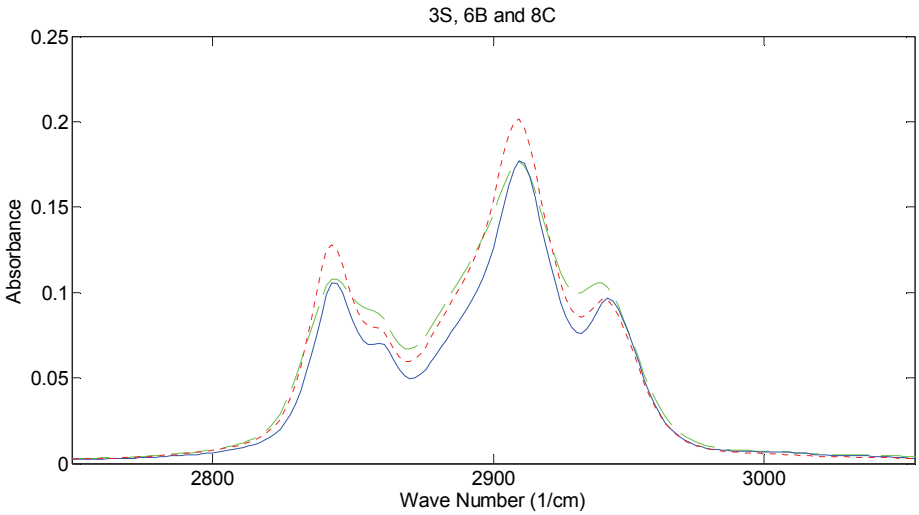


Figure A.22: Comparison of a biodegraded oil (6B), a nondegraded oil (3S) and a condensate (8C) in the 2750-3050 cm^{-1}- region of the FTIR spectrum. The condensate (8C) is marked as the blue, solid line, the biodegraded oil (6B) is marked as the green, dashed line and the nondegraded oil (3S) is marked as the red, dotted line.

Appendix B. WOGC chromatograms

In this section the WOGC chromatograms of all the 20 crude oils in the data set used in this thesis is presented. In the WOGC chromatograms the x-axis is the time window in which the different compounds elute, ranging from 0 to 135 minutes, while the y-axis is the voltage detected by the FID detector, ranging from 0 up to around 1250000 µV. The range is different for the different figures, some range from 0 to 80000 µV while others range from 0 to 1250000 µV.

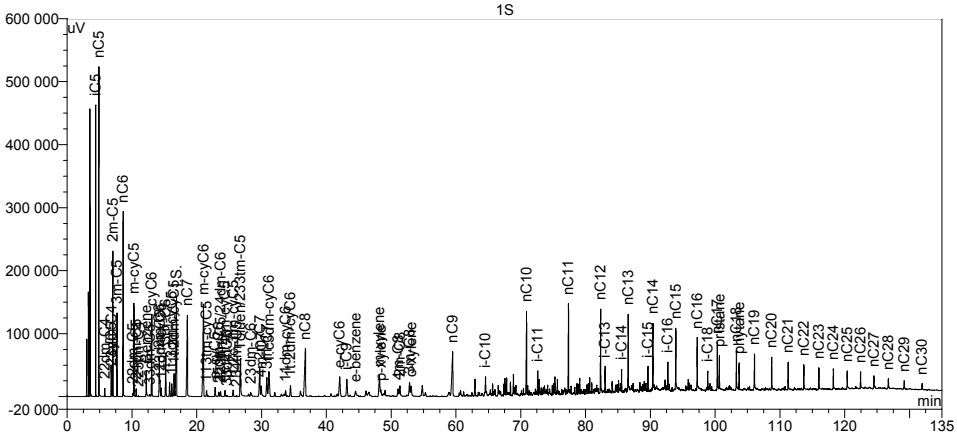


Figure B.1: WOGC chromatogram of the crude oil labelled 1S in the crude oil dataset.

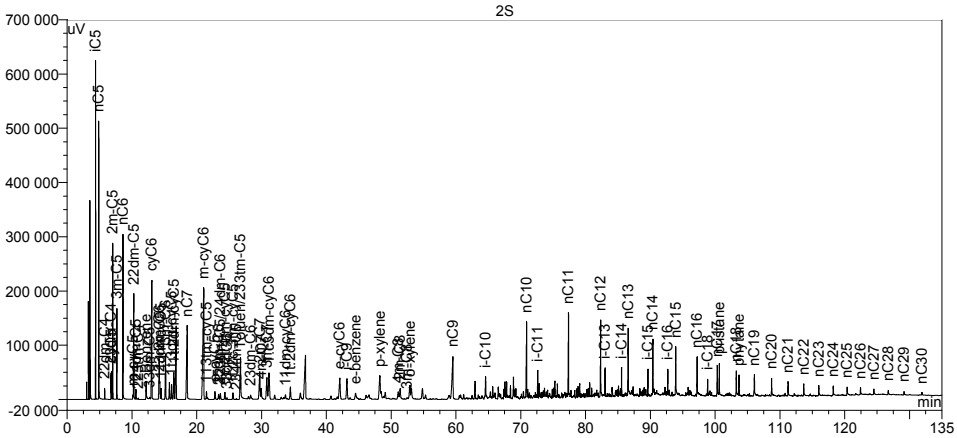


Figure B.2: WOGC chromatogram of the crude oil labelled 2S in the crude oil dataset.

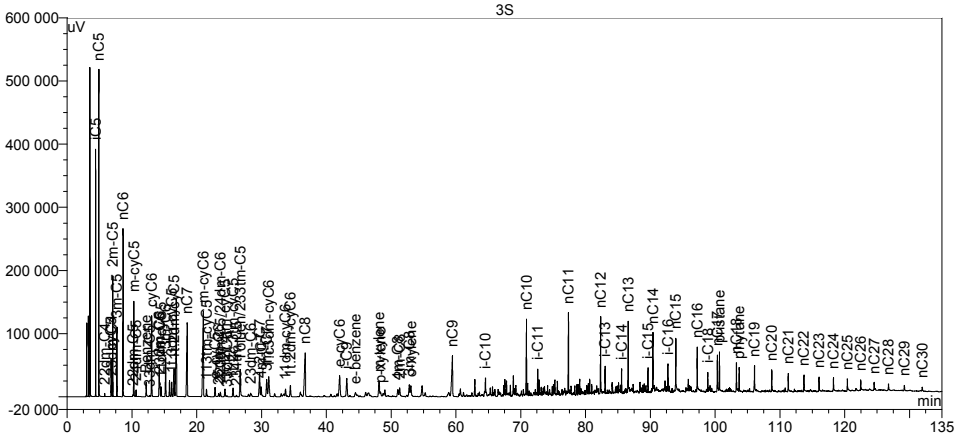


Figure B.3: WOGC chromatogram of the crude oil labelled 3S in the crude oil dataset.

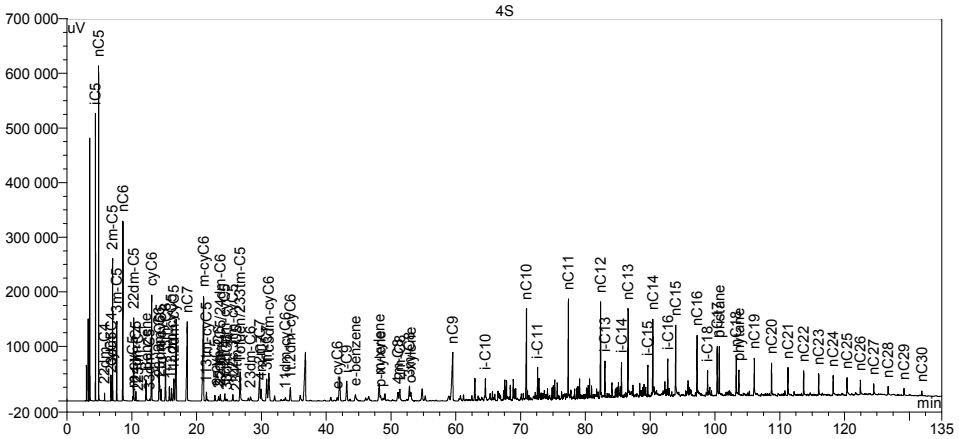


Figure B.4: WOGC chromatogram of the crude oil labelled 4S in the crude oil dataset.

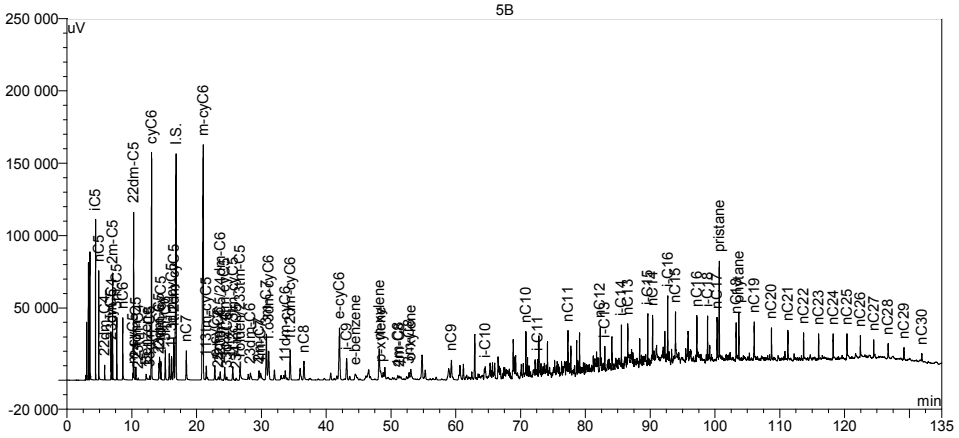


Figure B.5:WOGC chromatogram of the crude oil labelled 5B in the crude oil dataset.

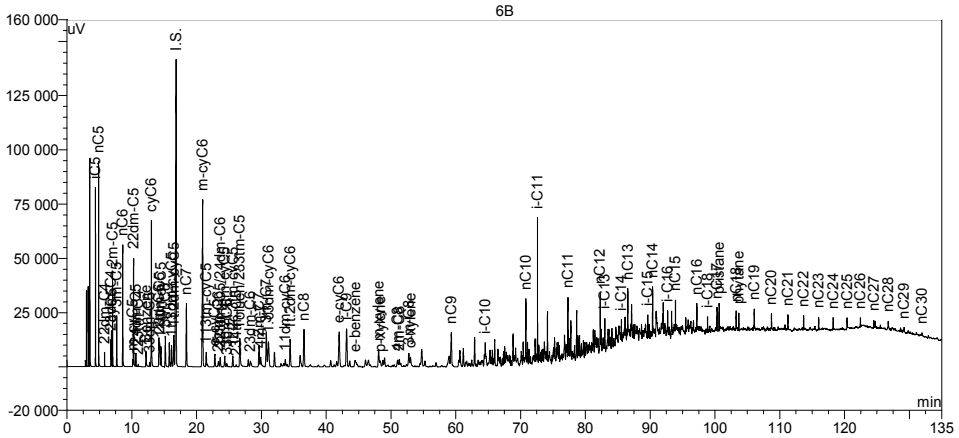


Figure B.6:WOGC chromatogram of the crude oil labelled 6B in the crude oil dataset.

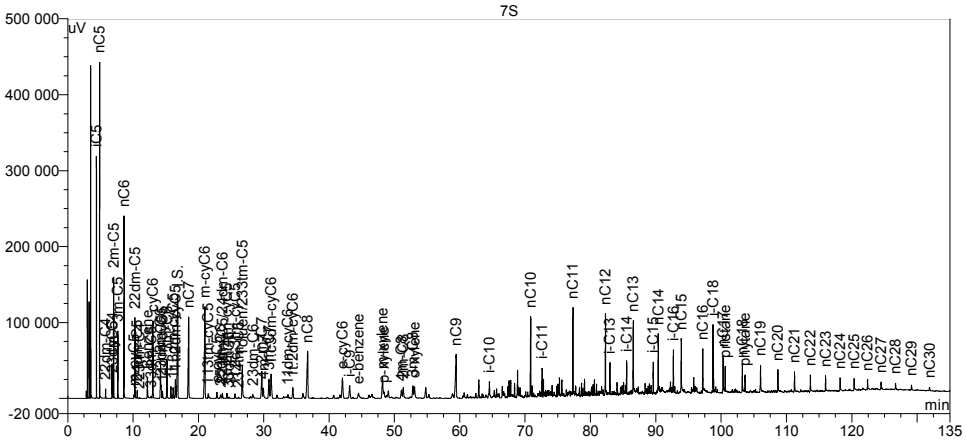


Figure B.7: WOGC chromatogram of the crude oil labelled 7S in the crude oil dataset.

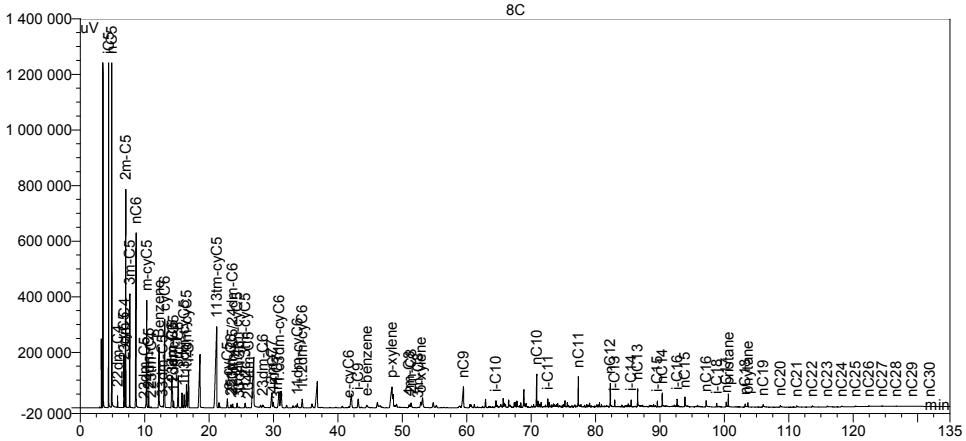


Figure B.8:WOGC chromatogram of the crude oil labelled 8C in the crude oil dataset.

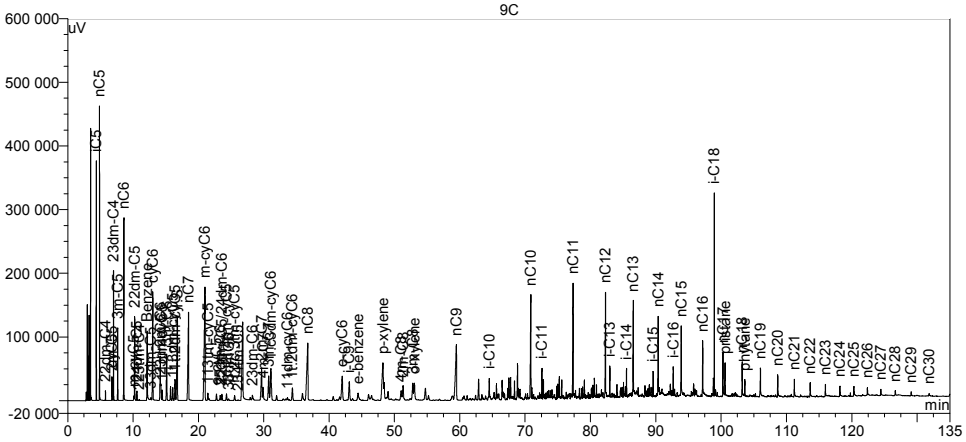


Figure B.9:WOGC chromatogram of the crude oil labelled 9C in the crude oil dataset.

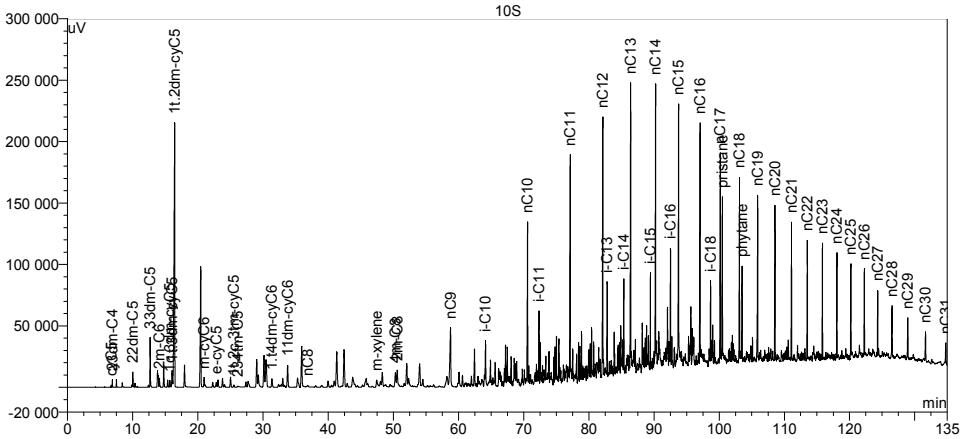


Figure B.10: WOGC chromatogram of the crude oil labelled 10S in the crude oil dataset.

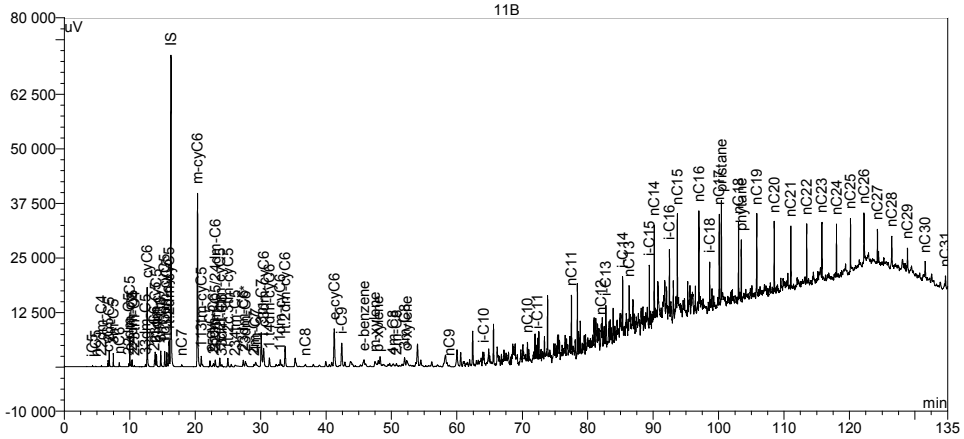


Figure B.11: WOGC chromatogram of the crude oil labelled 11B in the crude oil dataset.

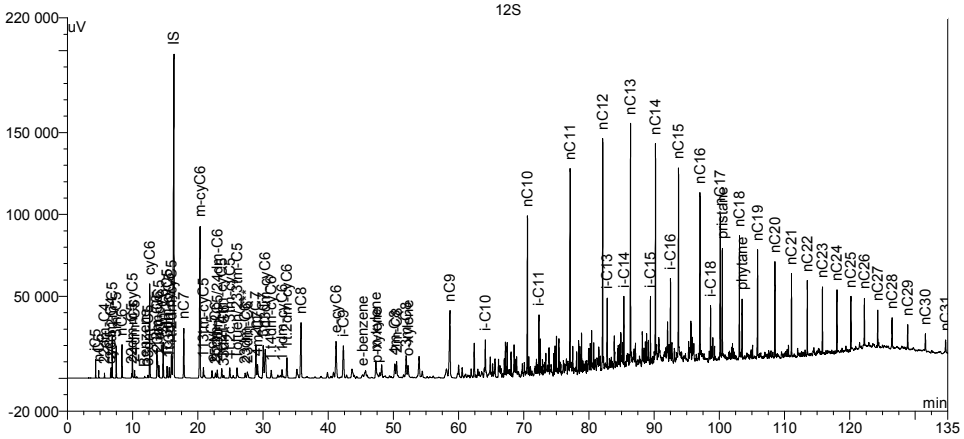


Figure B.12: WOGC chromatogram of the crude oil labelled 12S in the crude oil dataset.

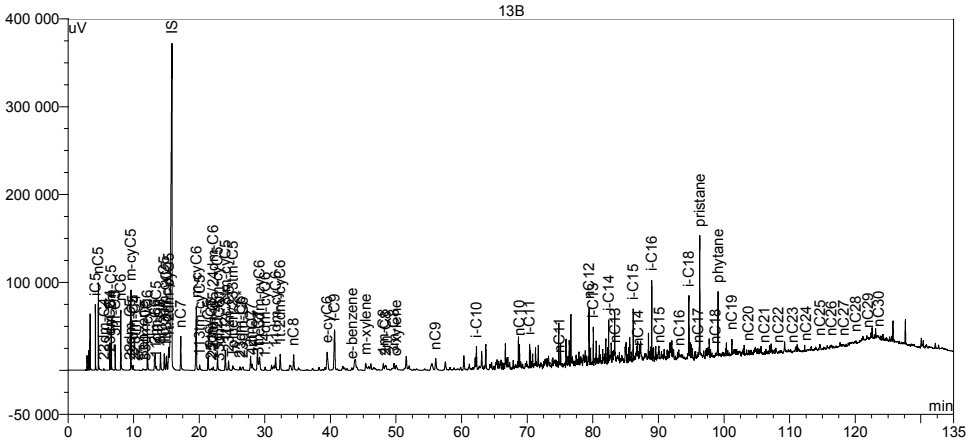


Figure B.13: WOGC chromatogram of the crude oil labelled 13B in the crude oil dataset.

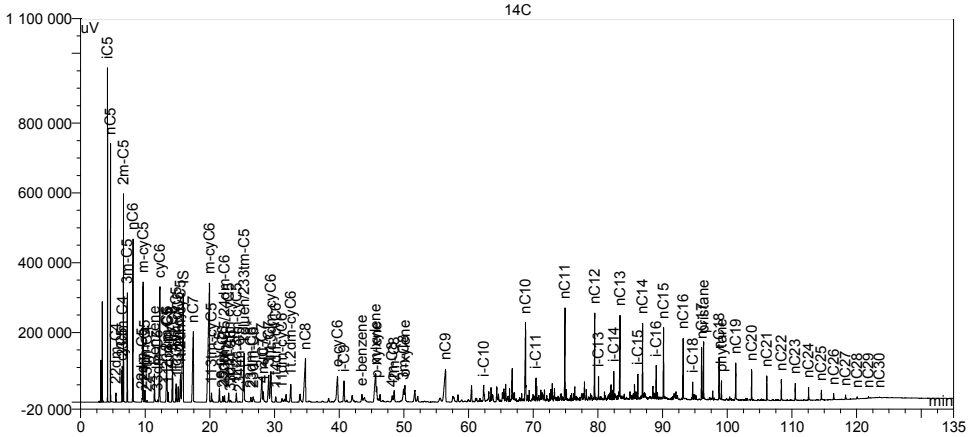


Figure B.14: WOGC chromatogram of the crude oil labelled 14C in the crude oil dataset.

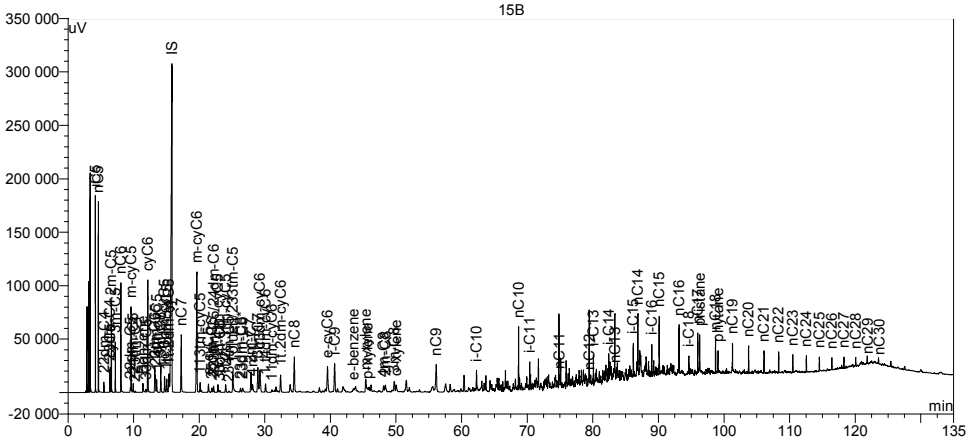


Figure B.15: WOGC chromatogram of the crude oil labelled 15B in the crude oil dataset.

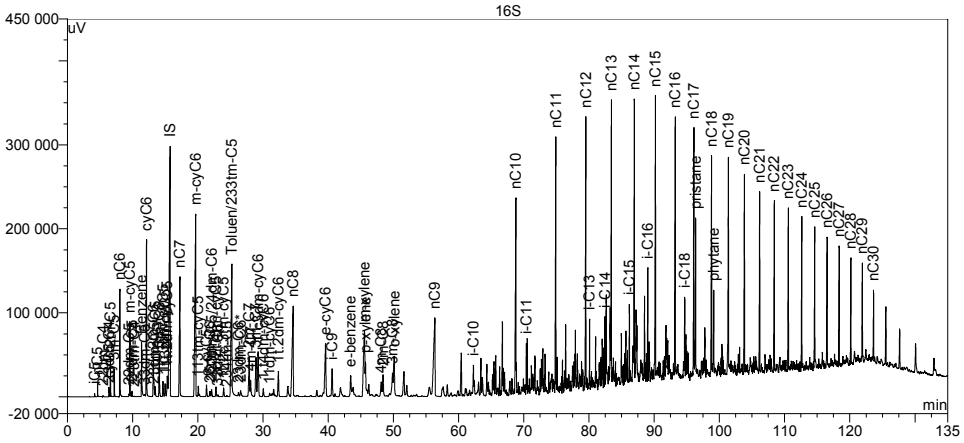


Figure B.16: WOGC chromatogram of the crude oil labelled 16S in the crude oil dataset.

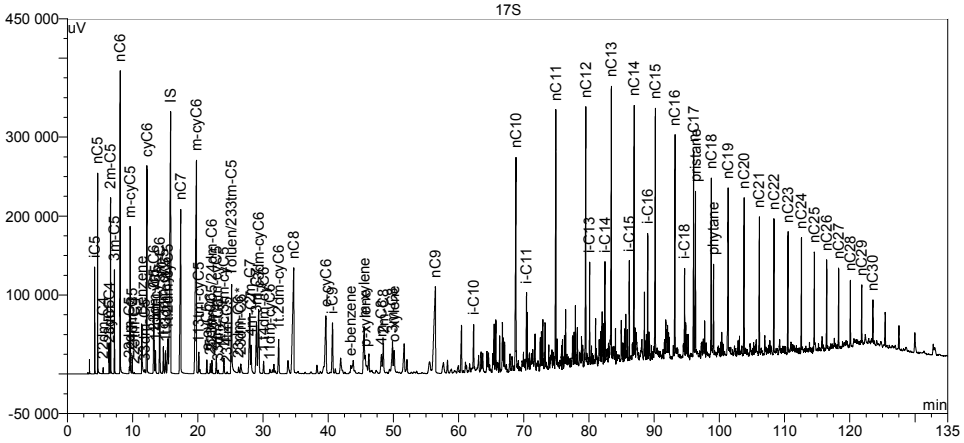


Figure B.17: WOGC chromatogram of the crude oil labelled 17S in the crude oil dataset.

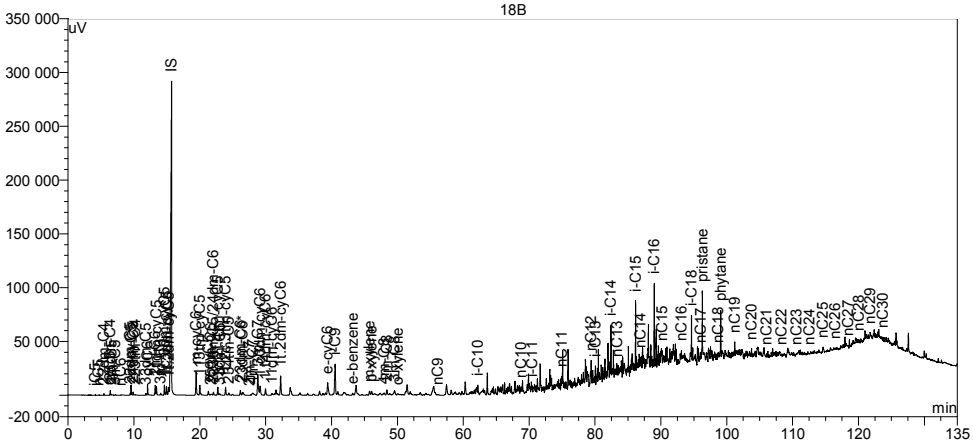


Figure B.18: WOGC chromatogram of the crude oil labelled 18B in the crude oil dataset.

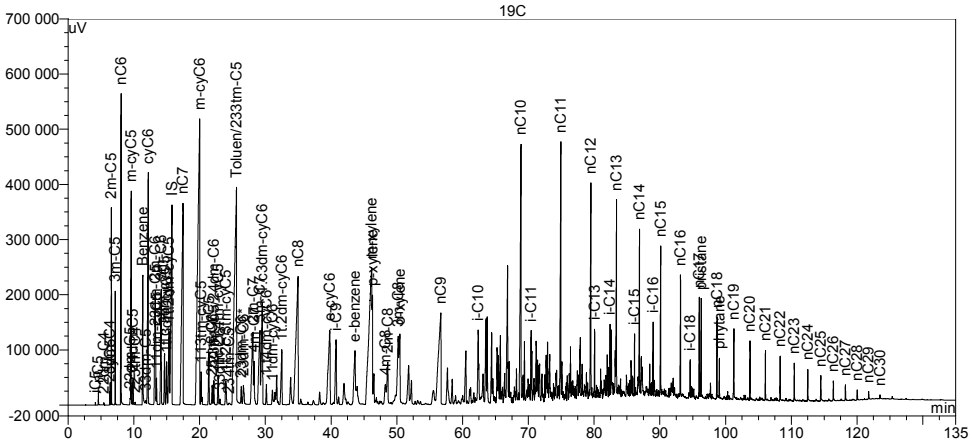


Figure B.19: WOGC chromatogram of the crude oil labelled 19C in the crude oil dataset.

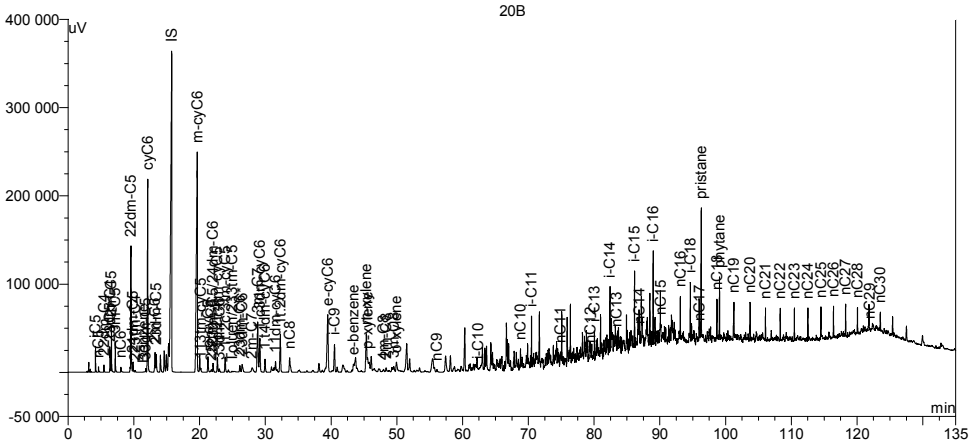


Figure B.20: WOGC chromatogram of the crude oil labelled 20B in the crude oil dataset.

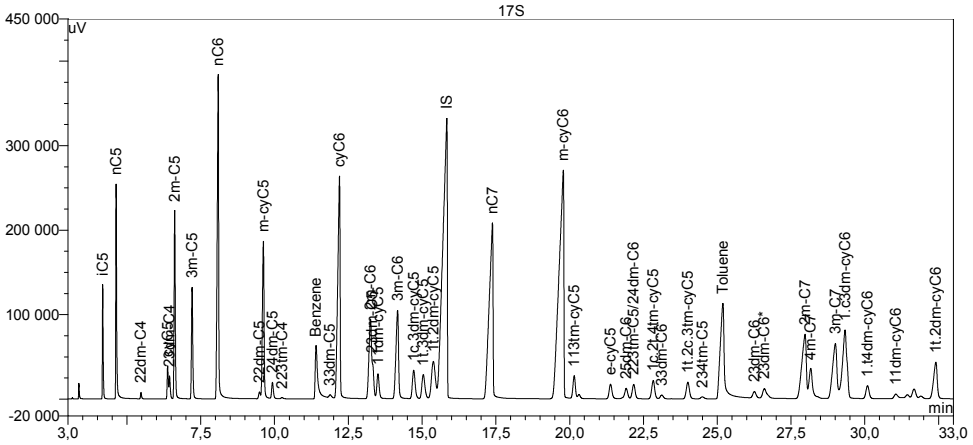


Figure B.21: WOGC chromatogram of the crude oil labelled 17S in the crude oil dataset, a closer look at the first 30 minutes of the chromatogram.

Table B.1: Variable coding and variable names for the GC data.

Variable Coding	**Variable name**
iC5	iso-pentane
nC5	n-pentane
22dm-C4	2,2-dimethylbutane
cyC5	cyclopentane
23dm-C4	2,3-dimethylbutane
2m-C5	2-methylpentane
3m-C5	3-methylpentane
nC6	n-hexane
22dm-C5	2,2-dimethylpentane
m-cyC5	methylcyclopentane

24dm-C5	2,4-dimethylpentane
223tm-C4	2,2,3-trimethylbutane
Benzene	Benzene
33dm-C5	3,3-dimethylpentane
cyC6	cyclohexane
2m-C6	2-methylhexane
23dm-C5	2,3-dimethylpentane
11dm-cyC5	1,1-dimethylcyclopentane
3m-C6	3-methylhexane
1c.3dm-cyC5	cis-1-3-dimethylcyclopentane
1t.3dm-cyC5 1t.2dm-cyC5	trans-1-3-dimethylcyclopentane trans-1-2-dimethylcyclopentane
nC7	n-heptane
m-cyC6	Methylcyclohexane
113tm-cyC5	1,1,3-trimethylcyclopentane
e-cyC5	Ethylcyclopentane
25dm-C6	2,5-dimethylhexane
223tm-C5/24dm-C6	2,2,3-trimethylpentane/2,4-dimethylhexane
1c.2t.4tm-cyC5	cis-1-trans-2-4-trimethylcyclopentane
33dmC6	3,3-dimethylhexane
1t.2c.3tm-cyC5	trans-1-cis-2-3-methylcyclopentane
234tm-C5	2,3,4-trimethylpentane
Toluene/233tm-C5	Toluene/2,3,3-trimethylpentane

23dm-C6	2,3-dimethylhexane
2m-C7	2-methylheptane
4m-C7	4-methylheptane
3m-C7	3-methylheptane
1.c3dm-cyC6	cis-1-3-dimethylcyclohexane
1.14dm-cyC6	trans-1-4-dimethylcyclohexane
11dm-cyC6	1,1-dimethylcyclohexane
1t.2dm-cyC6	trans-1-2-dimethylcyclohexane
nC8	n-octane
e-cyC6	Ethylcyclohexane
i-C9	iso-nonane
e-benzene	Ethylbenzene
m-xylene	meta-xylene
p-xylene	para-xylene
4m-C8	4-methyloctane
2m-C8	2-methyloctane
3m-C8	3-methyloctane
o-xylene	orto-xylene
nC9	n-nonane
i-C10	iso-decane
nC10	n-decane
i-C11	iso-undecane
nC11	n-undecane

nC12	n-dodecane
i-C13	iso-tridecane
i-C14	iso-tetradecane
nC13	n-tridecane
i-C15	iso-pentadecane
nC14	n-tetradecane
i-C16	iso-hexadecane
nC15	n-pentadecane
nC16	n-hexadecane
i-C18	iso-octadecane
nC17	n-heptadecane
pristane	pristane
nC18	n-octadecane
phytane	phytane
nC19	n-nonadecane
nC20	n-icosane
nC21	n-henicaosane
nC22	n-docosane
nC23	n-tricosane
nC24	n-tetracosane
nC25	n-pentacosane
nC26	n-hexacosane
nC27	n-heptacosane

nC28	n-octacosane
nC29	n-nonacosane
nC30	n-triacontane

Appendix C. Model characteristics

In this section some of the model characteristic for the multivariate models generated in this thesis is presented. In chapter C1.1 to C1.9 the characteristics of the models based on FTIR is shown, while in chapter C2.1 to C2.9 the characteristics of the models based on GC is shown.

In this section the term Standard Error of Cross Validation (SECV) is used, and is described as:

$$SECV = \sqrt{\frac{\sum_{i=1}^{N}\left(\hat{y}_{CV,i} - y_i\right)^2}{N}}, \qquad \text{Equation C.1}$$

where N= number of objects (samples), A= number of Principal Components, $\hat{y}_i$ = y predicted, y_i = y observed. The first term in the nominator is the predicted response for sample i, when this sample is kept out of the model building during the cross validation step.

The method of plotting SECV against number of components is used in order to determine how many components one should use in a model. A low SECV indicate that the predictive quality of a model is good. As seen in Figure C.1, the model with 4 components give the lowest SECV, but the model with 3 components should be considered as the best choice. The reason for that is that the more components in a

model, the higher the possibility of modelling noise, which could decrease the predictive quality of the model. Other methods for determining the predictive quality should also be applied, for example external validation, in order to determine how many PC's should be used.

C1.1 Models based on FTIR data

C1.1.1 Density based on FTIR

Table C1.1: Explained variance and Cross Validation Standard Deviation for the model for density based on FTIR data.

Number of Latent Variables	Explained variance	Cross Validation Standard Deviation
1	84.58%	0.48
2	95.58%	1.29
3	96.03%	1.06
4	97.06%	1.57
5	98.51%	1.93
6	99.01%	1.87

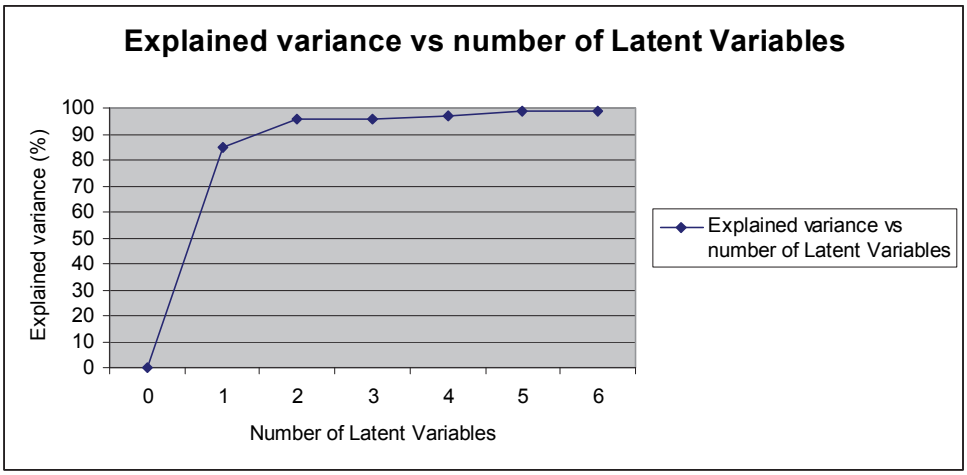


Figure C1.1: Explained variance plotted against the number of Latent Variables in the model for the model for density based on FTIR data.

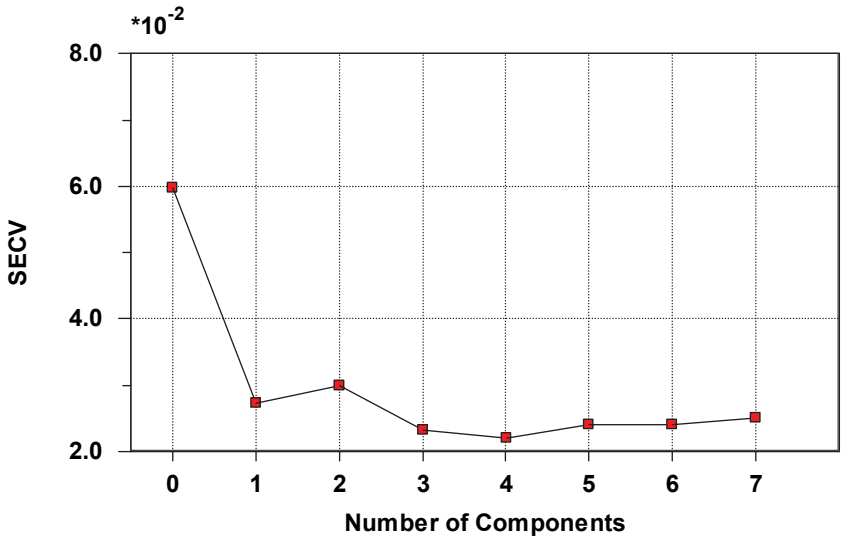


Figure C1.2: SECV plotted against number of components for the model for density based on FTIR data.

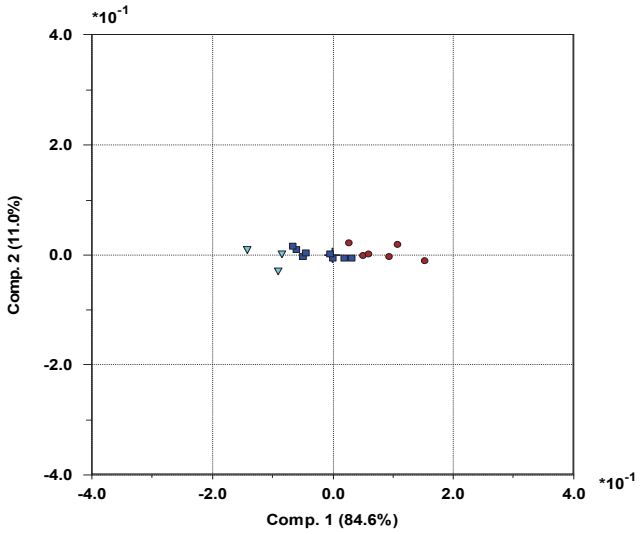


Figure C1.3: Scores for the model for density based on FTIR data. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates.

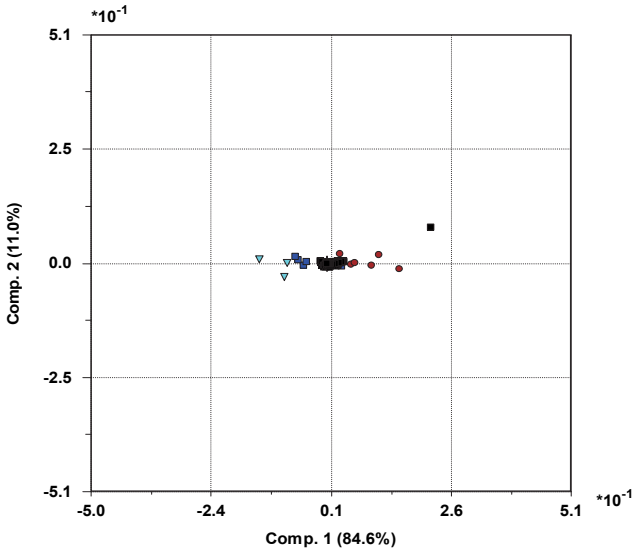


Figure C1.4: Biplot for the model for density based on FTIR data. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates. The black squares are the variables, with FTIR data in the centre and density on the far right.

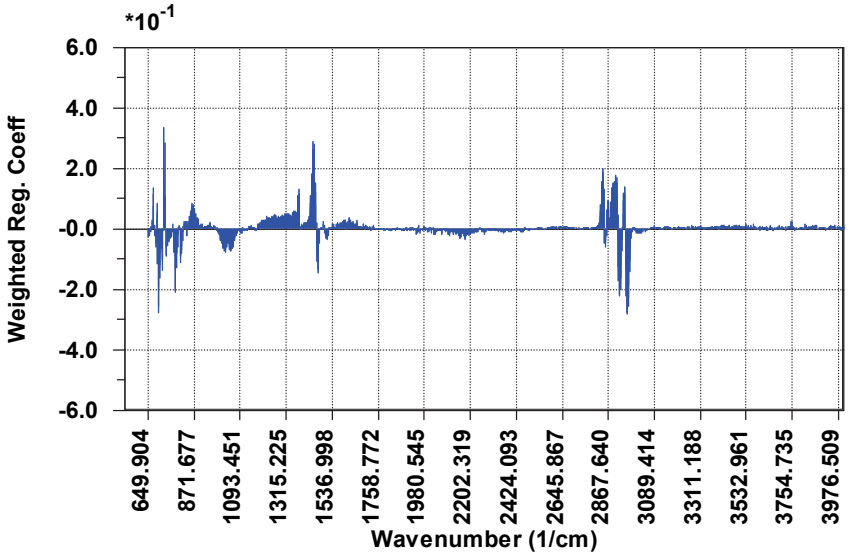


Figure C1.5: Weighted regression coefficients for the model for density based on FTIR data.

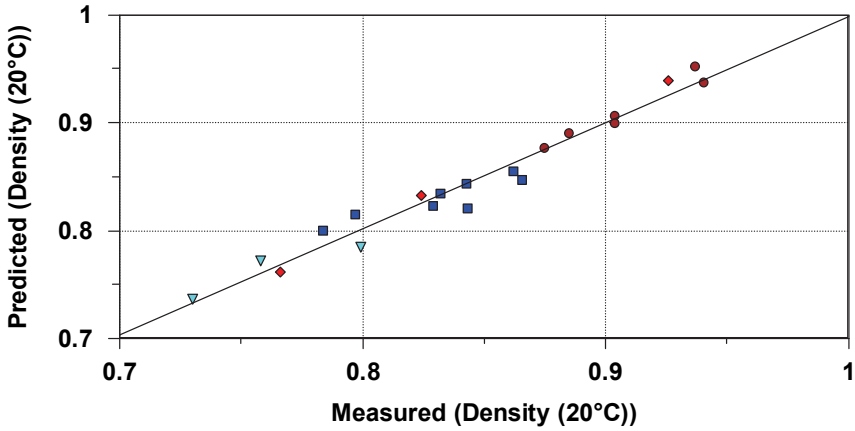


Figure C.1.6: Predicted value plotted against measured value for the model for density based on FTIR data. R^2 =0.964. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates, red diamonds=validation objects.

C1.1.2 Velocity of sound based on FTIR

Table C1.2: Explained variance and Cross Validation Standard Deviation for the model for velocity of sound based on FTIR data.

Number of components	Explained variance	Cross Validation Standard Deviation
1	85.10 %	0.44
2	88.97 %	1.09
3	97.21 %	1.01
4	98.52 %	1.58
5	99.18 %	2.10
6	99.60 %	2.03
7	99.76 %	1.92

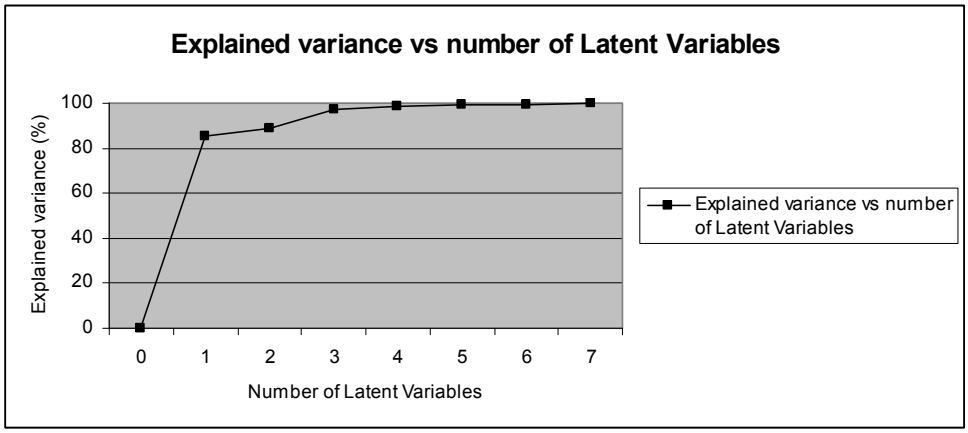


Figure C1.7: Explained variance plotted against the number of Latent Variables for the model for velocity of sound based on FTIR data.

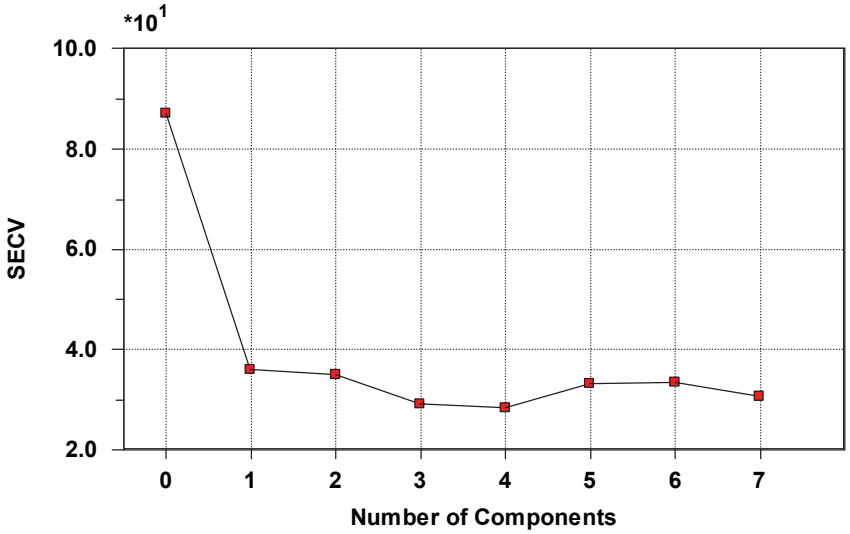


Figure C1.8: SECV plotted against number of components for the model for velocity of sound based on FTIR data.

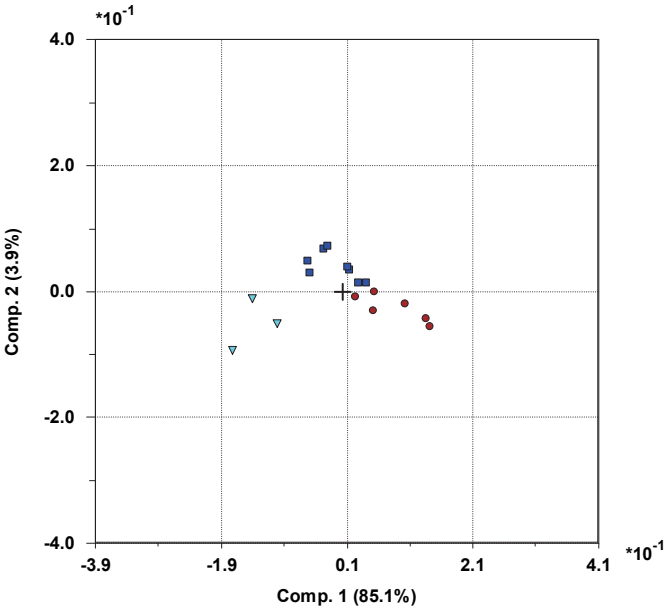


Figure C1.9: Scores for the model for velocity of sound based on FTIR data. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates.

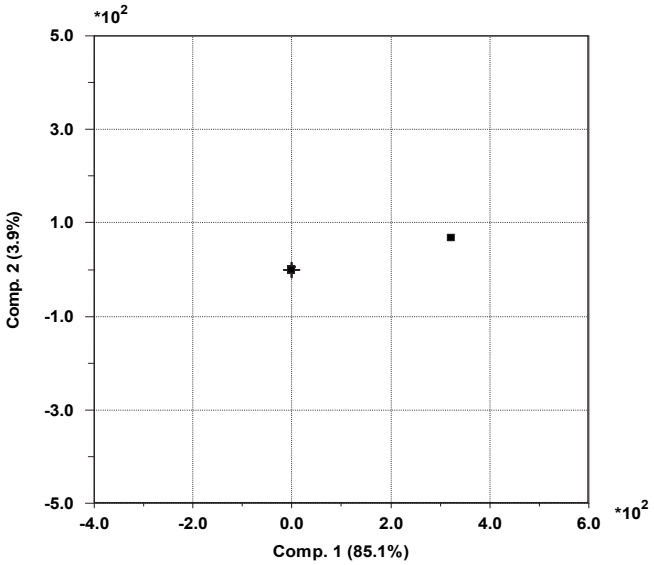


Figure C1.10: Biplot for the model for velocity of sound based on FTIR data. The black squares are the variables, with FTIR data and objects concentrated in the centre while the velocity of sound is placed on the far right.

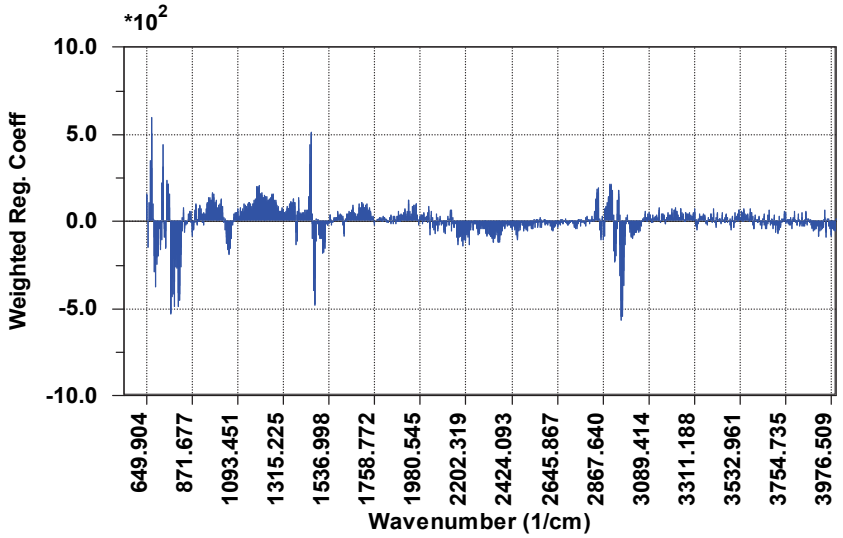


Figure C1.11: Weighted regression coefficients for the model for velocity of sound based on FTIR data.

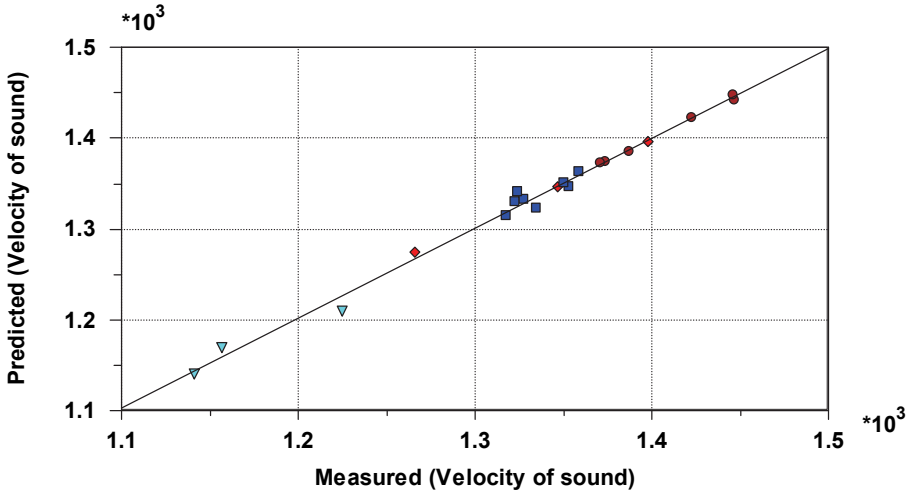


Figure C1.12: Predicted value plotted against measured value for the model for velocity of sound based on FTIR data. R^2=0.992. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates, red diamonds=validation objects.

C1.1.3 Static permittivity based on FTIR

Table C1.3: Explained variance and Cross Validation Standard Deviation for the model for static permittivity (ε_s) based on FTIR data.

Number of components	Explained variance	Cross Validation Standard Deviation
1	90.38 %	0.41
2	91.80 %	1.22
3	97.15 %	0.94
4	98.52 %	1.18
5	98.93 %	3.04
6	99.10 %	2.21
7	99.64 %	2.05

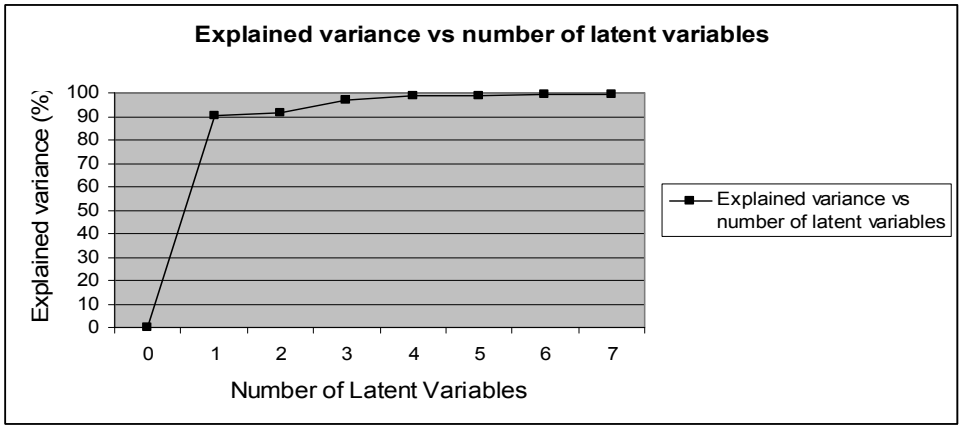


Figure C1.13: Explained variance plotted against the number of Latent Variables for the model for static permittivity (ε_s) based on FTIR data.

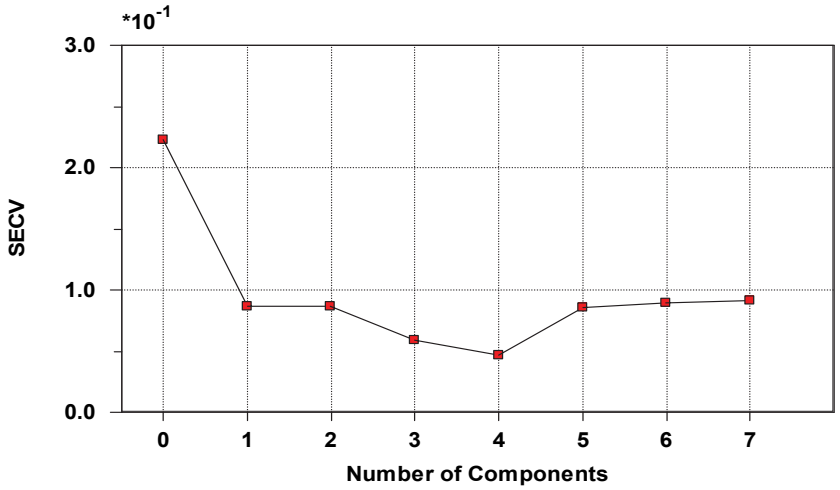


Figure C1.14: SECV plotted against number of components for the model for static permittivity (ε_s) based on FTIR data.

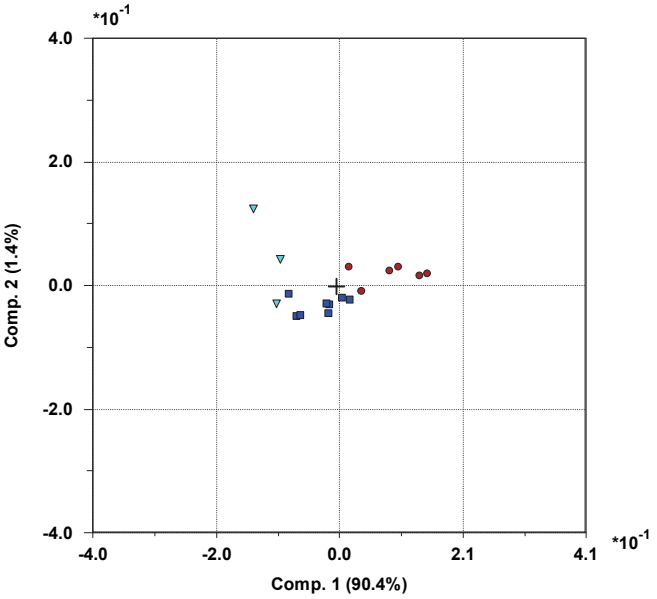


Figure C1.15: Scores for the model for static permittivity (ε_s) based on FTIR data. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates.

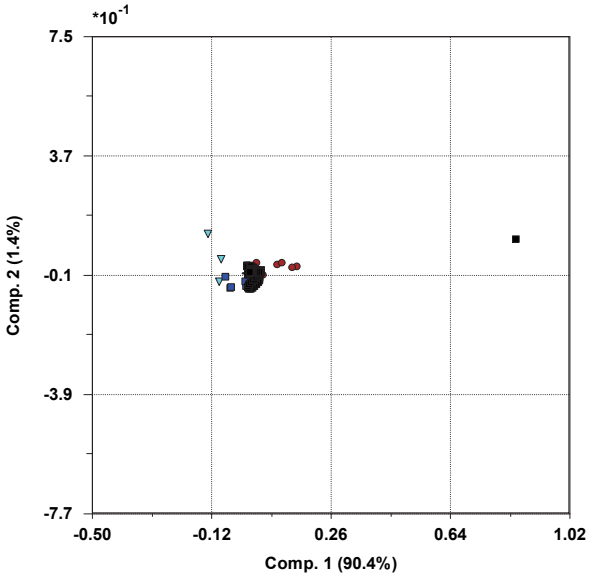


Figure C1.16: Biplot for the model for static permittivity (ε_s) based on FTIR data. Brown circles= biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates. The black squares are the variables, with FTIR data and objects concentrated in the centre while static permittivity is placed on the far right.

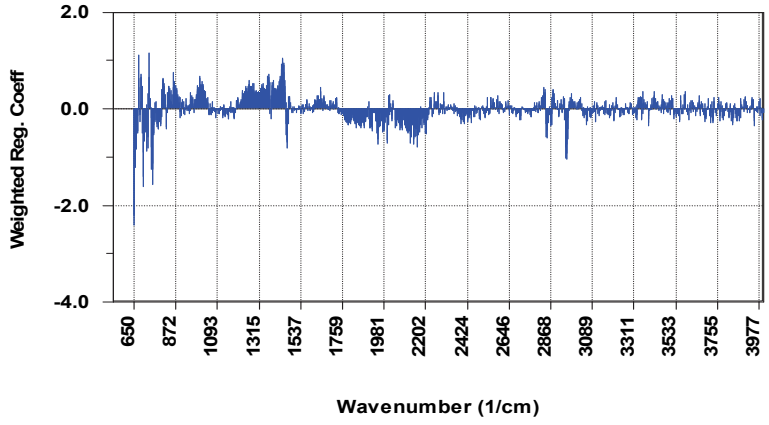


Figure C1.17: Weighted regression coefficients for the model for static permittivity (ε_s) based on FTIR data

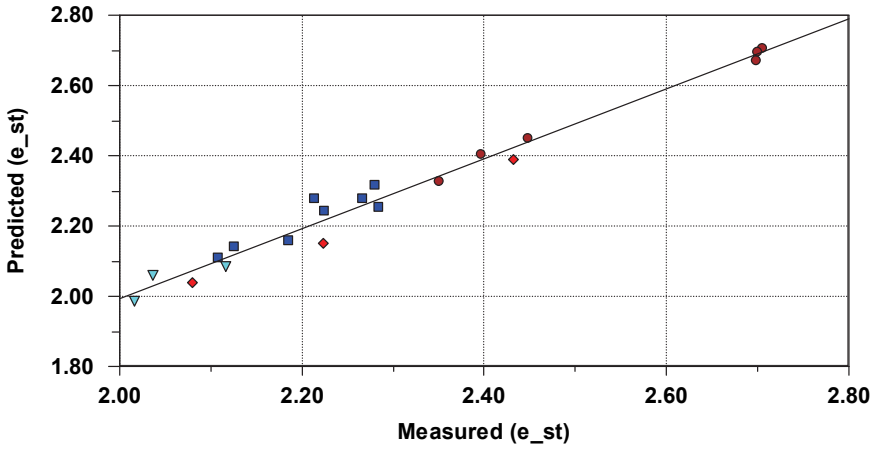


Figure C1.18: Predicted value plotted against measured value for the model for static permittivity (ε_s) based on FTIR data. R^2=0.978. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates, red diamonds=validation objects.

C1.1.4 High frequency permittivity based on FTIR

Table C1.4: Explained variance and Cross Validation Standard Deviation for the model for high frequency permittivity (ε_∞) based on FTIR data.

Number of components	Explained variance	Cross Validation Standard Deviation
1	87.11 %	0.48
2	89.07 %	1.76
3	94.99 %	1.45
4	96.58 %	1.83
5	98.29 %	2.12
6	98.51 %	1.89
7	99.21 %	1.94

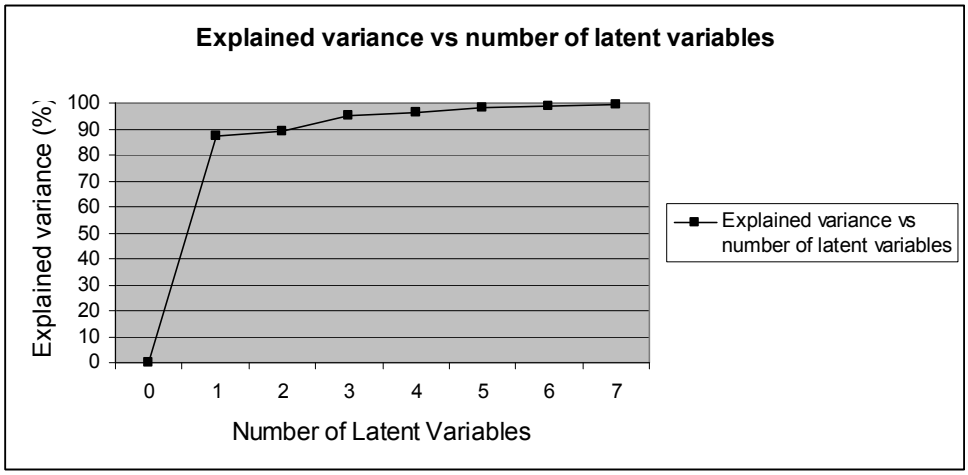


Figure C1.19: Explained variance plotted against the number of Latent Variables for the model for high frequency permittivity (ε_∞) based on FTIR data.

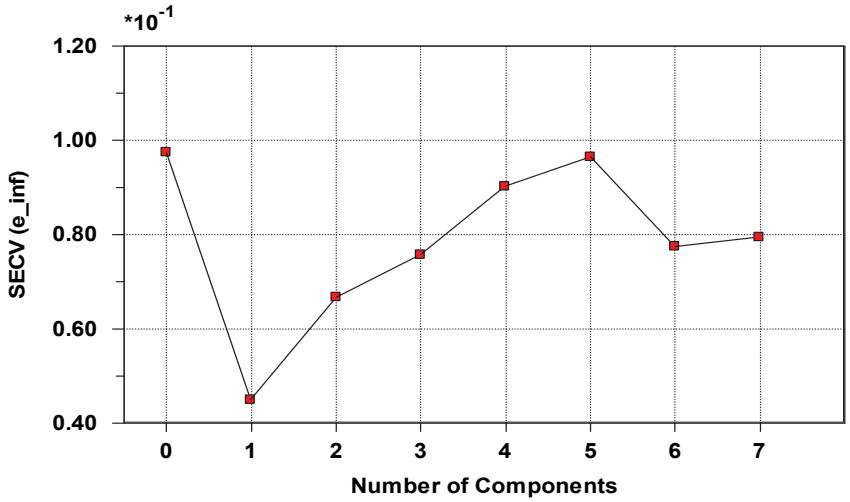


Figure C1.20: SECV plotted against number of components for the model for high frequency permittivity (ε_∞) based on FTIR data.

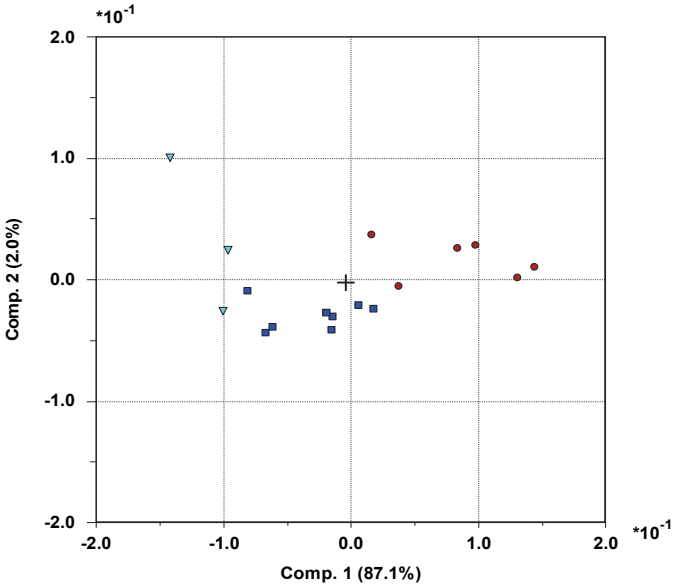


Figure C1.21: Scores for the model for high frequency permittivity (ε_∞) based on FTIR data. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates.

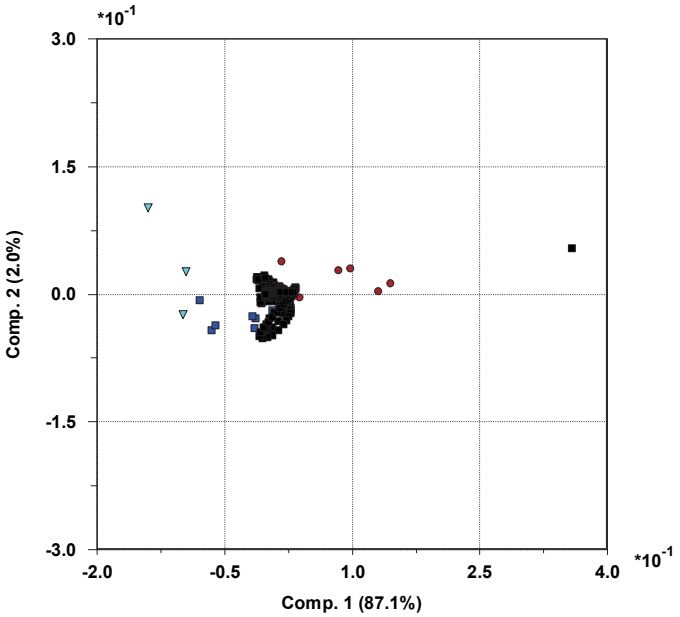


Figure C1.22: Biplot for the model for high frequency permittivity (ε_∞) based on FTIR data. Brown circles= biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates. The black squares are the variables, with FTIR data and objects concentrated in the centre while high frequency permittivity is placed on the far right.

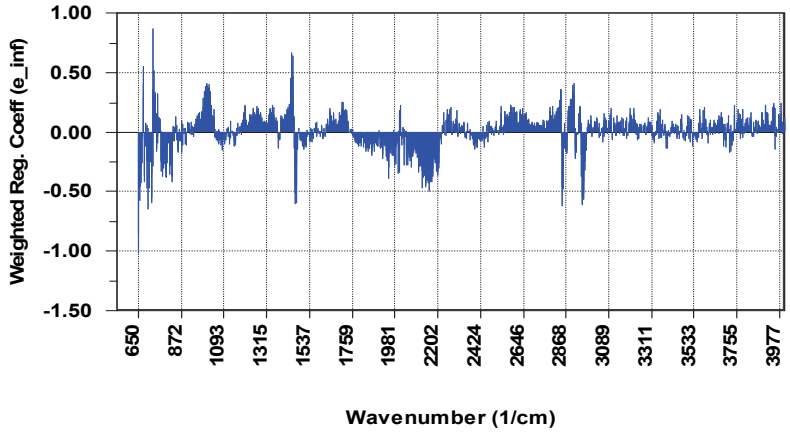


Figure C1.23: Weighted regression coefficients for the model for high frequency permittivity (ε_∞) based on FTIR data.

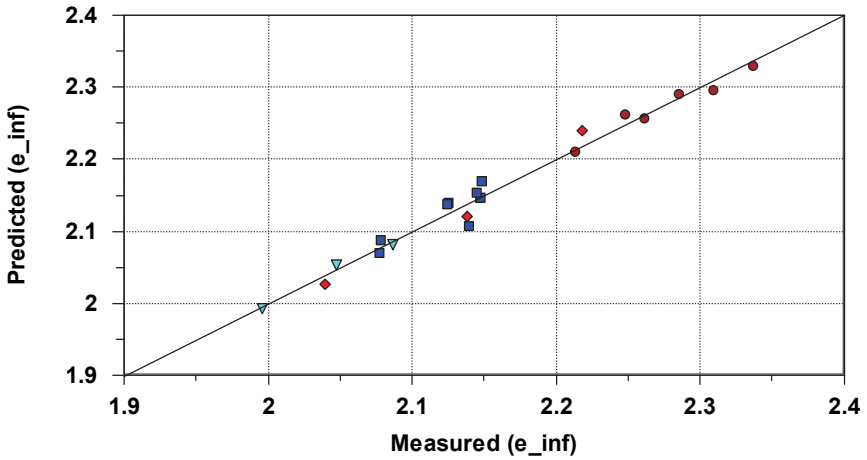


Figure C1.24: Predicted value plotted against measured value for the model for high frequency permittivity (ε_∞) based on FTIR data. R^2=0.980. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates, red diamonds=validation objects.

C1.1.5 α based on FTIR

Table C1.5: Explained variance and Cross Validation Standard Deviation for the permittivity variable α based on FTIR data.

Number of components	Explained variance	Cross Validation Standard Deviation
1	22.23 %	1.24
2	31.47 %	1.7
3	69.71 %	1.45
4	77.51 %	1.63
5	87.84 %	1.99
6	92.51 %	2.24
7	95.29 %	1.97
8	96.45 %	2.15
9	97.74 %	2.16

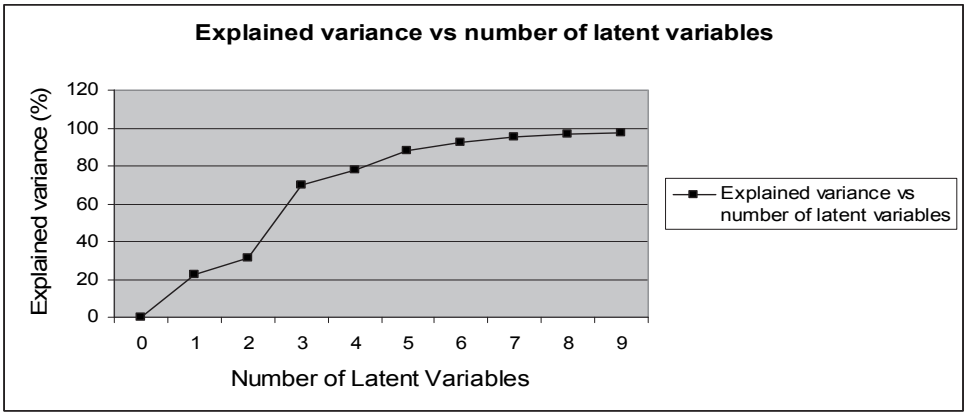


Figure C1.25: Explained variance plotted against the number of Latent Variables for the model for the permittivity variable α based on FTIR data.

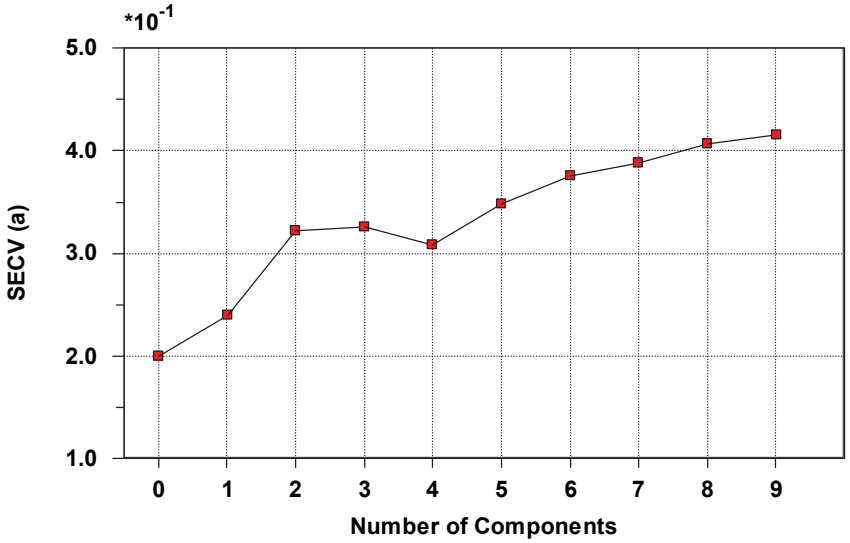


Figure C1.26: SECV plotted against number of components for the model for the permittivity variable α based on FTIR data.

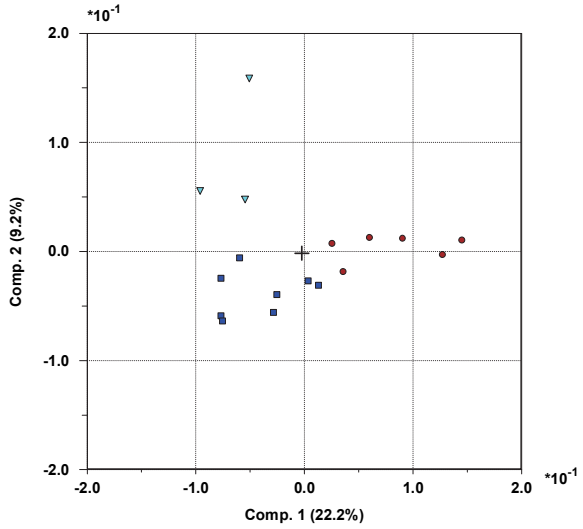


Figure C1.27: Scores for the model for the permittivity variable α based on FTIR data. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates.

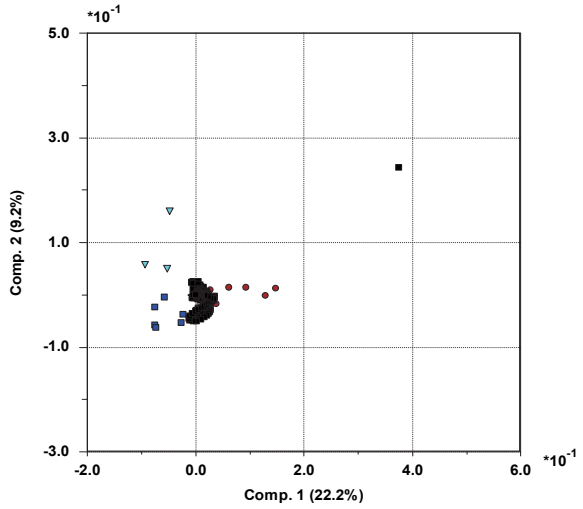


Figure C1.28: Biplot for the model for the permittivity variable α based on FTIR data. Brown circles= biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates. The black squares are the variables, with FTIR data and objects concentrated in the centre while the permittivity variable α is placed somewhat up and right from the centre.

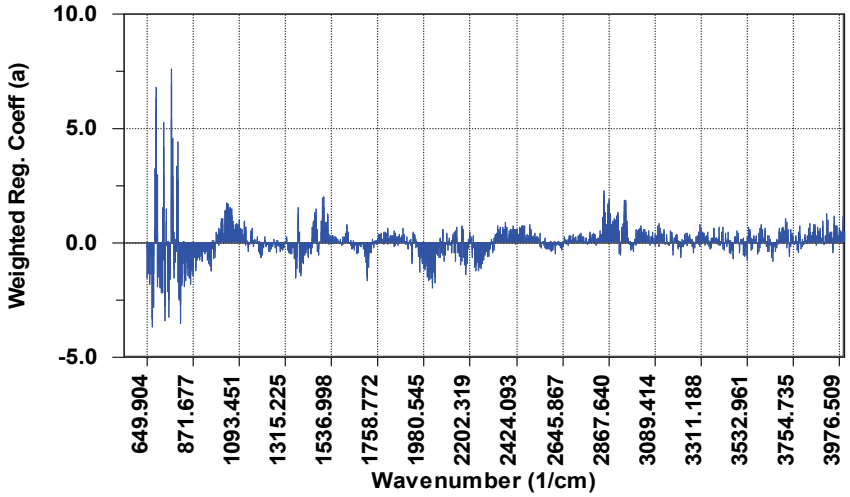


Figure C1.29: Weighted regression coefficients for the model for the permittivity variable α based on FTIR data.

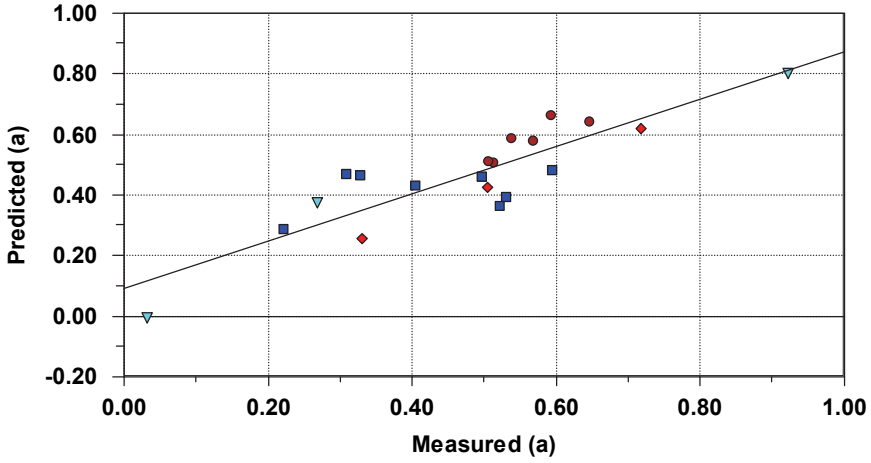


Figure C1.30: Predicted value plotted against measured value for the model for the permittivity variable α based on FTIR data. R^2=0.775. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates, red diamonds=validation objects.

C1.1.6 τ based on FTIR

Table C1.6: Explained variance and Cross Validation Standard Deviation for the permittivity variable τ based on FTIR data.

Number of components	Explained variance	Cross Validation Standard Deviation
1	23.77 %	1.09
2	32.01 %	1.26
3	71.14 %	1.57
4	83.92 %	1.57
5	88.18 %	2.19
6	91.75 %	2.16
7	95.26 %	2.1
8	97.73 %	2.07
9	99.65 %	2.15

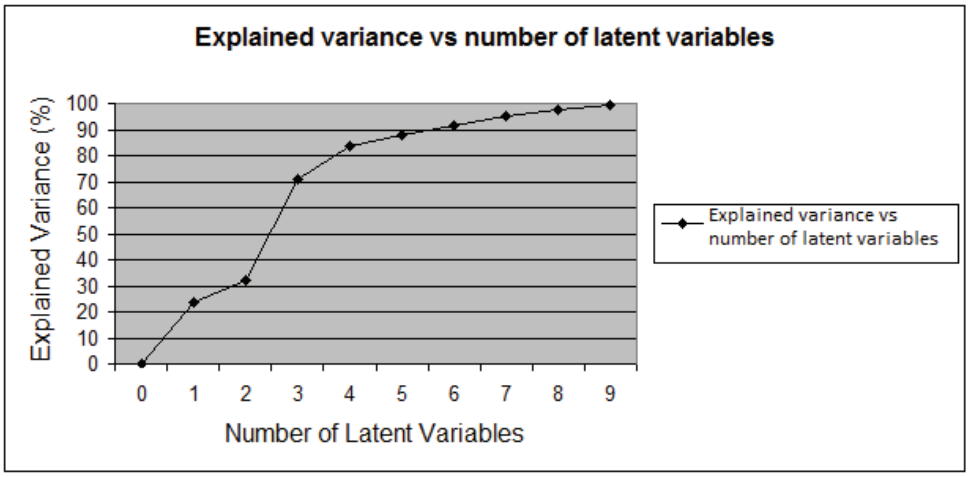


Figure C1.31: Explained variance plotted against the number of Latent Variables for the model for the permittivity variable τ based on FTIR data.

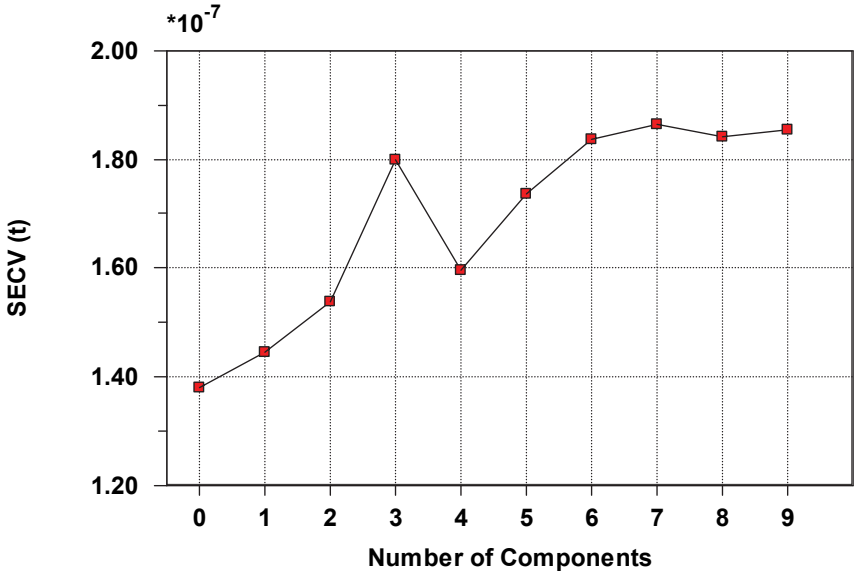


Figure C1.32: SECV plotted against number of components for the model for the permittivity variable τ based on FTIR data.

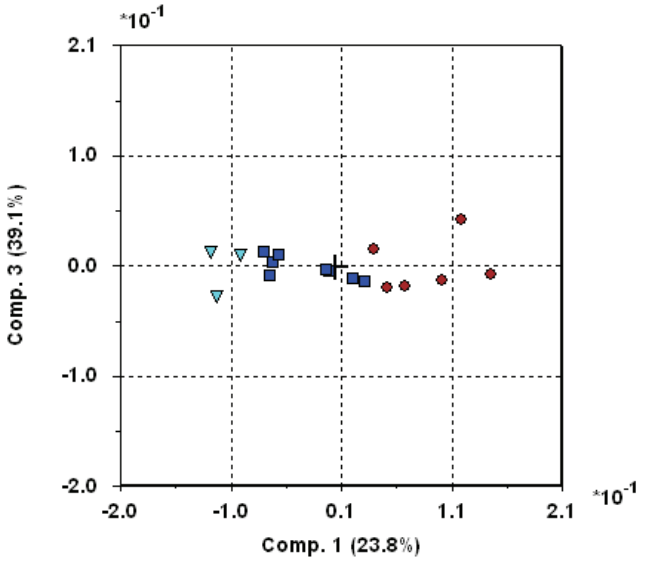


Figure C1.33: Scores for the model for the permittivity variable τ based on FTIR data, Component 1 vs Component 3. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates.

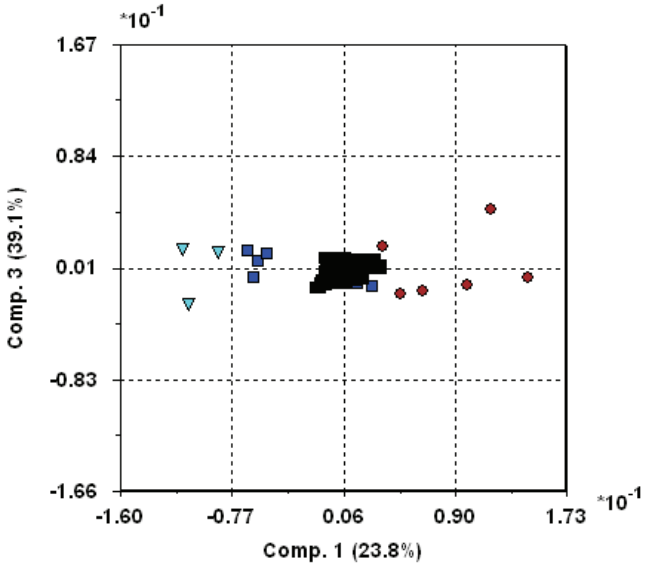


Figure C1.34: Biplot for the model for the permittivity variable τ based on FTIR data. Brown circles= biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates. The black squares are the variables, with FTIR data and objects concentrated in the centre. The permittivity variable τ is placed in the centre as well.

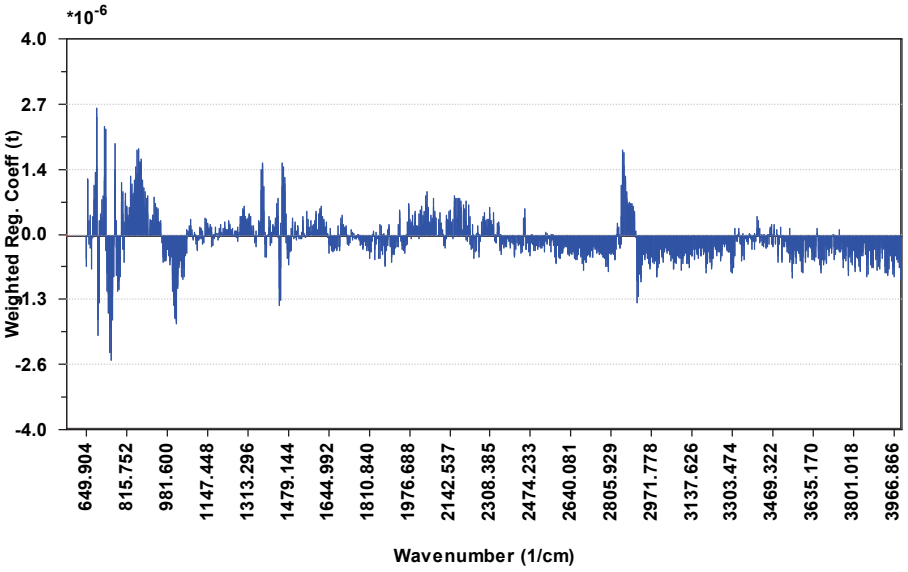


Figure C1.35: Weighted regression coefficients for the model for the permittivity variable τ based on FTIR data.

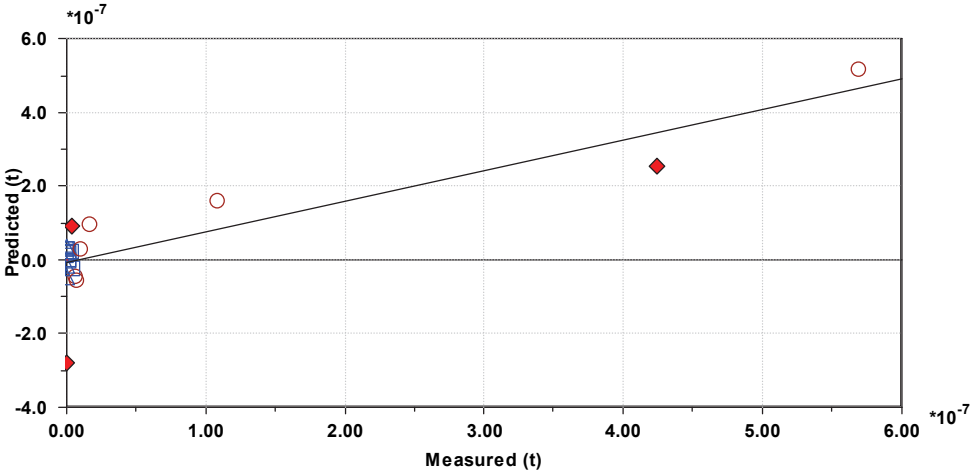


Figure C1.36: Predicted value plotted against measured value for the model for the permittivity variable τ based on FTIR data. R^2=0.721. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates, red diamonds=validation objects.

C1.1.7 TAN based on FTIR

Table C1.7: Explained variance and Cross Validation Standard Deviation for TAN based on FTIR data.

Number of components	Explained variance	Cross Validation Standard Deviation
1	68.61%	0.59
2	77.45%	1.09
3	89.99%	0.84
4	93.88%	1.07
5	95.74%	1.12
6	98.09%	1.23
7	98.35%	1.07
8	98.58%	1.11
9	99.62%	1.19
10	99.92%	1.12

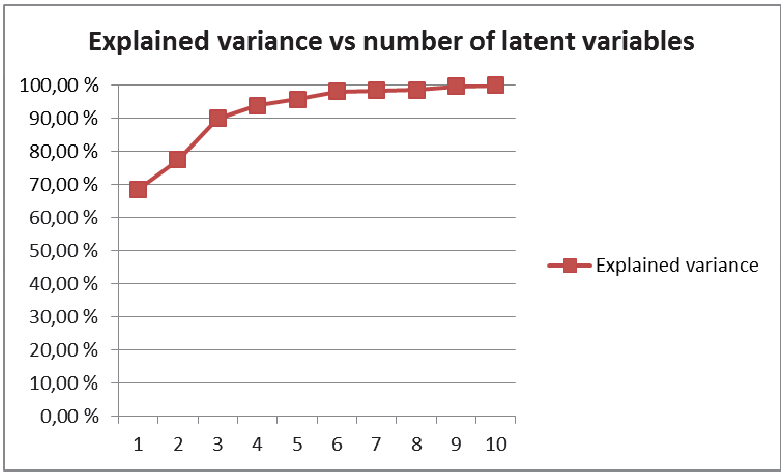


Figure C1.37: Explained variance plotted against the number of Latent Variables for the model for TAN based on FTIR data.

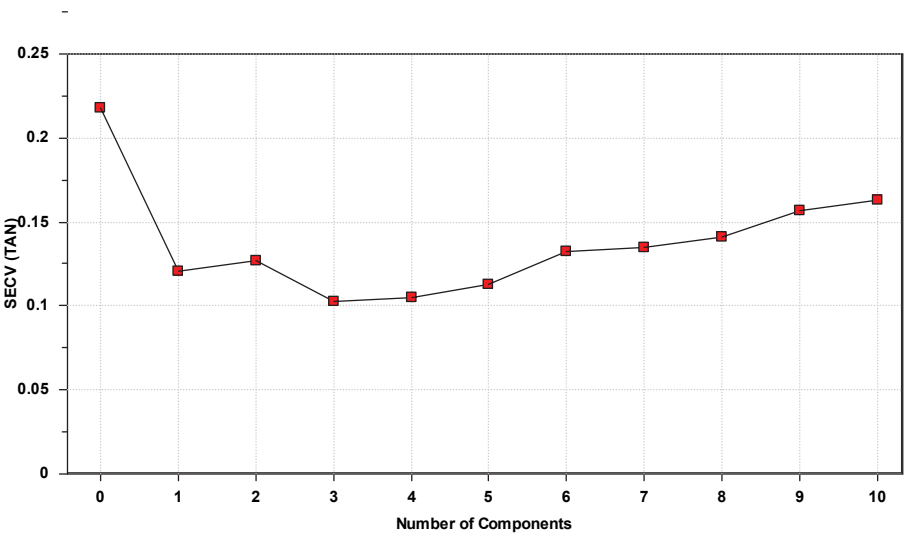


Figure C1.38: SECV plotted against number of components for the model for TAN based on FTIR data.

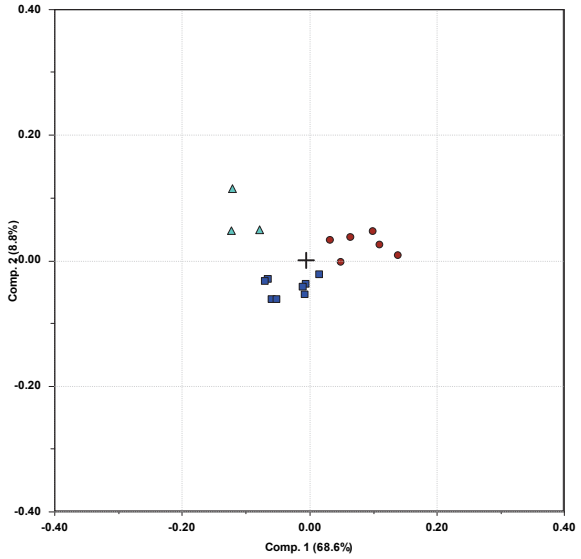


Figure C1.39: Scores for the model for TAN based on FTIR data, Component 1 vs Component 2. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates.

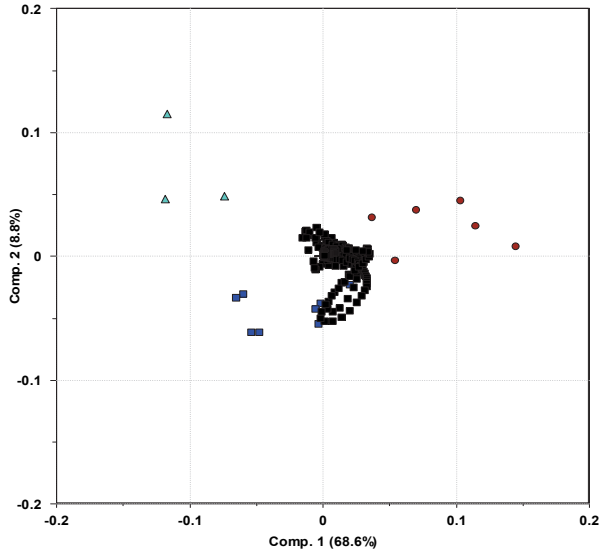


Figure C1.40: Biplot for the model for TAN based on FTIR data. Brown circles= biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates. The black squares are the variables, with FTIR data concentrated in the centre.

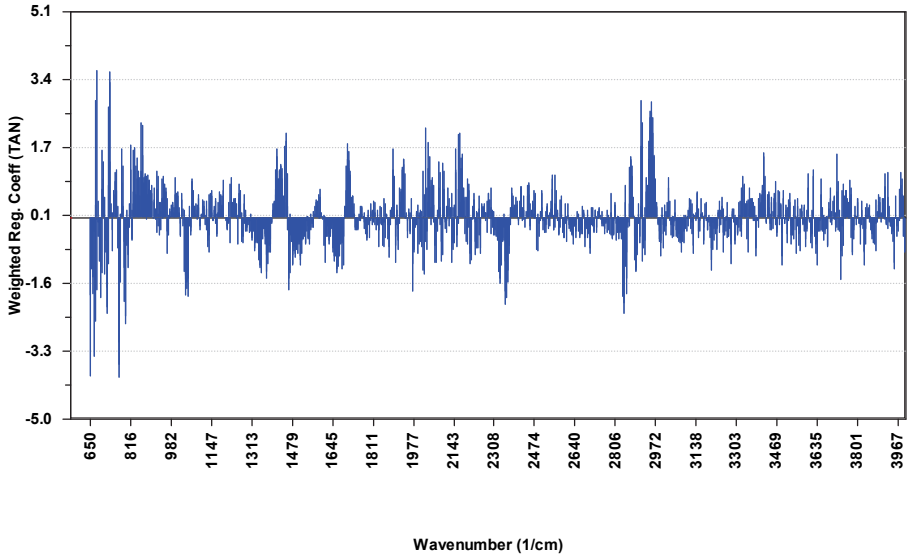


Figure C1.41: Weighted regression coefficients for the model for TAN based on FTIR data.

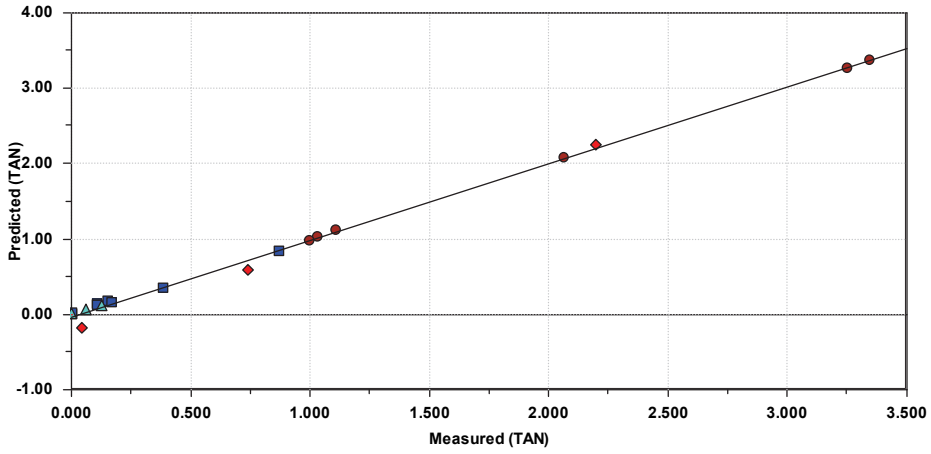


Figure C1.42: Predicted value plotted against measured value for the model for TAN based on FTIR data. R^2=0.997. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates, red diamonds=validation objects.

C1.1.8 Viscosity based on FTIR

Table C1.8: Explained variance and Cross Validation Standard Deviation for viscosity based on FTIR data.

Number of components	Explained variance	Cross Validation Standard Deviation
1	85.64%	0.45
2	88.04%	1.12
3	95.59%	0.83
4	96.68%	0.76

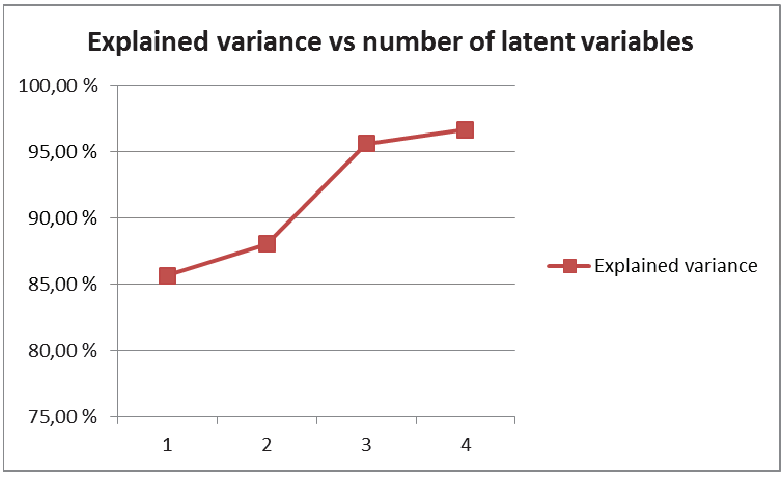


Figure C1.43: Explained variance plotted against the number of Latent Variables for the model for viscosity based on FTIR data.

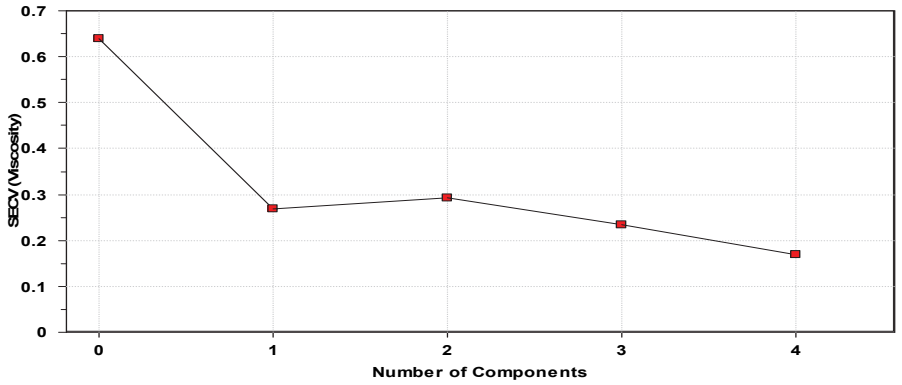


Figure C1.44: SECV plotted against number of components for the model for viscosity based on FTIR data.

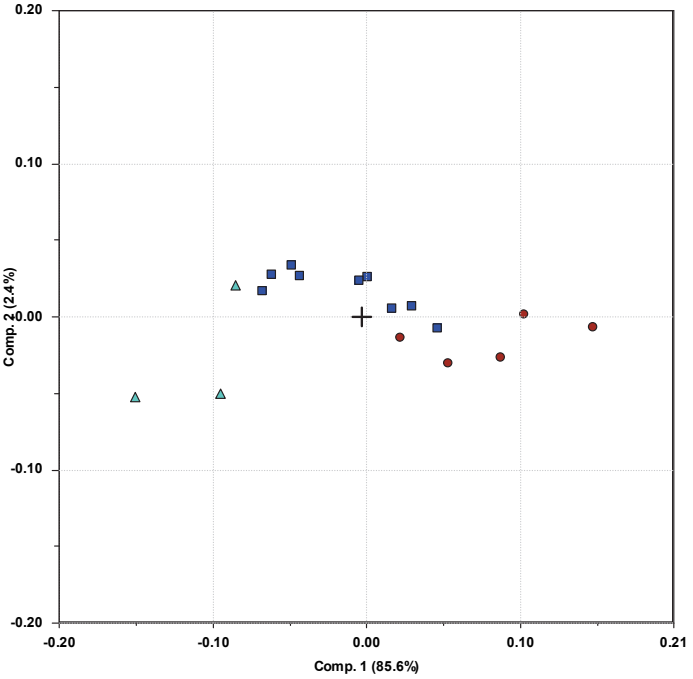


Figure C1.45: Scores for the model for viscosity based on FTIR data, Component 1 vs Component 2. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates.

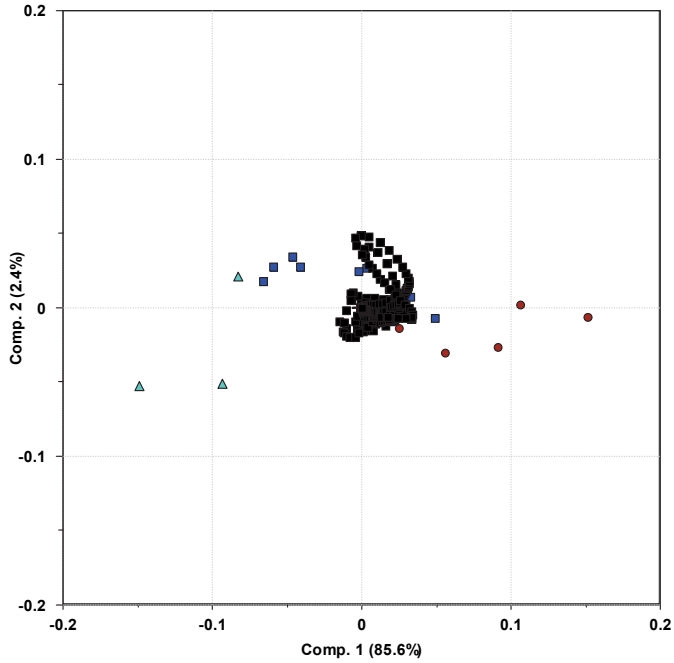


Figure C1.46: Biplot for the model for viscosity based on FTIR data. Brown circles= biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates. The black squares are the variables, with FTIR data concentrated in the centre.

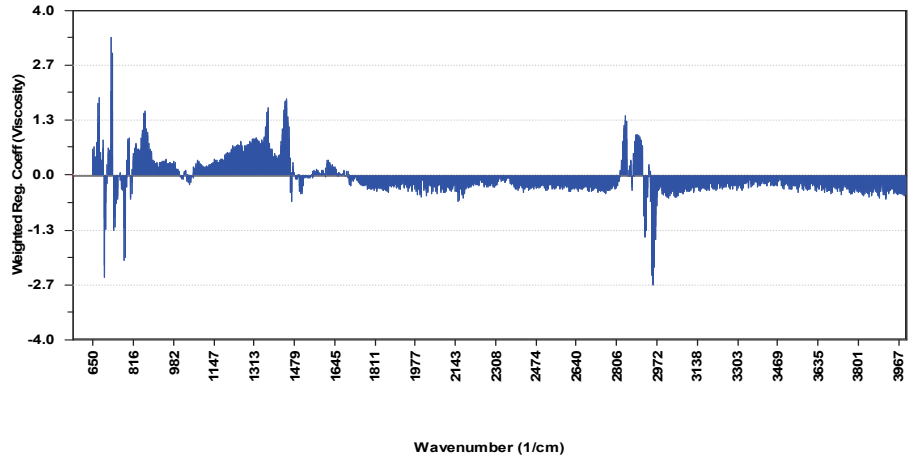


Figure C1.47: Weighted regression coefficients for the model for viscosity based on FTIR data.

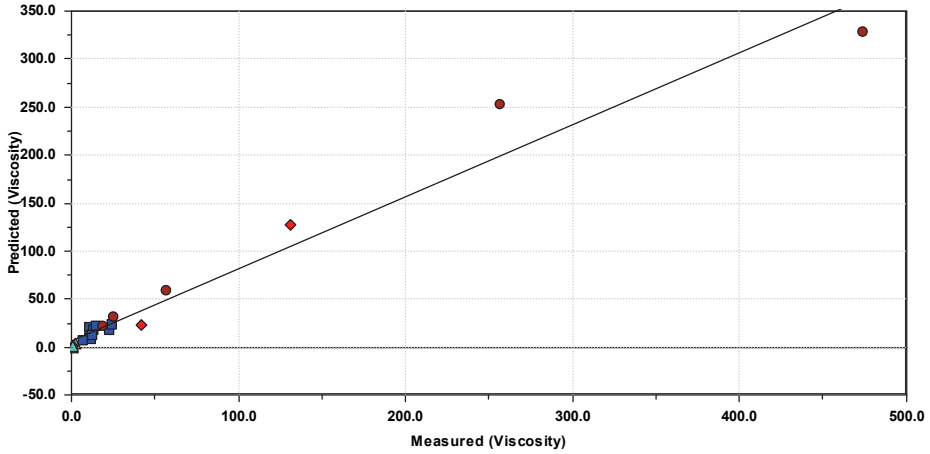


Figure C1.48: Predicted value plotted against measured value for the model for viscosity based on FTIR data. R^2=0.964. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates, red diamonds=validation objects.

C1.1.9 Viscosity based on FTIR

Table C1.9: Explained variance and Cross Validation Standard Deviation for asphaltene content based on FTIR data.

Number of components	Explained variance	Cross Validation Standard Deviation
1	88.73%	0.38
2	92.29%	1.06
3	96.12%	0.91
4	96.85%	0.96

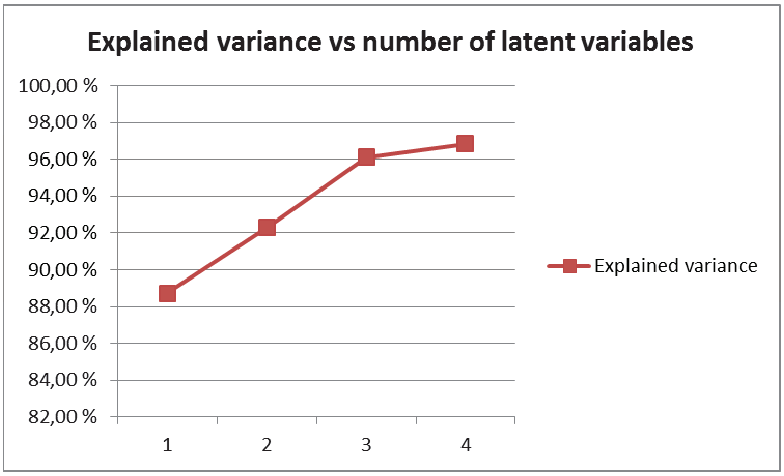


Figure C1.49: Explained variance plotted against the number of Latent Variables for the model for asphaltene content based on FTIR data.

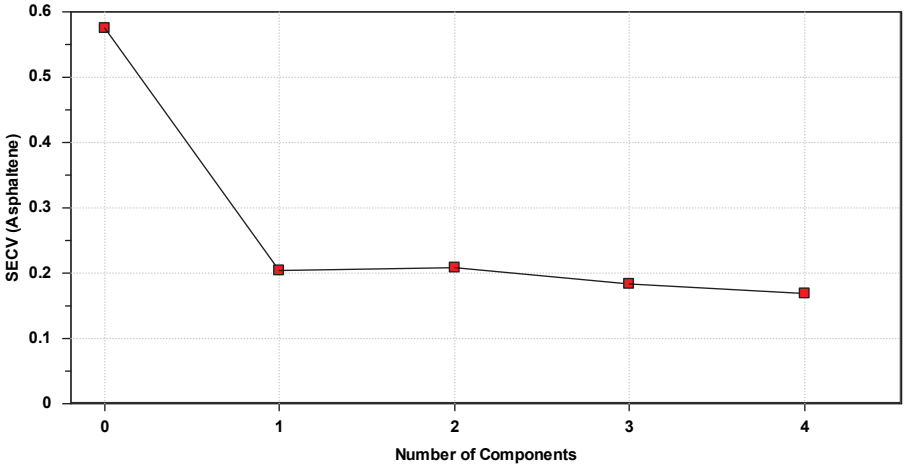


Figure C1.50: SECV plotted against number of components for the model for asphaltene content based on FTIR data.

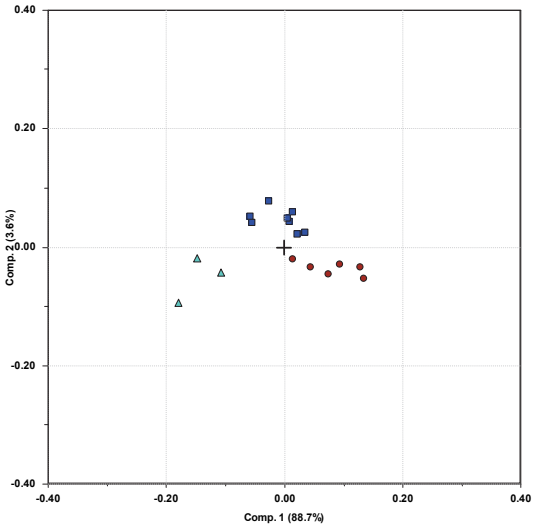


Figure C1.51: Scores for the model for asphaltene content based on FTIR data, Component 1 vs Component 2. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates.

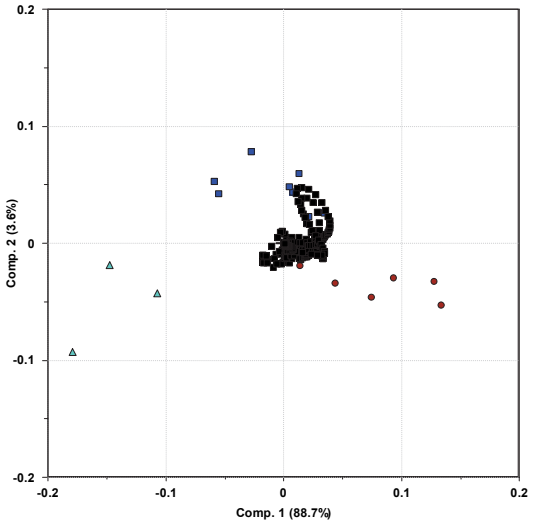


Figure C1.52: Biplot for the model for asphaltene content based on FTIR data. Brown circles= biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates. The black squares are the variables, with FTIR data concentrated in the centre.

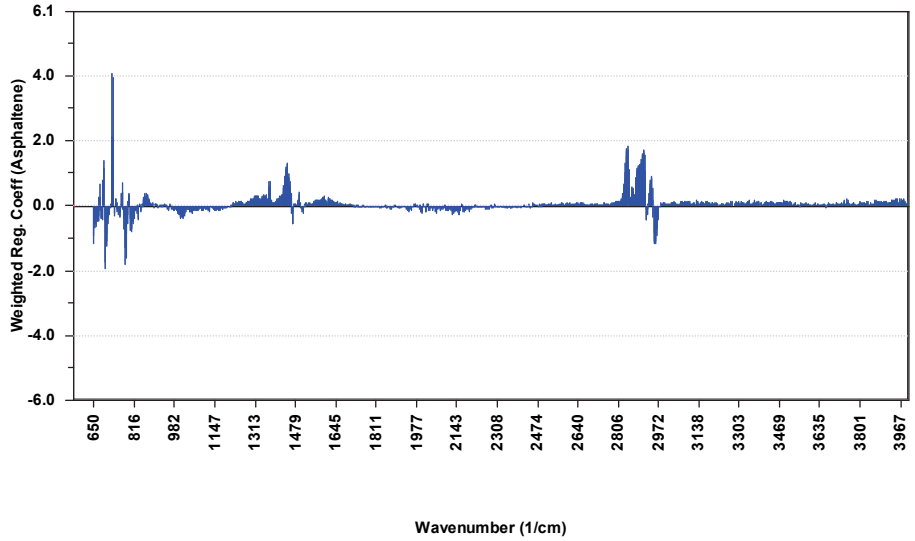


Figure C1.53: Weighted regression coefficients for the model for asphaltene content based on FTIR data.

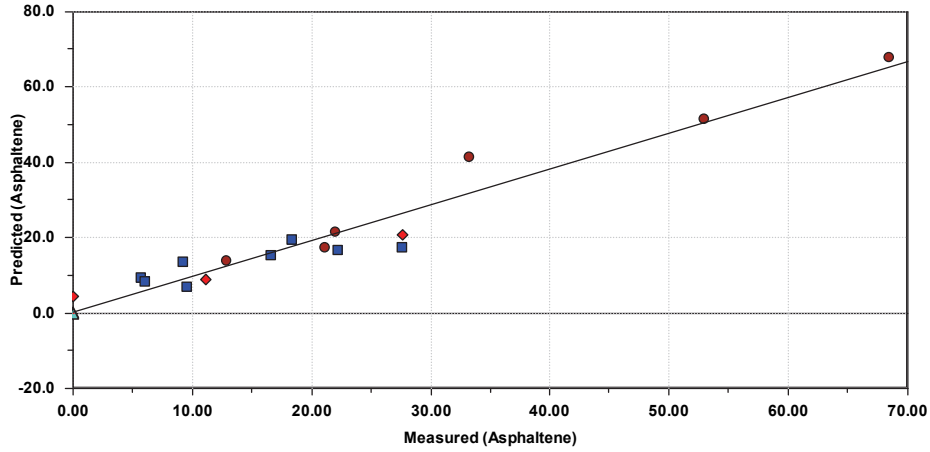


Figure C1.54: Predicted value plotted against measured value for the model for asphaltene content based on FTIR data. R^2=0.944. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates, red diamonds=validation objects.

C2.1 Models based on GC data

C2.1.1 Density based on GC

Table C2.1: Explained variance and Cross Validation Standard Deviation for the model for density based on GC data.

Number of components	Explained variance	Cross Validation Standard Deviation
1	72.20 %	0.64
2	81.24 %	1.10
3	88.39 %	1.05
4	93.35 %	0.94
5	95.64 %	1.03
6	97.66 %	1.02

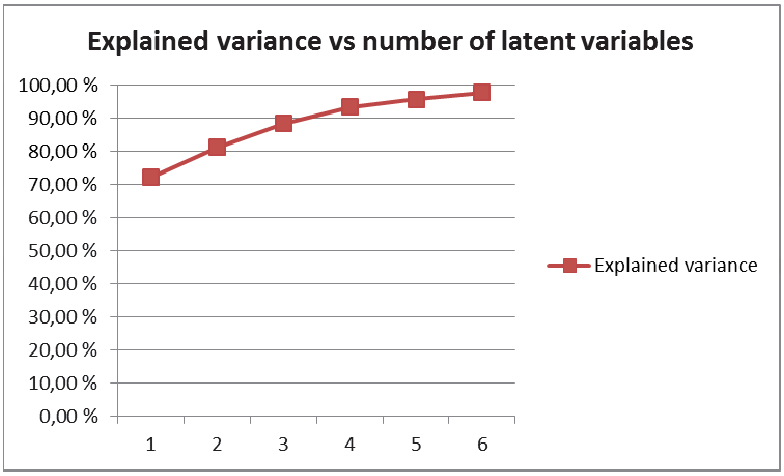


Figure C2.1: Explained variance plotted against the number of Latent Variables for the model for density based on GC data.

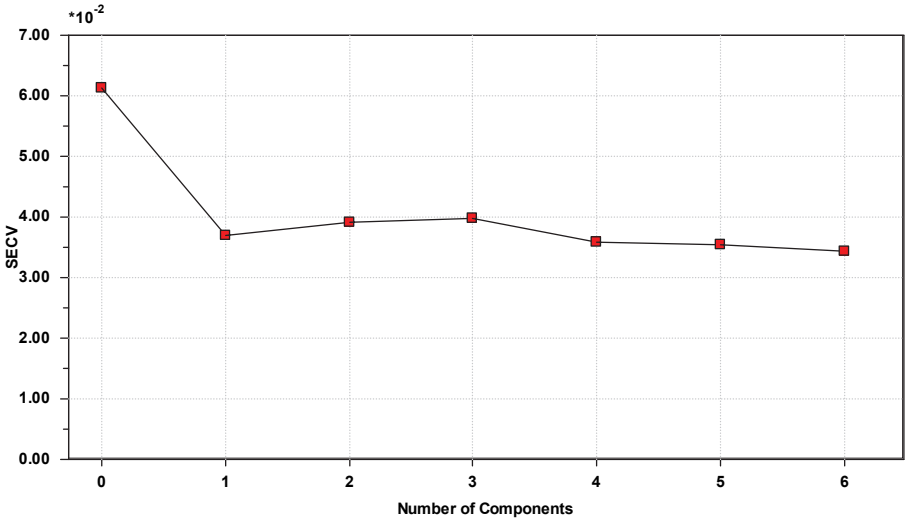


Figure C2.2: SECV plotted against number of components for the model for density based on GC data.

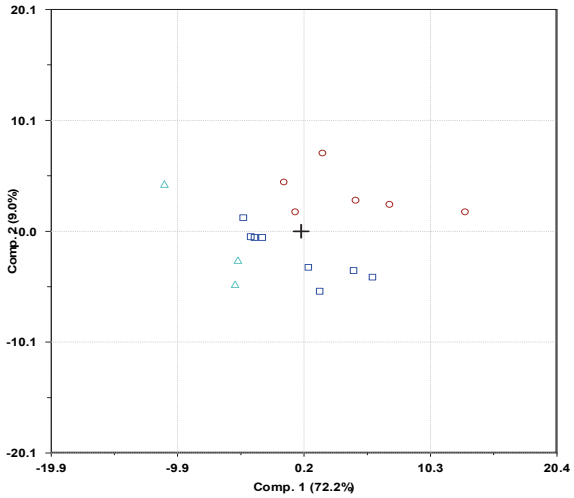


Figure C2.3: Scores for the model for density based on GC data, Component 1 vs Component 2. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates.

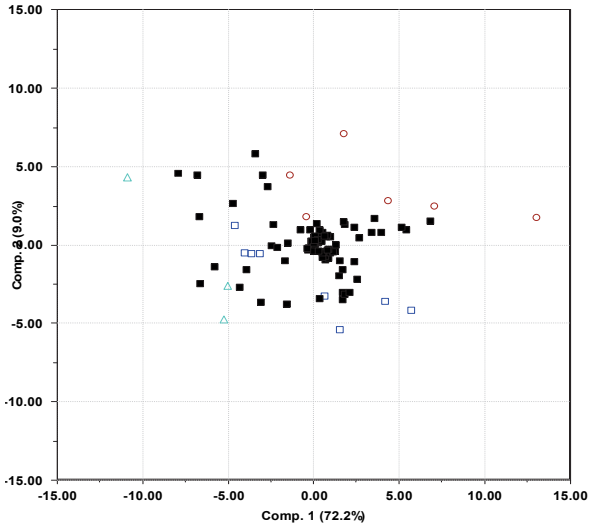


Figure C2.4: Biplot for the model for density based on GC data. Brown circles= biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates. The black squares are the variables, with GC data and objects concentrated in the centre while density is placed somewhat up and right from the centre.

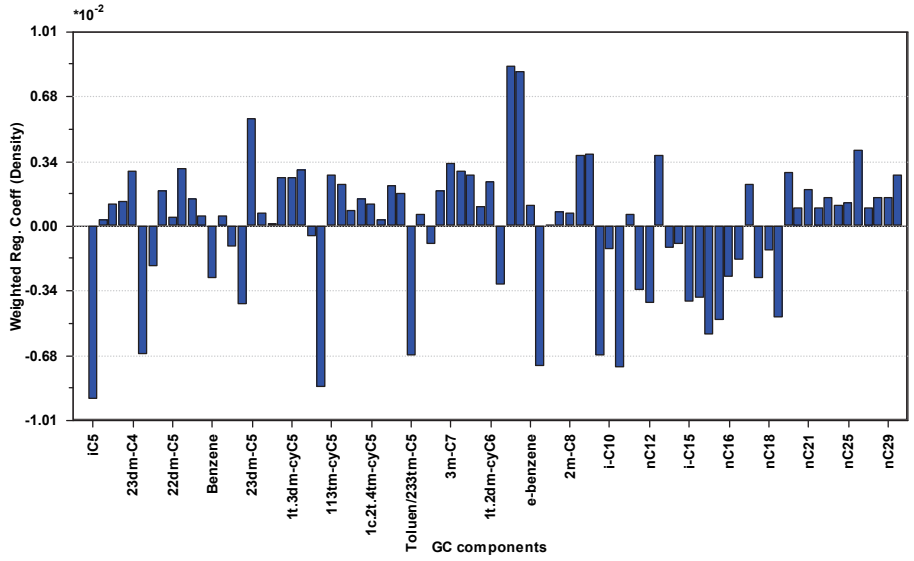


Figure C2.5: Weighted regression coefficients for the model for density based on GC data.

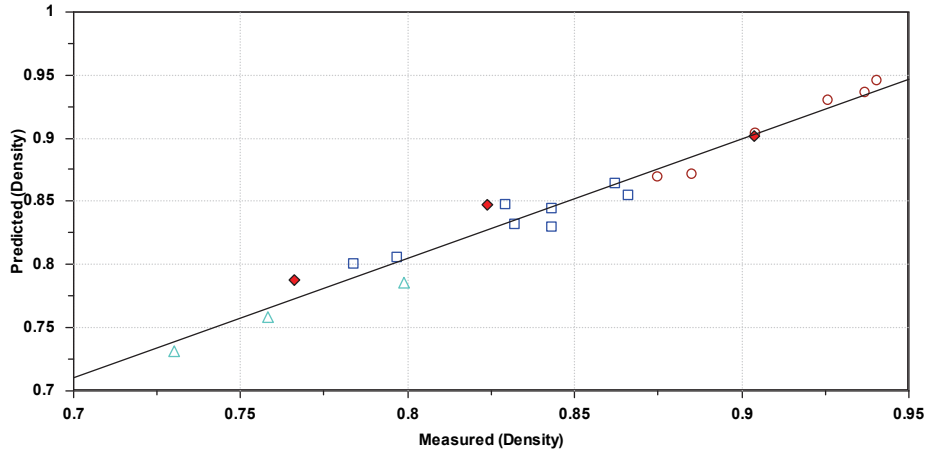


Figure C2.6: Predicted value plotted against measured value for the model density based on GC data. R^2=0.967. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates, red diamonds=validation objects.

C2.1.2 Velocity of sound based on GC

Table C2.2: Explained variance and Cross Validation Standard Deviation for the model for velocity of sound based on GC data.

Number of components	Explained variance	Cross Validation Standard Deviation
1	65.34%	0.74
2	73.02%	1.21
3	91.16%	0.94
4	95.95%	0.83
5	96.38%	0.80

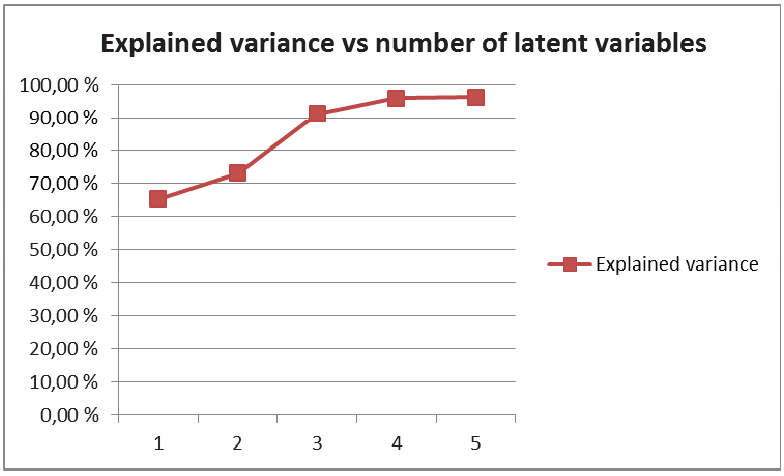


Figure C2.7: Explained variance plotted against the number of Latent Variables for the model for velocity of sound based on GC data.

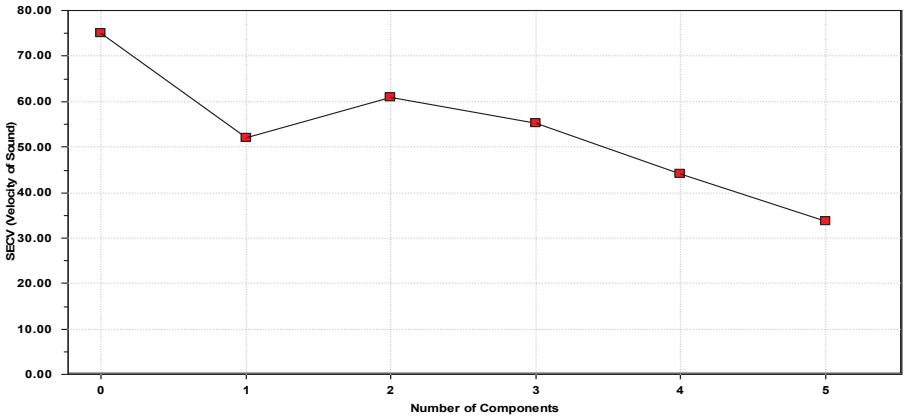


Figure C2.8: SECV plotted against number of components for the model for velocity of sound based on GC data.

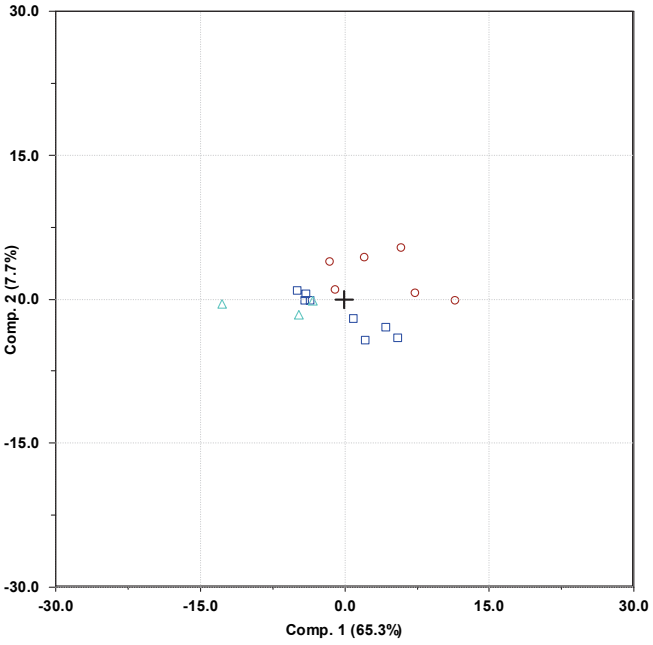


Figure C2.9: Scores for the model for velocity of sound based on GC data, Component 1 vs Component 2. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates.

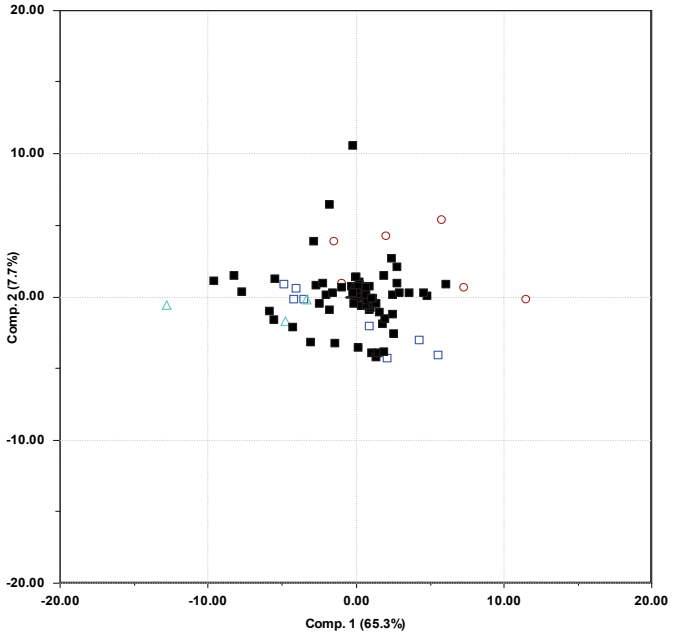


Figure C2.10: Biplot for the model for velocity of sound based on GC data. Brown circles= biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates. The black squares are the variables.

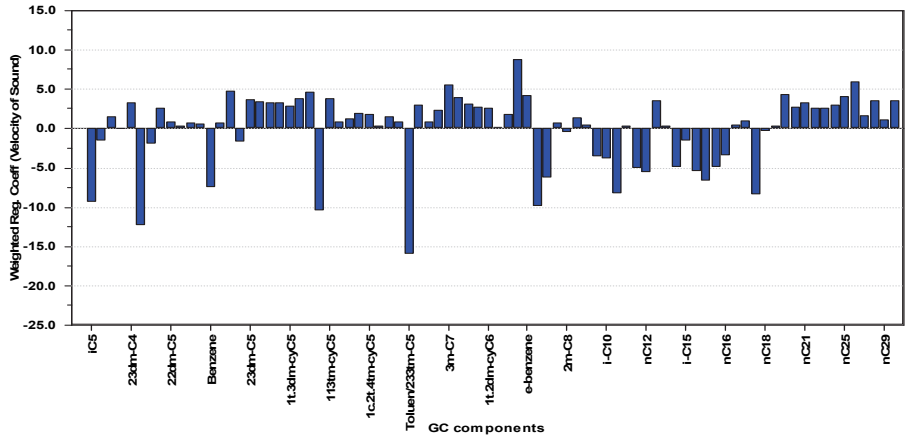


Figure C2.11: Weighted regression coefficients for the model for velocity of sound based on GC data.

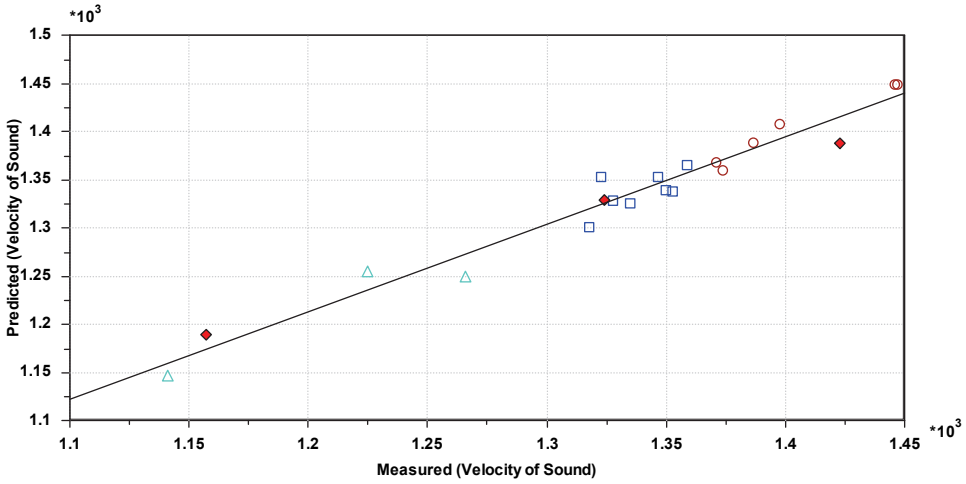


Figure C2.12: Predicted value plotted against measured value for the model for velocity of sound based on GC data. R^2=0.960. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates, red diamonds=validation objects.

C2.1.3 Static permittivity based on GC, asphaltene content and TAN

Table C2.3: Explained variance and Cross Validation Standard Deviation for the model for the permittivity variable static permittivity based on GC data, asphaltene content and TAN.

Number of components	Explained variance	Cross Validation Standard Deviation
1	82.41%	0.53
2	92.86%	0.79
3	94.87%	0.95
4	95.76%	1.10
5	98.35%	1.02
6	98.65%	1.04
7	98.89%	1.04
8	99.32%	1.00

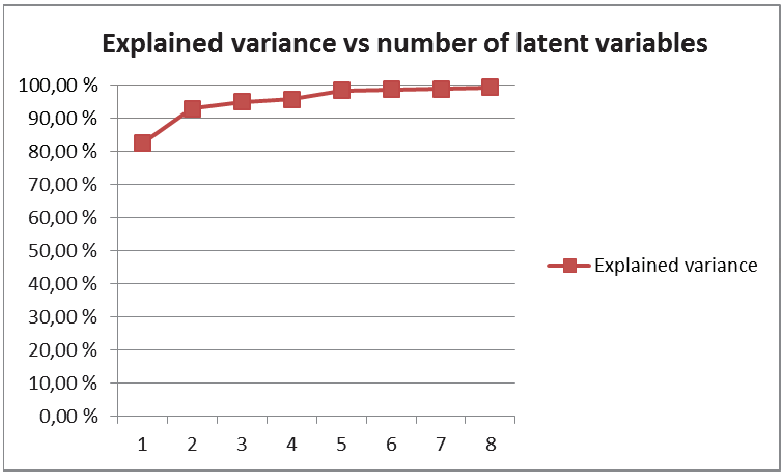


Figure C2.13: Explained variance plotted against the number of Latent Variables for the model for the permittivity variable static permittivity based on GC data, asphaltene content and TAN.

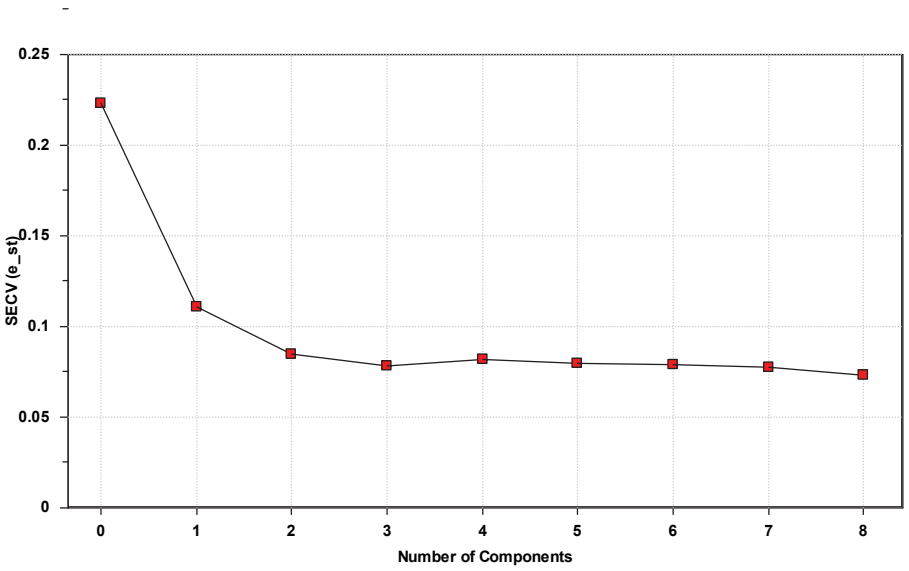


Figure C2.14: SECV plotted against number of components for the model for the permittivity variable static permittivity based on GC data, asphaltene content and TAN.

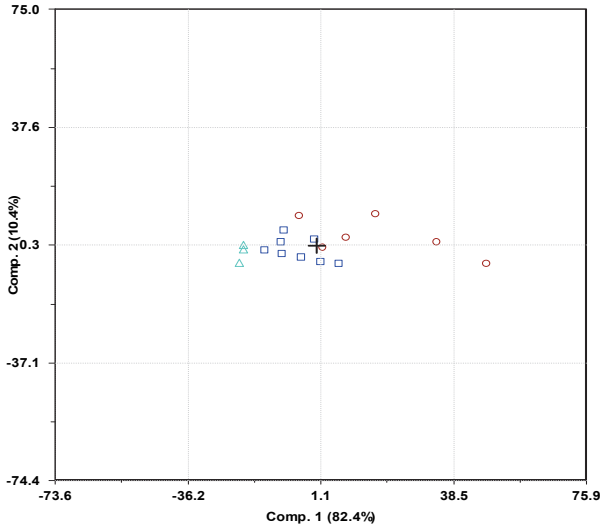


Figure C2.15: Scores for the model for the permittivity variable static permittivity based on GC data, asphaltene content and TAN, Component 1 vs Component 2. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates.

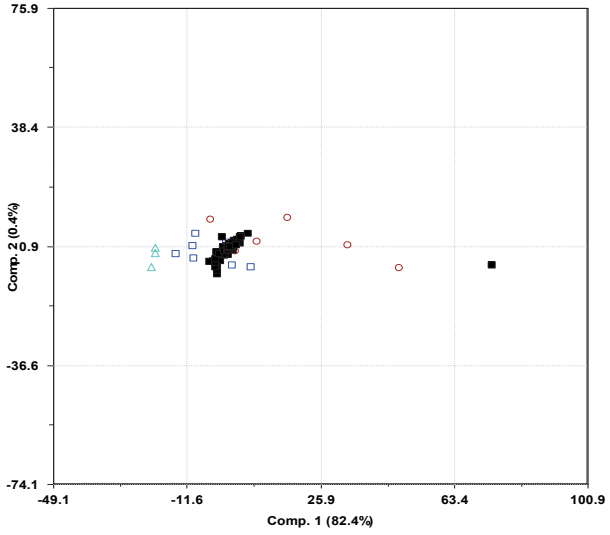


Figure C2.16: Biplot for the model for the permittivity variable static permittivity based on GC data, asphaltene content and TAN. Brown circles= biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates. The black squares are the variables, with the GC variables grouped in the middle together with the TAN variable, and the asphaltene variable placed on the far right.

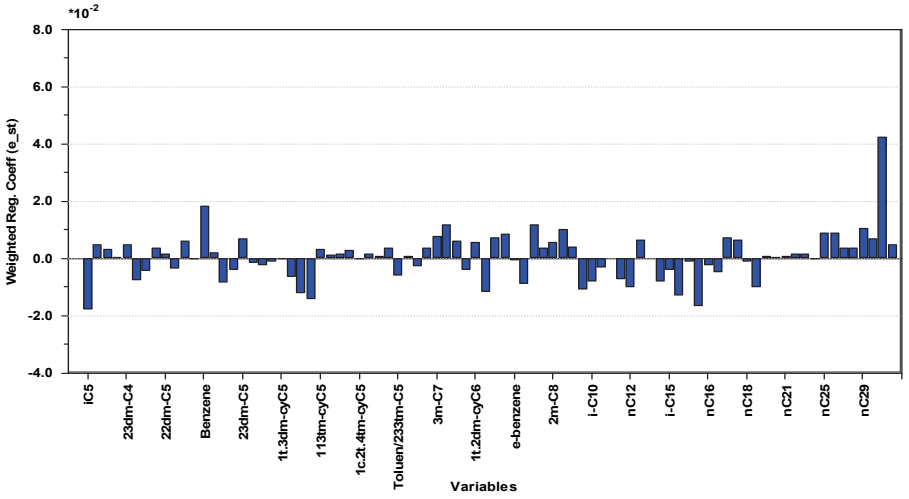


Figure C2.17: Weighted regression coefficients for the model for permittivity variable static permittivity based on GC data, asphaltene content and TAN. TAN and asphaltene on the far left, with TAN having a large positive effect on the model.

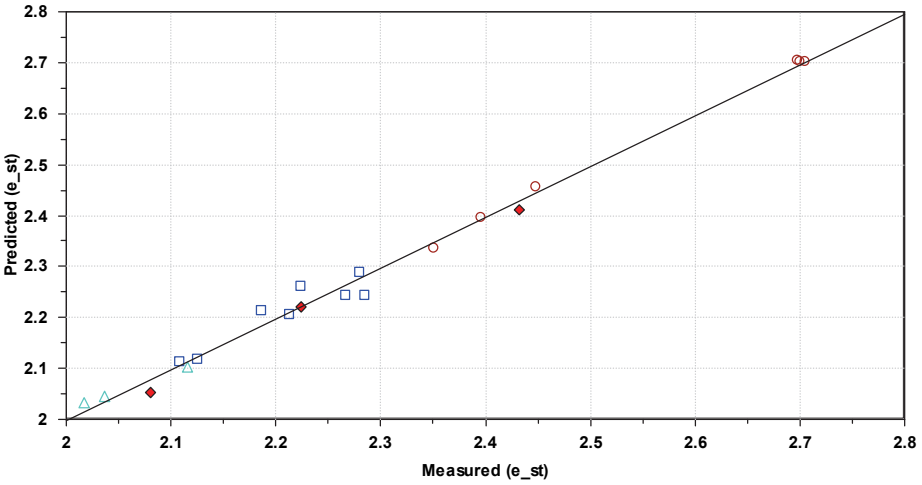


Figure C2.18: Predicted value plotted against measured value for the model for the permittivity variable static permittivity based on GC data, asphaltene content and TAN. R^2=0.993. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates, red diamonds=validation objects.

C2.1.4 High frequency permittivity based on GC, asphaltene content and TAN

Table C2.4: Explained variance and Cross Validation Standard Deviation for the model for the permittivity variable high frequency permittivity based on GC data, asphaltene content and TAN.

Number of components	Explained variance	Cross Validation Standard Deviation
1	74.08%	0.64
2	89.92%	0.76
3	91.53%	0.96
4	96.33%	1.06
5	97.19%	0.95
6	97.85%	1.30
7	98.31%	1.01
8	99.08%	1.01
9	99.36%	1.32

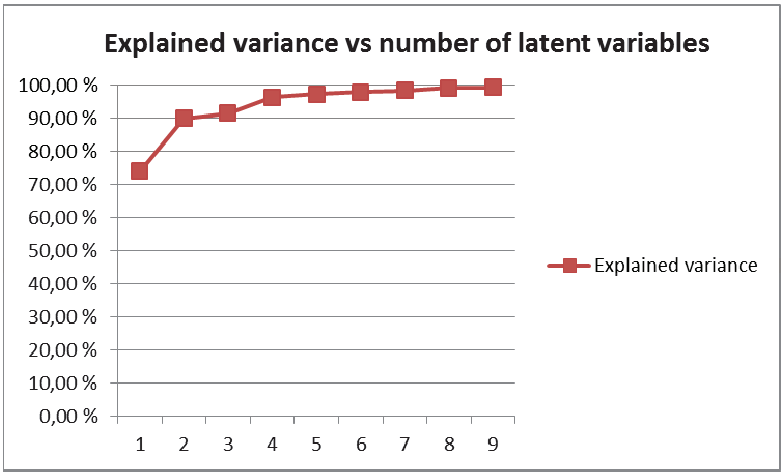


Figure C2.19: Explained variance plotted against the number of Latent Variables for the model for the permittivity variable high frequency permittivity based on GC data, asphaltene content and TAN.

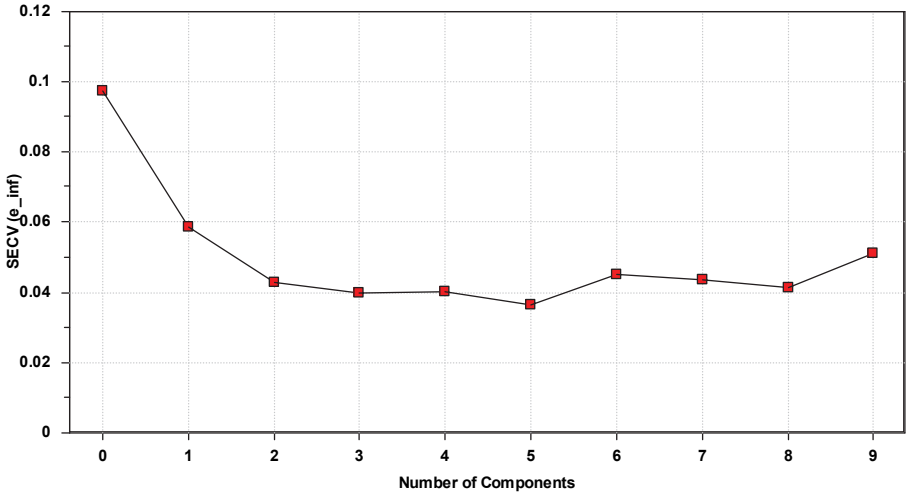


Figure C2.20: SECV plotted against number of components for the model for the permittivity variable high frequency permittivity based on GC data, asphaltene content and TAN.

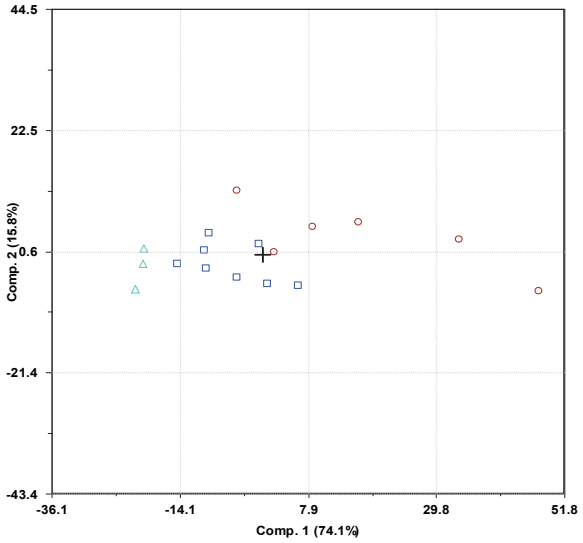


Figure C2.21: Scores for the model for the permittivity variable high frequency permittivity based on GC data, asphaltene content and TAN, Component 1 vs Component 2. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates.

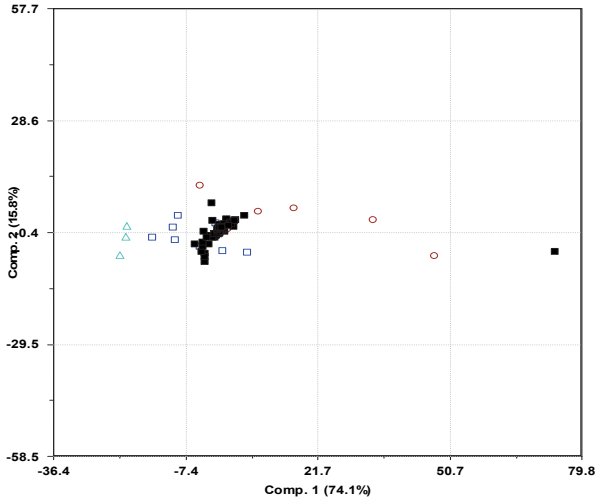


Figure C2.22: Biplot for the model for the permittivity variable high frequency permittivity based on GC data, asphaltene content and TAN. Brown circles= biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates. The black squares are the variables, with the GC variables grouped in the middle together with the TAN variable, and the asphaltene variable placed on the far right.

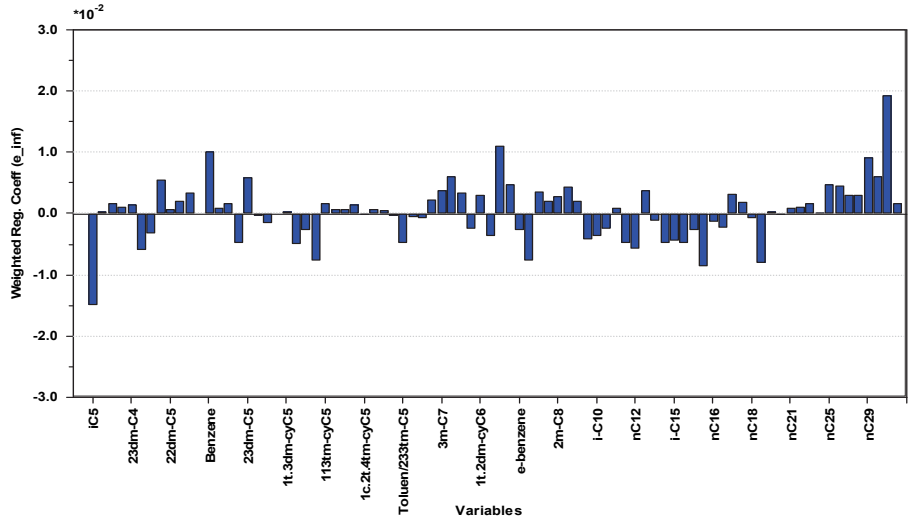


Figure C2.23: Weighted regression coefficients for the model for permittivity variable high frequency permittivity based on GC data, asphaltene content and TAN. TAN and asphaltene on the far right, with TAN having a large positive effect on the model.

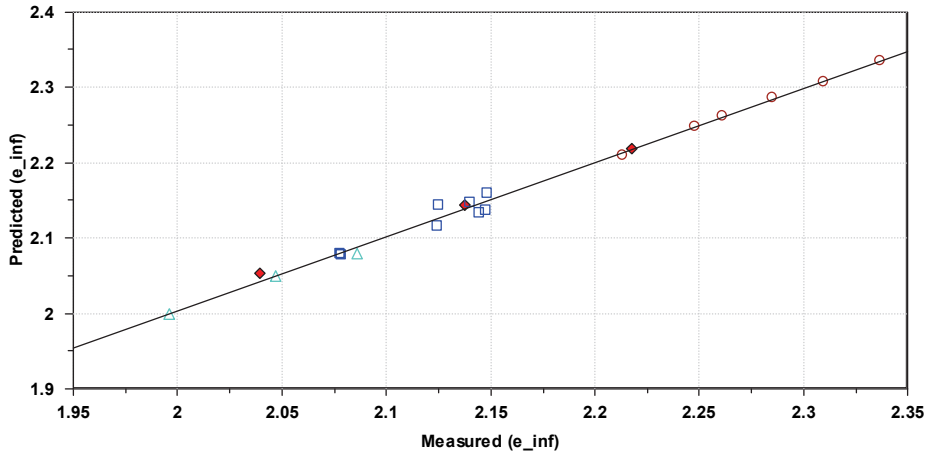


Figure C2.24: Predicted value plotted against measured value for the model for the permittivity variable high frequency permittivity based on GC data, asphaltene content and TAN. R^2=0.993. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates, red diamonds=validation objects.

C2.1.5 α based on GC

Table C2.5: Explained variance and Cross Validation Standard Deviation for the model for the permittivity variable α based on GC data.

Number of components	Explained variance	Cross Validation Standard Deviation
1	37.12%	1.14
2	59.32%	1.01
3	63.45%	0.91
4	69.13%	1.14
5	75.58%	1.15
6	83.24%	0.98

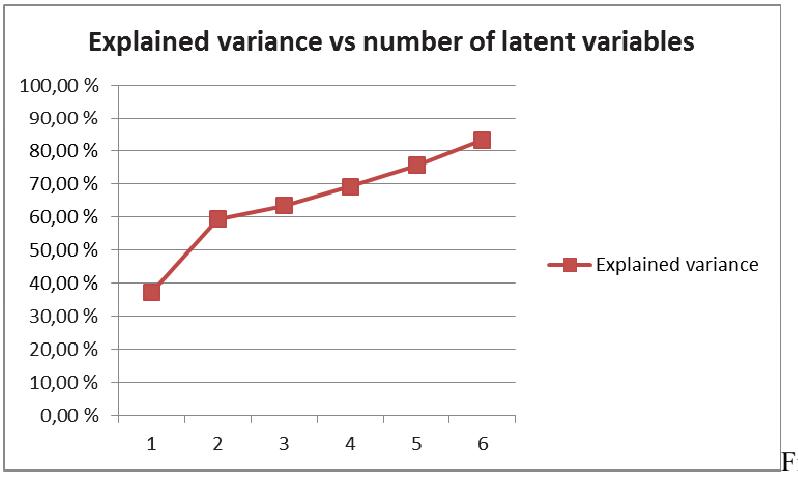


Figure C2.25: Explained variance plotted against the number of Latent Variables for the model for the permittivity variable α based on GC data.

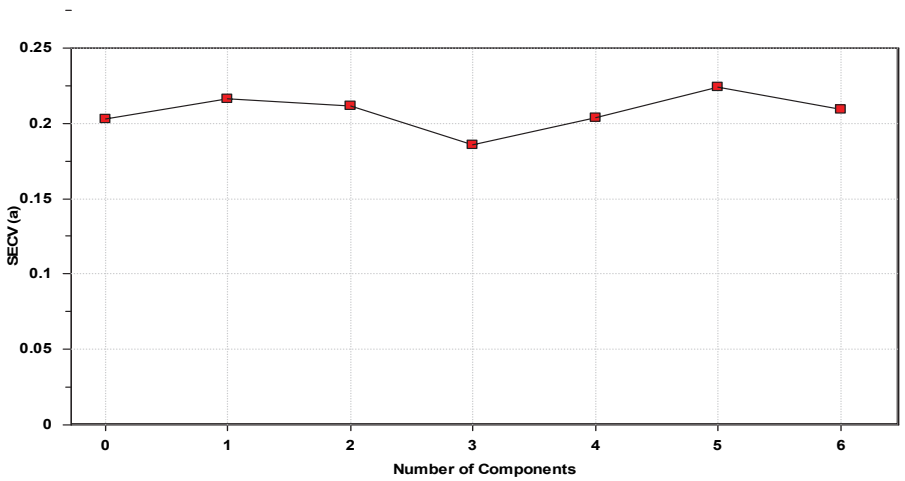


Figure C2.26: SECV plotted against number of components for the model for the permittivity variable α based on GC data.

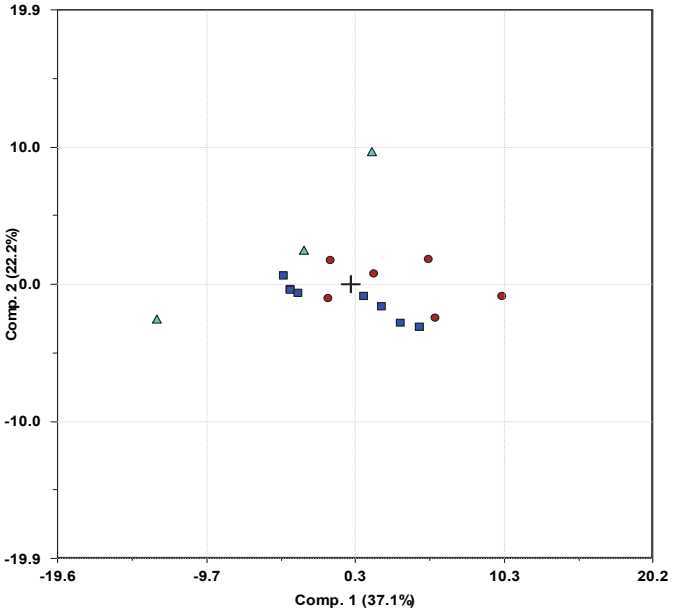


Figure C2.27: Scores for the model for the permittivity variable α based on GC data. Component 1 vs Component 2. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates.

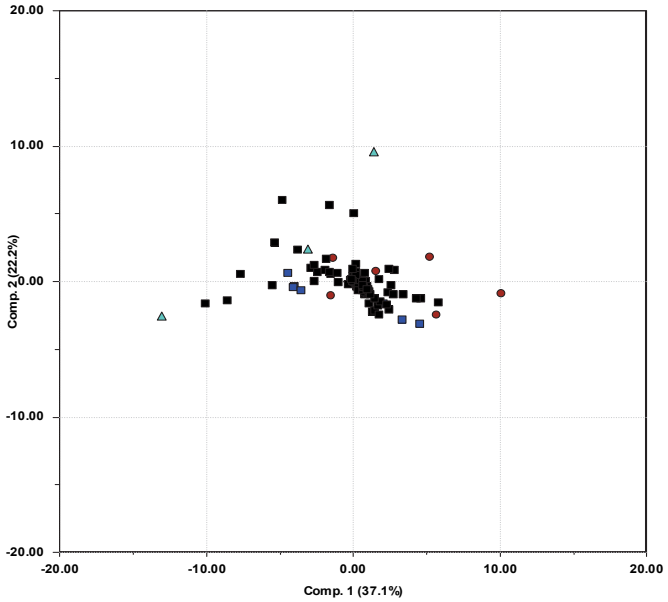


Figure C2.28: Biplot for the model for the permittivity variable α based on GC data. Brown circles= biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates. The black squares are the variables, with the GC variables grouped in the middle.

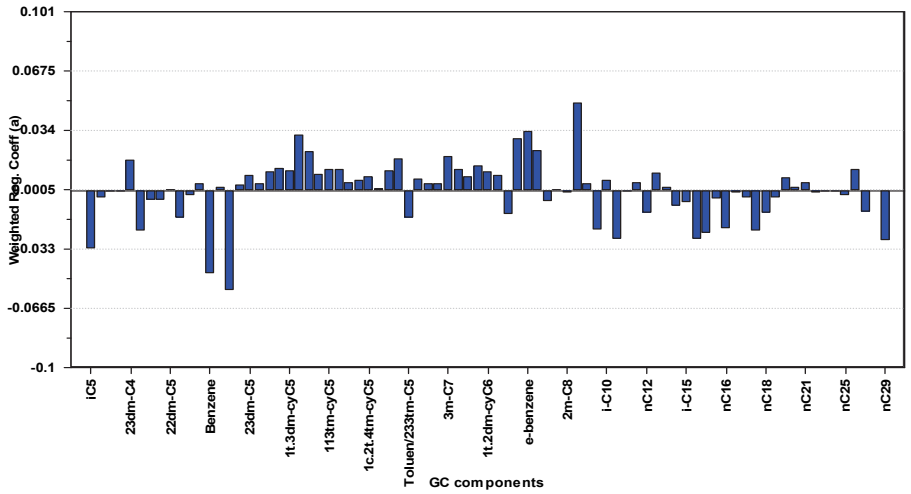


Figure C2.29: Weighted regression coefficients for the model for permittivity variable α based on GC data.

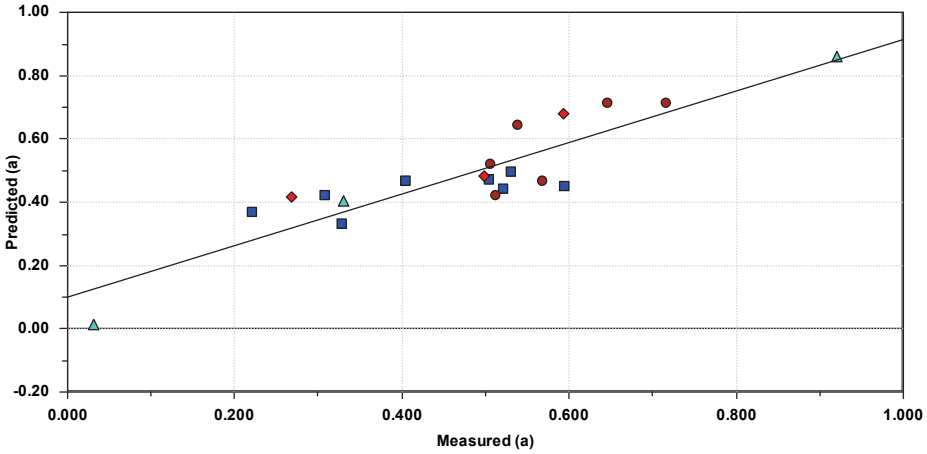


Figure C2.30: Predicted value plotted against measured value for the model for the permittivity variable α based on GC data. R^2=0.808. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates, red diamonds=validation objects.

C2.1.6 τ based on GC

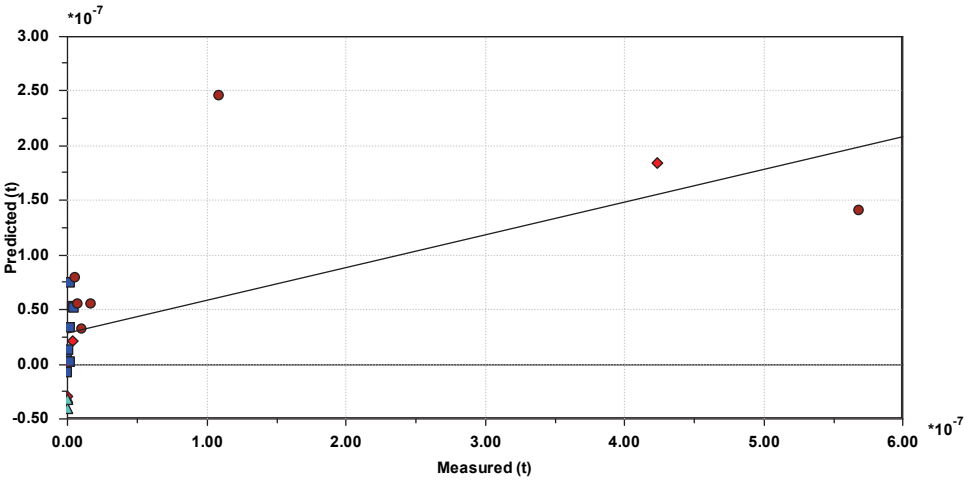


Figure C2.31: Predicted value plotted against measured value for the model for the permittivity variable τ based on GC data, TAN and asphaltene content. R^2=0.393. 1 Latent variable, explained variance=25.67%, Cross Validation Standard Deviation=0.96. This model could be used for separating between high and low value, the low value objects could then be applied to the second model in this section. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates, red diamonds=validation objects.

Table C2.6: Explained variance and Cross Validation Standard Deviation for the low value model for the permittivity variable τ based on GC data, TAN and asphaltene content.

Number of components	Explained variance	Cross Validation Standard Deviation
1	44.57%	0.85
2	59.28%	1.17
3	81.05%	1.13
4	88.85%	0.92
5	90.98%	0.87
6	93.58%	1.39
7	99.64%	1.10

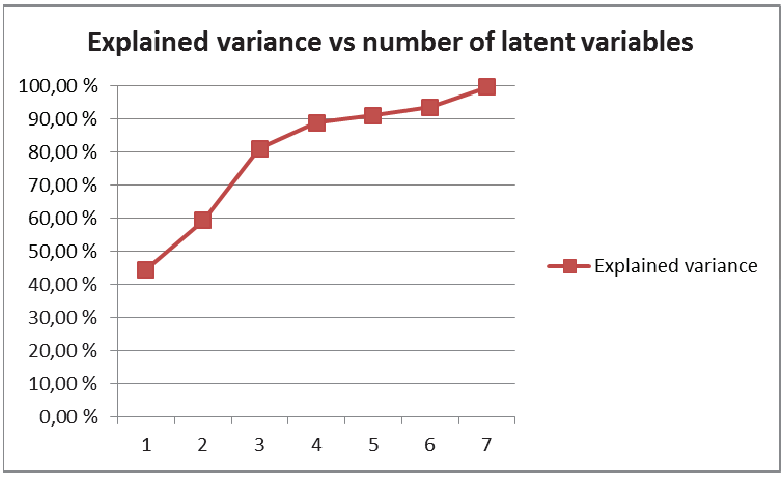


Figure C2.32: Explained variance plotted against the number of Latent Variables for the low value model for the permittivity variable τ based on GC data, TAN and asphaltene content.

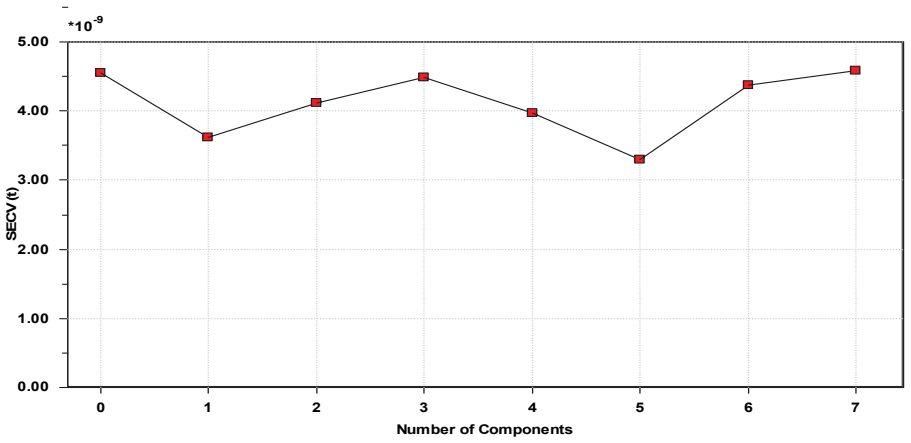


Figure C2.33: SECV plotted against number of components for the low value model for the permittivity variable τ based on GC data, TAN and asphaltene content.

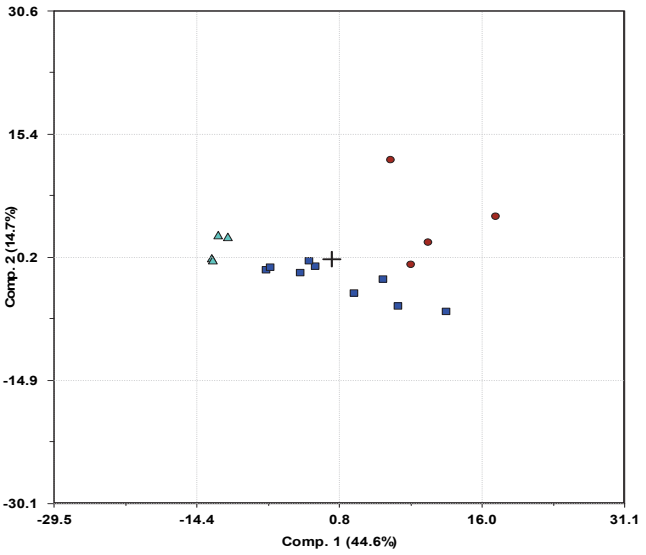


Figure C2.34: Scores for the low value model for the permittivity variable τ based on GC data, TAN and asphaltene content. The high value objects are omitted from this model, in the score plot they would be placed in the upper right quadrant, showing clear outlier character. Component 1 vs Component 2. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates.

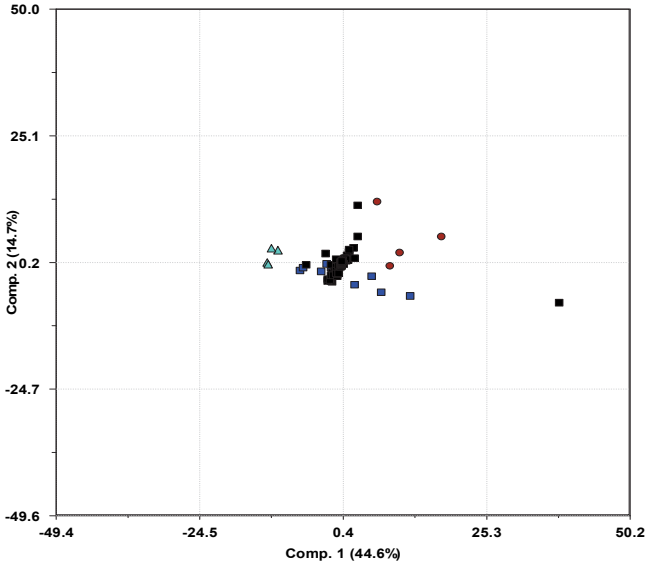


Figure C2.35: Biplot for the low value model for the permittivity variable τ based on GC data, TAN and asphaltene content. Brown circles= biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates. The black squares are the variables, with the GC variables and TAN grouped in the middle and asphaltene content on the far right.

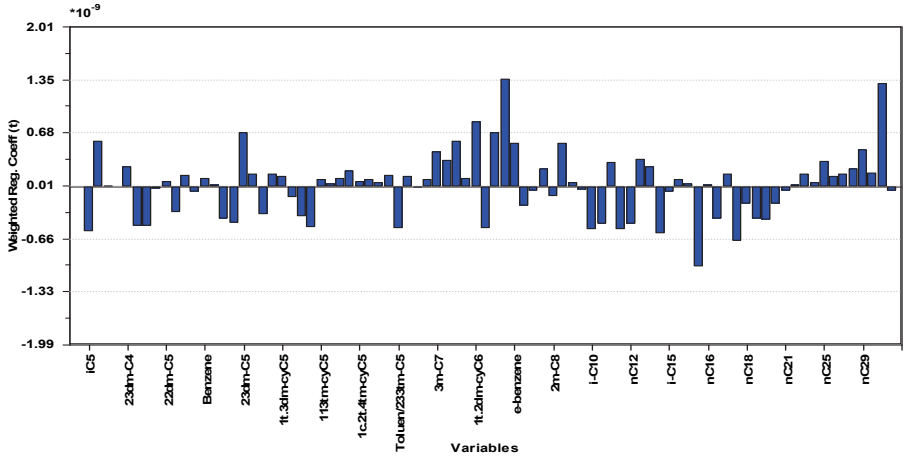


Figure C2.36: Weighted regression coefficients for the low value model for the permittivity variable τ based on GC data, TAN and asphaltene content. TAN and asphaltene on the far right, with TAN having a large positive effect on the model.

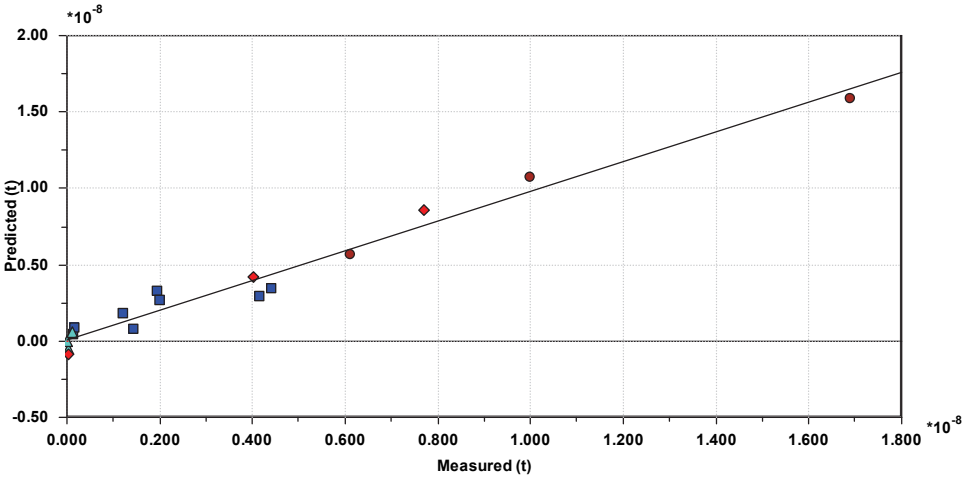


Figure C2.37: Predicted value plotted against measured value for the low value model for the permittivity variable τ based on GC data, TAN and asphaltene content. R^2=0.969. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates, red diamonds=validation objects.

C2.1.7 TAN based on GC and asphaltene content.

Table C2.7: Explained variance and Cross Validation Standard Deviation for the model for TAN based on GC data and asphaltene content.

Number of components	Explained variance	Cross Validation Standard Deviation
1	27.92%	0.96
2	45.02%	1.07
3	58.12%	1.41
4	67.33%	1.16
5	81.61%	0.77
6	86.51%	0.83
7	90.55%	0.96
8	95.10%	1.39
9	95.92%	1.30
10	97.55%	1.26
11	98.94%	1.12

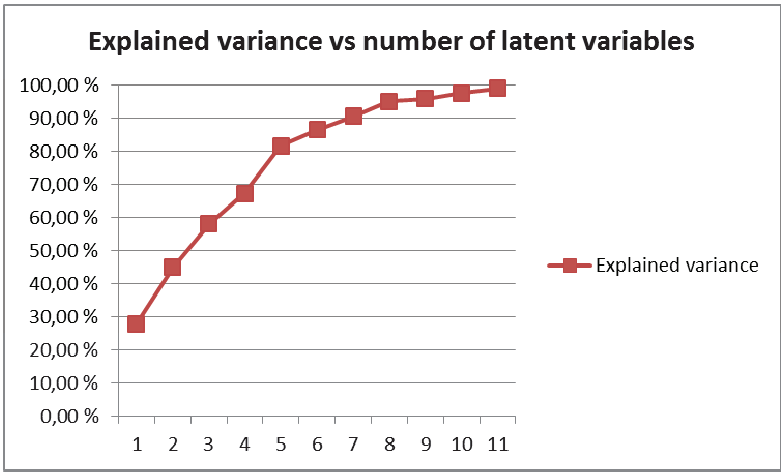


Figure C2.38: Explained variance plotted against the number of Latent Variables for the model for TAN based on GC data and asphaltene content.

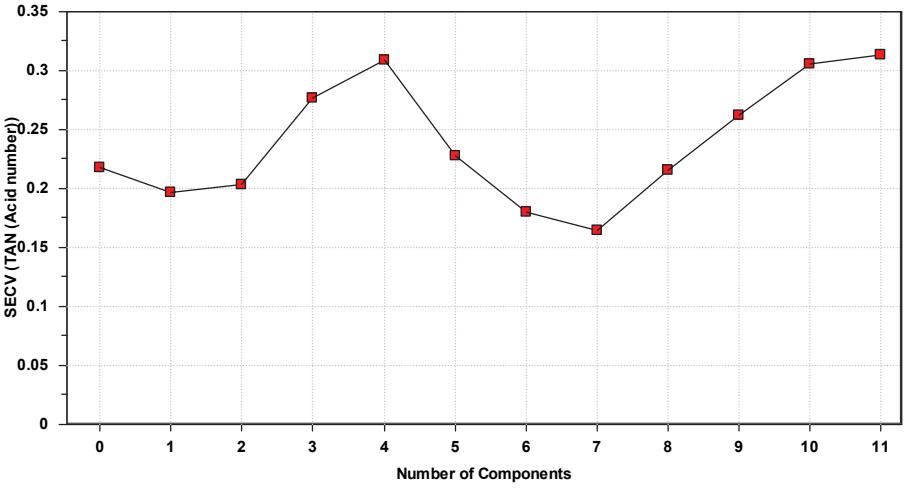


Figure C2.39: SECV plotted against number of components for the model for TAN based on GC data and asphaltene content.

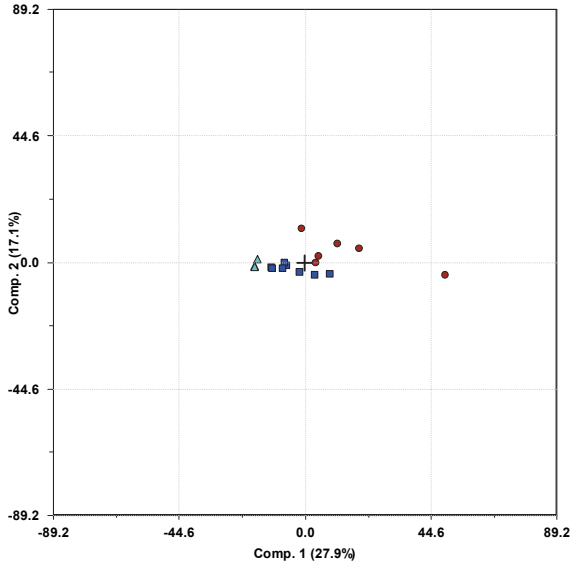


Figure C2.40: Scores for the model for TAN based on GC data and asphaltene content. Component 1 vs Component 2. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates.

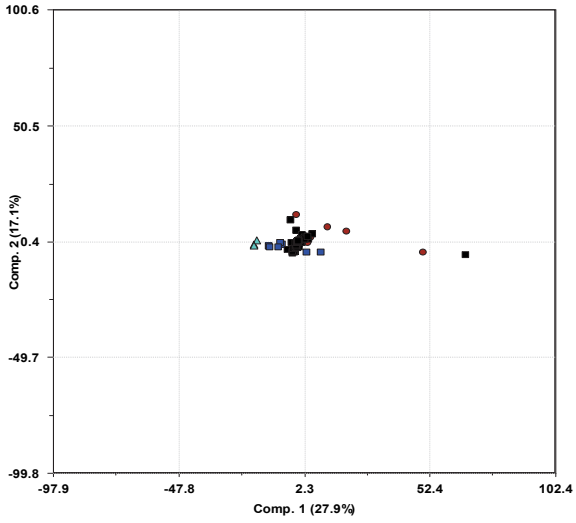


Figure C2.41: Biplot for the model TAN based on GC data and asphaltene content. Brown circles= biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates. The black squares are the variables, with the GC variables grouped in the middle and asphaltene content on the far right.

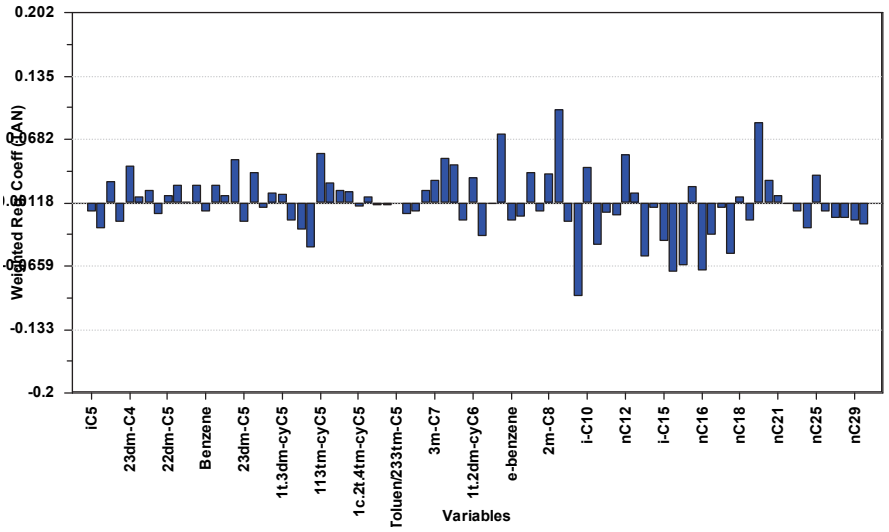


Figure C2.42: Weighted regression coefficients for the model for TAN based on GC data and asphaltene content. Asphaltene content on the far right, having small to little effect on the model.

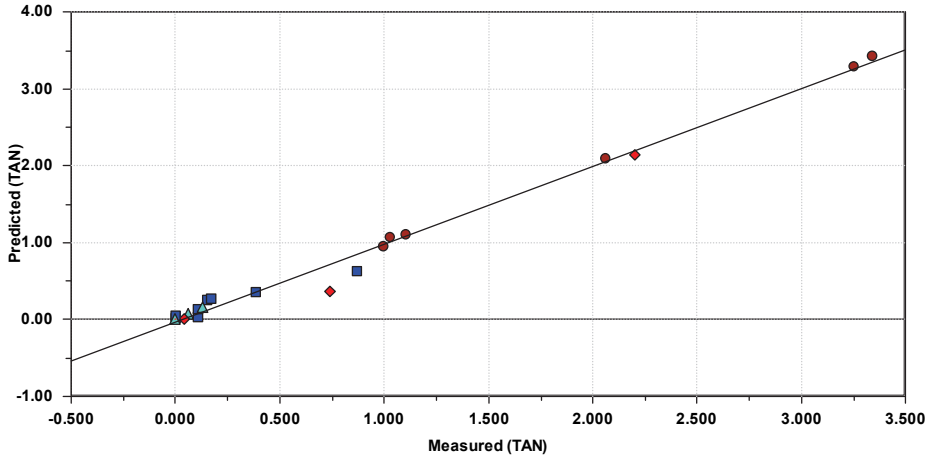


Figure C2.43: Predicted value plotted against measured value for the model for TAN based on GC data and asphaltene content. R^2=0.989. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates, red diamonds=validation objects.

C2.1.8 Viscosity based on GC, TAN and asphaltene content

Table C2.8: Explained variance and Cross Validation Standard Deviation for the model for viscosity based on GC data, TAN and asphaltene content.

Number of components	Explained variance	Cross Validation Standard Deviation
1	85.81%	0.43
2	90.62%	0.93
3	92.16%	1.08
4	93.41%	1.18
5	95.64%	1.07
6	96.66%	1.19
7	97.52%	1.12
8	98.85%	1.10

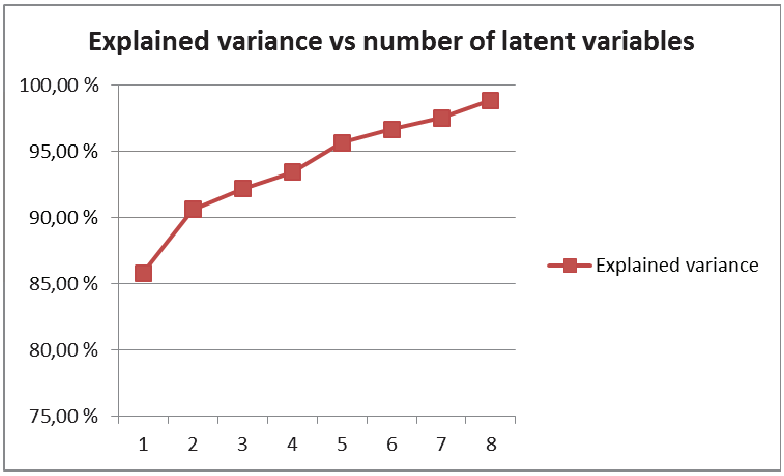


Figure C.2.44: Explained variance plotted against the number of Latent Variables for the model for viscosity based on GC data, TAN and asphaltene content.

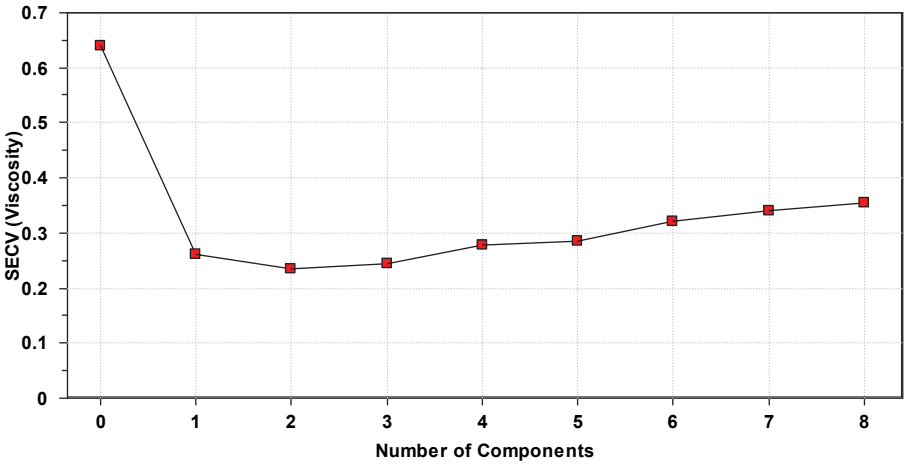


Figure C2.45: SECV plotted against number of components for the model for viscosity based on GC data, TAN and asphaltene content.

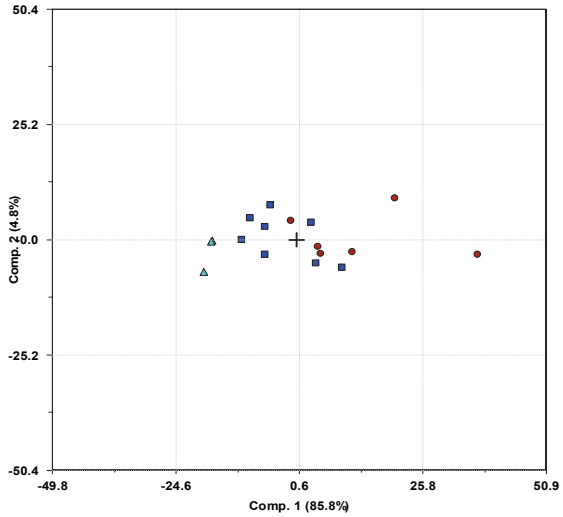


Figure C2.46: Scores for the model for viscosity based on GC data, TAN and asphaltene content. Component 1 vs Component 2. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates.

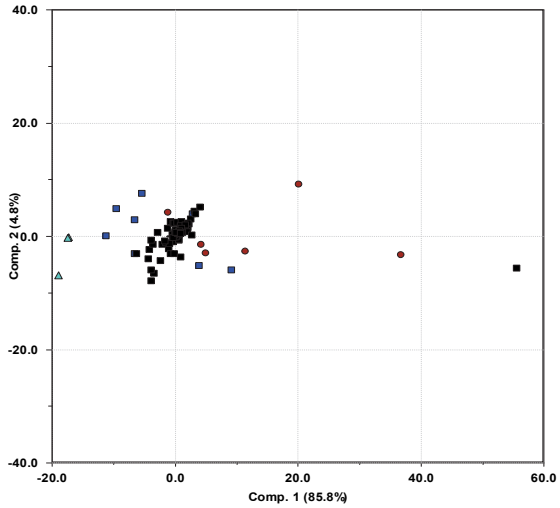


Figure C2.47: Biplot for the model viscosity based on GC data, TAN and asphaltene content. Brown circles= biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates. The black squares are the variables, with the GC variables and TAN grouped in the middle and asphaltene content on the far right.

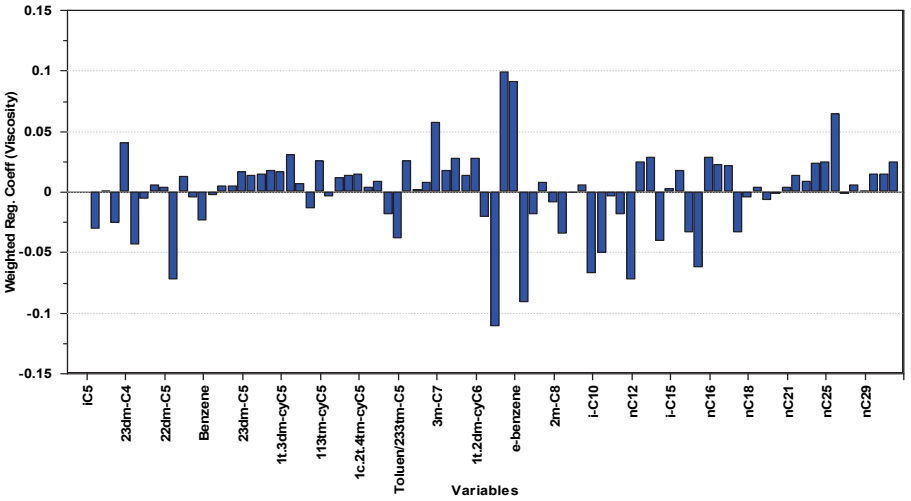


Figure C2.48: Weighted regression coefficients for the model for viscosity based on GC data, TAN and asphaltene content. TAN and asphaltene content on the far right, having some positive effect on the model.

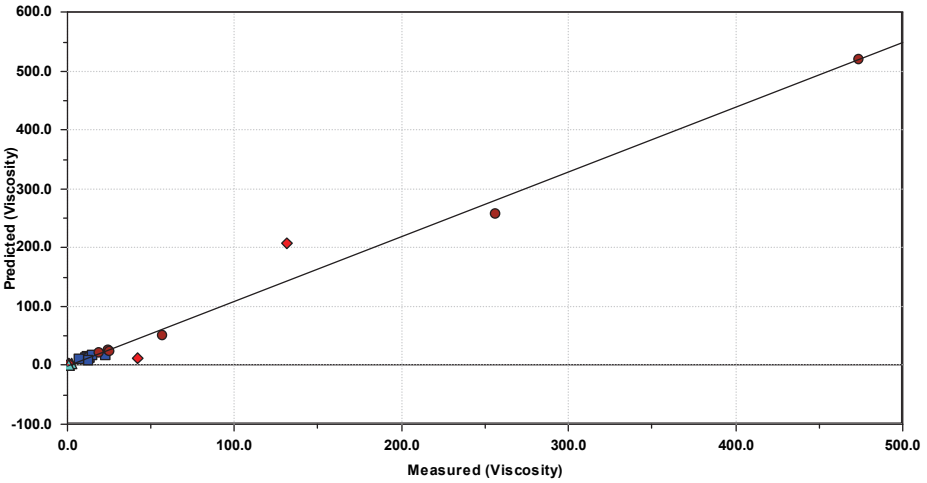


Figure C2.49: Predicted value plotted against measured value for the model for viscosity based on GC data, TAN and asphaltene content. R^2=0.981. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates, red diamonds=validation objects.

C2.1.9 Asphaltene content based on GC

Table C2.9: Explained variance and Cross Validation Standard Deviation for the model for asphaltene content based on GC data.

Number of components	Explained variance	Cross Validation Standard Deviation
1	40.72%	0.92
2	54.98%	1.06
3	78.72%	1.37
4	90.63%	0.99

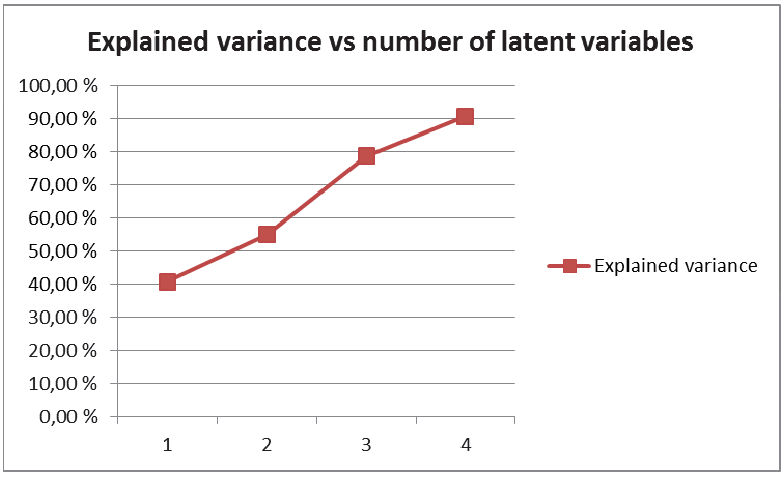


Figure C.2.50: Explained variance plotted against the number of Latent Variables for the model for asphaltene content based on GC data.

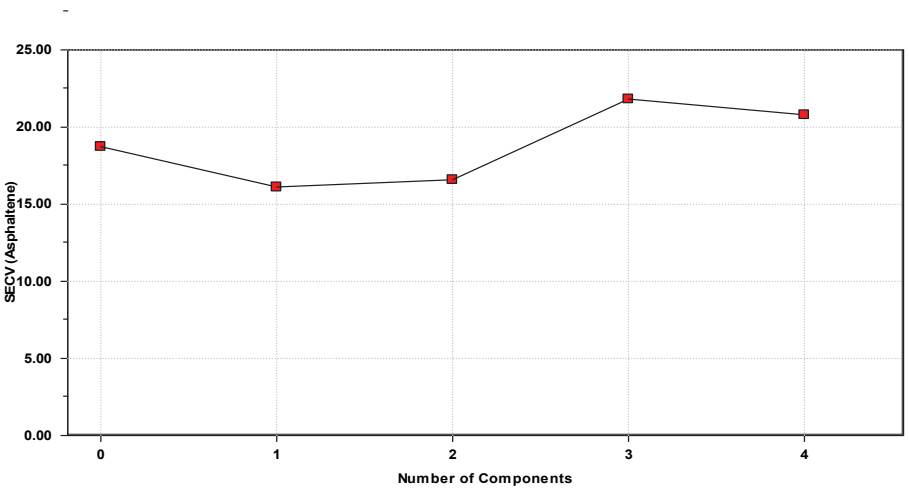


Figure C2.51: SECV plotted against number of components for the model for asphaltene content based on GC data.

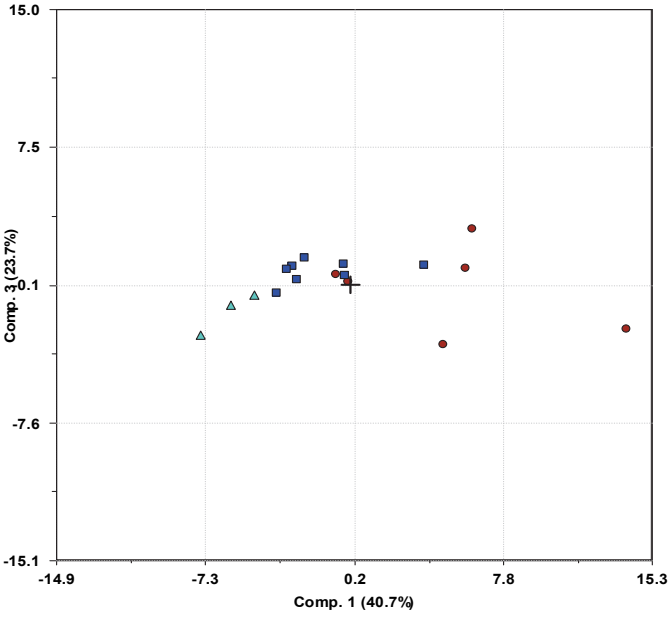


Figure C2.52: Scores for the model for asphaltene content based on GC data. Component 1 vs Component 3. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates.

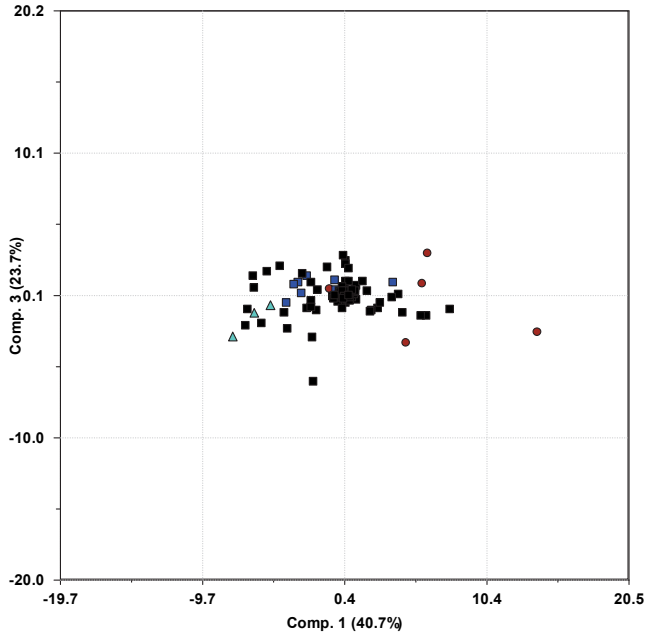


Figure C2.53: Biplot for the model for asphaltene content based on GC data. Brown circles= biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates. The black squares are the variables, with the GC variables grouped in the middle.

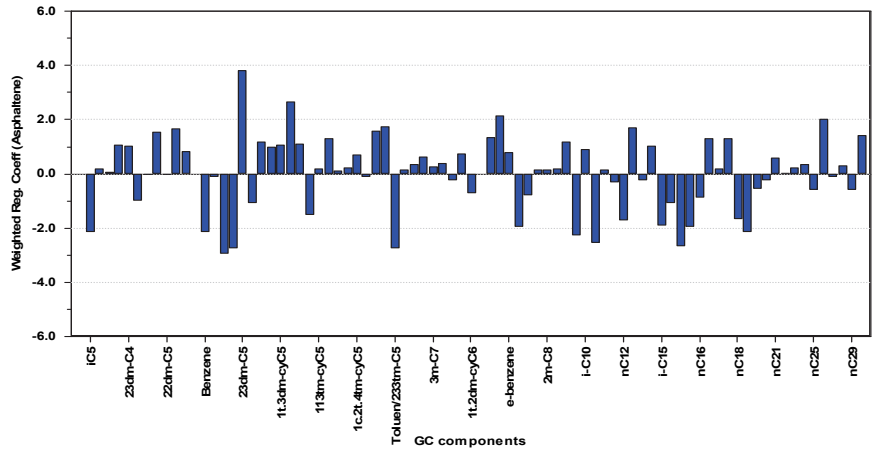


Figure C2.54: Weighted regression coefficients for the model for asphaltene content based on GC data.

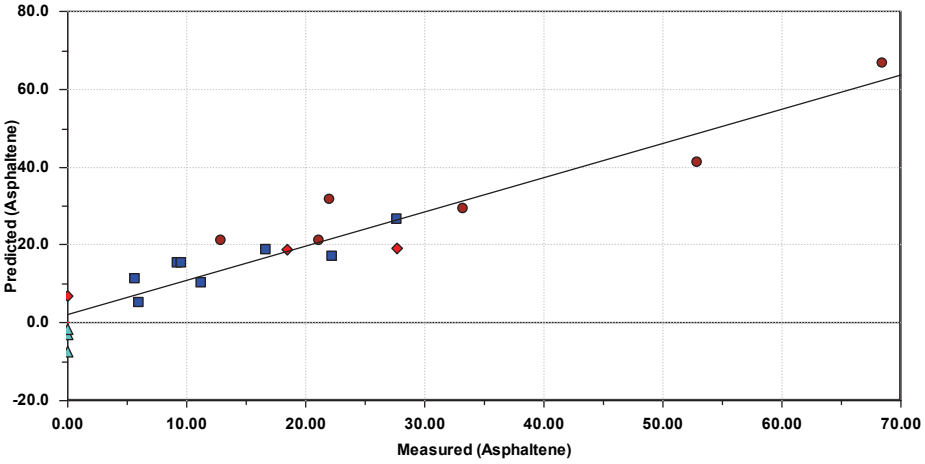


Figure C2.55: Predicted value plotted against measured value for the model for asphaltene content based on GC data. R^2=0.893. Brown circles=biodegraded oils, blue squares=nondegraded oils, cyan triangles=condensates, red diamonds=validation objects.