Characterization and Classification of Crude Oils Using a Combination of Spectroscopy and Chemometrics

Karakterisering en classificatie van ruwe aardolie met behulp van gecombineerde spectroscopie en chemometrie

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof. dr. J.C. Stoof,
ingevolge het besluit van het college voor promoties in het openbaar te
verdedigen op dinsdag 15 december 2009 des ochtends te 10.30 uur

door

Peter de Peinder

geboren op 29 augustus 1967 te Tiel

Promotor: Prof. dr. ir. B. M. Weckhuysen

Co-promotor: Dr. T. Visser

The research presented in this thesis was financially supported by Shell Global Solutions Nederland B.V.

Voor mijn ouders en Joop

ISBN: 978-90-39351895

Contents

Chapter 1

General Introduction

1.1. Introduction

Crude oil, or petroleum[1] as it is often called in the United States, is the world’s most important fossil fuel today and the basic material for an immense range of products in society. Crude oil is primarily used for the production of transportation fuels, such as gasoline, diesel and kerosene, but it is also the major source for the synthesis of a range of polymers, paraffins and bitumen. This makes crude oil of invaluable economical importance [1-6]. The current amount of exploitable quantities of conventional crude oil is estimated at 1200 billion barrels, while the amount of so-called unconventional crude oils, such as the Canadian tar sands, is expected to provide around 1300 billion barrels of crude oil equivalent [1, 7]. As schematically depicted in Fig. 1, the world's remaining crude oil resources are mainly non conventional, which means that they can't be extracted using traditional oil well methods.

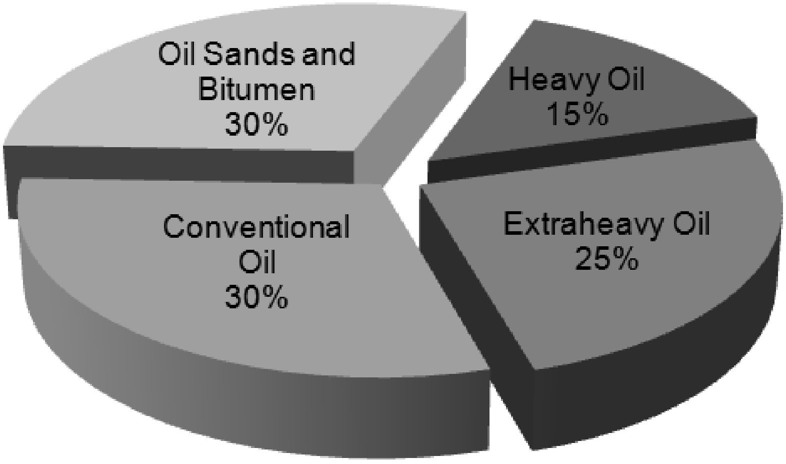


Fig. 1: Distribution of the total world crude oil resources [1].

Although, the proven quantities are huge, oil sources are obviously not inexhaustible. The daily consumption anno 2009, for instance, is around 90 million barrels and this number is still increasing due to the growing demand for energy in rapidly expanding economies like China and India and the still

---

[1] The name petroleum is derived from the Greece ‘Petra Oleum’, which means rock oil.

increasing world population. According to the International Energy Agency this will lead to shortness in the exploitable amounts of crude oils around 2030 and hence to a proportional increase of the price of energy [2-4]. This scenario and the growing environmental concern, *i.e.*, the supposed global warming as a result of the burning of fossil fuels, has urged the search for alternative and sustainable energy sources like bio-fuels, wind mills, solar cells and nuclear power plants. In parallel, it has stimulated research to a more efficient and 'greener' use of crude oils and related products. As a result, for example, improved processes have been developed to facilitate cracking of heavy fractions and hence to explore tar sands and substances that were initially regarded as 'waste' products like coke and bitumen. Likewise, studies have been undertaken to reduce the emission of polluting elements like sulfur, vanadium, copper and nickel that are released upon combustion of fossil fuels. Finally, a variety of physical, chemical and spectroscopic characterization techniques have been developed to determine the qualitative and quantitative composition of a crude oil in detail, since this type of information is inevitable for a complete and optimal use of the crude oil feed stock.

1.2. The origin of crude oils

There have been many hypotheses for the origin of crude oil. It has been proposed, for instance, that crude oil is formed from methane from the earth's interior, where rock melting may have generated it abiotically [8, 9]. Nowadays, the most commonly accepted theory is that crude oil originates from plant and animal life of 100 to 600 million years ago, where it has been formed from the preserved sediments of zooplankton, algae and animals on the bottom of prehistoric seas [2-5]. The presence of porphyrins, organic molecules that are structurally very similar to chlorophyll in plants and hemoglobin in animal blood, are used as evidence for the anaerobic conditions during the formation of oil [10]. The local oxygen poor conditions enabled anaerobic bacteria to digest the organic material into a solid substance that is named 'earth wax' or 'kerogen' and which is defined as the ancestor of crude oil. The anaerobic conditions explain the very low concentration of oxygen containing molecules in crude oil. The continuous formation of new sediments on top of the kerogen layers resulted not only in an increasing pressure, but also in a considerable temperature raise. As a result, the kerogen transformed into crude oils at temperatures up to 100℃ and to earth gas at higher temperatures. A typical depth for the oil formation is 4-6 km, but oil might migrate after formation to much shallower depths.

1.3. The composition of crude oils

The local circumstances during the formation of crude oils are variable and for that reason, the molecular composition of crude oils can be diverse. First of all, crude oils contain a large variety of organic compounds, which may exceed the number of 20.000 different molecular structures [11]. The main fraction is always hydrocarbons and ranges from 97 weight % in light crude oils to about 50 wt % in bitumen and tar sands [6]. Most of these are straight and branched alkanes, cycloalkanes and (polycyclic) aromatic compounds, but the precise composition depends on the geological conditions under which it was formed. Crude oils from the Middle-East for instance, contain a relatively large amount of paraffins, whereas crudes from Nigerian wells count a greater proportion of cyclic alkanes and aromatic species. In contrast, the heavy Chinese crudes have a very high concentration of polyaromatic ring structures. To illustrate the large variety in hydrocarbon composition of crude oils, the range of relative weight percentages for the compound classes paraffins, naphtenes, aromatics and asphaltenes have been listed in Table 1.

Table 1. Average and range of the relative percentage of four types of hydrocarbons, expressed in weight percentage, present in crude oils [2, 5].

Hydrocarbon type	Average (weight %)	Range (weight %)
Paraffins	30	15 - 60
Naphthenes	49	30 - 60
Aromatics	15	3 - 30
Asphaltenes	6	5-50

Asphaltenes are defined as the fraction of the crude oil that does not dissolve in n-heptane or n-hexane but in benzene or toluene. It is the heaviest and most polar fraction of the oil, which ends up in high concentrations in the residuals after extraction or distillation of the lighter fractions. For obvious reasons, asphaltenes form a major source of problems for the oil industry in recovery, blending, cracking, storage etc. [12]. The problems tend to become worse since nowadays the heavier oils need to be extracted as well and larger amounts of light fractions are extracted from crudes.

As already mentioned, oxygen containing compounds are hardly present in crude oils due to the anaerobic conditions during genesis. Nitrogen and sulfur compounds, however, form a substantial part, in particular the benzothiophenes. Finally, small amounts of inorganic matter are found. Next to mineral particles, which are obviously abundant in the tar sands, elemental carbon and sulfur are present next to (traces of) iron, vanadium, nickel and copper. Also here, the geological origin largely defines the precise concentrations and appearance, but on average, the amount per element varies over a rather small range as listed in Table 2.

Table 2. Chemical elements and their approximate concentrations, expressed in weight %, present in crude oils [2, 5].

Element	**Weigth %**
Carbon	83-87
Hydrogen	10-14
Nitrogen	0.1-2
Oxygen	0.1-1.5
Sulfur	0.5-9
Metals	0.001-0.1

1.4. Crude oil production and refining

In principle, three different techniques are used to extract crude oils from the oil fields. The primary method is straightforward pumping the liquid material from the well. This technique is relatively easy, fast and facilitated by the fact that a natural high pressure is often present underground. The second method is applied when the natural underground pressure is absent or insufficient. As the reservoir pressure drops towards the end of a field's life time, artificial lifting and water flooding and gas injection techniques are used to continue production. The third method is used in case the high viscosity of a crude oil does not admit the application of the other methods. It involves the injection of heat, solvents or miscible gasses, such as CO_2, to enhance the fluidity. Currently, tar sand is the main source of non-conventional oils, but, due to their consistency, extraction of the oil is much more challenging than from conventional wells. For tar sands, a high temperature treatment by injecting steam, solvents, and/or hot air into the sands is often applied and occasionally, a chemical transformation into smaller transportable hydrocarbons is applied. As a consequence, these processes require considerably more energy than conventional oil extraction and hence oil production from these sources is more expensive.

The first step after gaining crude oils is refining. Initially, this was done by straightforward distillation. However, since the obtained fractionation, based on differences in boiling point only, turned out to be not very economic, nowadays two linked processes are generally applied, i.e. fractionation and cracking. This is schematically shown in Fig. 2.

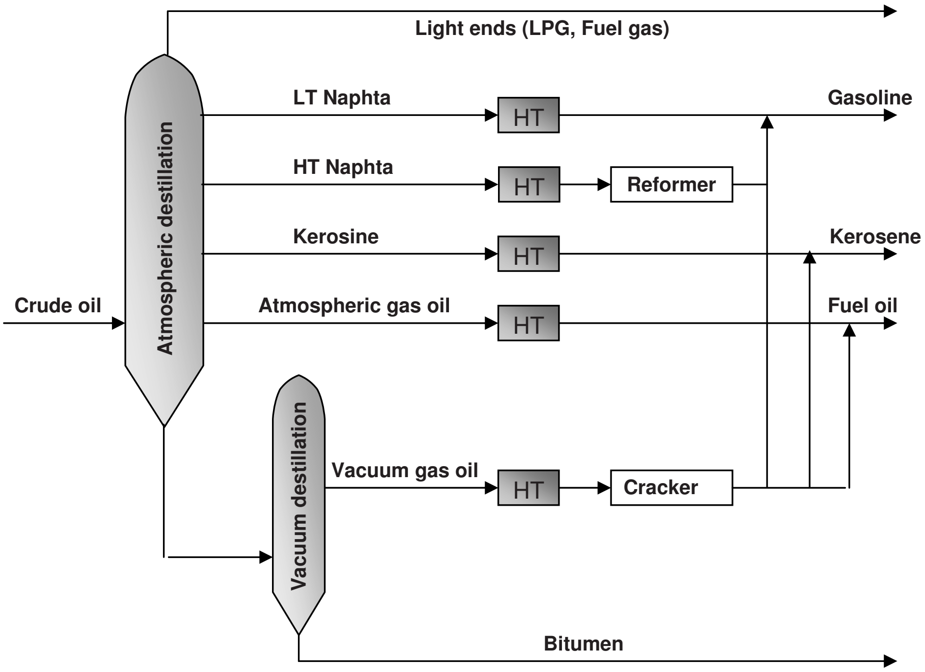


Fig. 2: Scheme of a crude oil refinery. After preheating, the crude oil is brought into the atmospheric distillation tower. Blocks indicated with HT are hydrotreaters.

After distillation, the different fractions are transported to transformer units where hydrogenation, desulfurization and cracking are carried out to optimize the yield of the fuels as indicated. Furthermore, the residuum of the atmospheric distillation is directed to a vacuum distillation unit of which the distilled fraction is further processed, while the bitumen residue is used for other applications. As such, bitumen is often seen as a side product of the oil industry. However, it is widely applied as an important component for road construction, roofing, waterproof adhesive and sealing. Therefore, it has a substantial economical value. Besides, huge amounts of bitumen are also found in nature, i.e., buried in (tar) sands and in so-called bitumen lakes. For that reason, oil companies are very careful to fully assess the bitumen from a crude oil to ensure that it meets certain specifications, before proceeding to approving for normal refinery production.

1.5. The characterization of crude oils

It is clear from the preceding, that crude oils can have various compositions and hence different chemical and physical properties. For many reasons, knowledge of these properties is very important. First of all, it defines the potential applicability and hence whether it is economically justified to exploit an oil reservoir. Furthermore, information on the composition of a crude oil is critical to the selection of refining of the downstream processes and of the refining conditions [11]. For example, knowledge of physical parameters like the viscosity, liquidity and the presence of solid particles is crucial to prevent problems like clogging or solidification during transportation and processing. For that reason, the determination of physicochemical profiles or so called crude oil assays is common practice in oil industries. As such, a crude oil assay is the chemical evaluation of a crude oil feedstock. These tests provide not only more general information such as for example the Saturated, Aromatics, Resins, Asphaltenes (SARA) distribution, but also a series of chemical and physical properties that are important for refinery engineers, oil traders and producers. The suitability of crude oils for bitumen manufacturing, for instance, is based on the determination of more than 10 so-called long residue (LR) and short residue (SR) properties. The LR properties are defined as the physical and chemical properties of the residues flashed at 370℃, whereas the SR properties represent a number of physicochemical parameters of the residues flashed at higher temperatures. Another important parameter of crude oils is the sulfur concentration. As previously described, crude oils originate from anaerobic bacterial reduction of sedimentary biological matter. As a result, crude oils contain hardly any oxygen, but sulfur can be present in considerable amounts up to 9 weight %. The presence of sulfur species has a number of direct drawbacks, such as corrosion of the production plant and clogging in pipelines [11]. However, most importantly, sulfur has a huge environmental impact due to the formation of sulfur oxides (SO_x) upon combustion of crude oil fuels. Therefore, next to the determination of other properties, sulfur speciation and characterization is also a major issue in crude oil production and refinery. For obvious reasons, these physicochemical tests can be quite complex and extensive. In practice, many tools are used for the characterization, depending on the application.

Next to relatively simple tests to determine for instance the density, viscosity and total sulfur content, also rather complex experiments are carried out such as the characterization of the boiling range fractions by means of physical or simulated distillation and the speciation of sulfur compounds in a series of compound classes by 2-dimensional gas chromatography. At present, the majority of these tests are carried out by standard methods developed by the American Society for Testing and Materials International (ASTM International) [13] and the Energy Institute, formerly known as the Institute of Petroleum (IP) [14]. However, most of these methods are rather slow, elaborate and expensive, requiring large sample quantities of the crude oils to be sent to the laboratory.

As a fast and viable alternative for these ASTM and IP tests, spectroscopic methods have been proposed already in 1995, since the corresponding spectra reflect the complete molecular composition of a crude oil [15]. Besides, spectroscopic techniques are able to supply information on interactions on a molecular scale. For that reason, the potentials of several spectroscopic methods for the characterization and classification of crude oils and crude oil products have been studied. Many of the results have been reported in the open literature, or filed in patents, comprising UV-Vis, infrared

(IR), near infrared (NIR), Raman, nuclear magnetic resonance (NMR) and mass (MS) spectrometry, often in combination with gas and liquid chromatography. Particularly, when combined with multivariate data analysis or chemometric modeling techniques spectroscopic data appear to be very useful to characterize and classify crude oils in many ways. In the next paragraphs, a brief description is given of the different spectroscopic methods that have been used in this thesis, including some examples from literature on the application of the technique in crude oil analysis. Furthermore, the basic principles of the chemometrical methods Partial Least Squares (PLS) regression and Principal Component Analysis (PCA) in relation to spectroscopic data are given.

1.5.1. UV-Vis spectroscopy

Ultraviolet-visible (UV-Vis) spectroscopy covers the wavelength range from 200-800 nm (50000-12500 cm^{-1}). It is used to register the absorptions that are the result of electronic transitions between atomic or molecular orbitals. The frequency of a UV-Vis absorption band is correlated to a molecular structure or an oxidation state. This makes UV-Vis spectroscopy a useful tool for structural analysis and hence a potential technique for the characterization of crude oils. In organic chemistry, UV-Vis spectroscopy is widely used to investigate the extent of multiple bond or aromatic conjugation within molecules. Solutions for many organic compounds can easily be prepared in solvents with limited to no absorption in the UV-Vis range, such as hexane or dichloromethane. Since vibrational and rotational levels of the molecular orbitals are superimposed upon the electronic levels, the absorption bands are rather broad. This makes structural elucidation of unknown compounds difficult, but the technique is very useful for quantitative measurements.

According to literature, the number of applications of UV-Vis spectroscopy in crude oil analysis is rather small [16-24]. Direct measurement is hampered by the high absorption coefficients of crude oils, which requires dissolution in a proper solvent prior to spectral acquisition. A typical UV-Vis spectrum of a crude oil is presented in Fig. 3. UV-Vis spectroscopy is found to be particularly valuable to determine the amount of asphaltenes [16-21, 23], although it should be noted that due to asphaltene aggregation a non linear behavior with the concentration can occur [22]. Another method to determine the asphaltenes content comprises measuring the absorbance at 750 and 800 nm in a solution or suspension in n-heptane [21]. This method is simple and straightforward applicable to crude oils as well as to bitumen. Furthermore, it has been demonstrated that complex peak structures can be revealed for crude oil UV-Vis spectra by mathematically pre-processing of the presumably featureless spectra [24]. UV-Vis spectroscopy is also applied to determine mono- and bicyclic hydrocarbons by characteristic absorptions at 200 and 230 nm [16]. Finally, it is used to study the binding of metals like Ni and V in tar sand and bitumen [17].

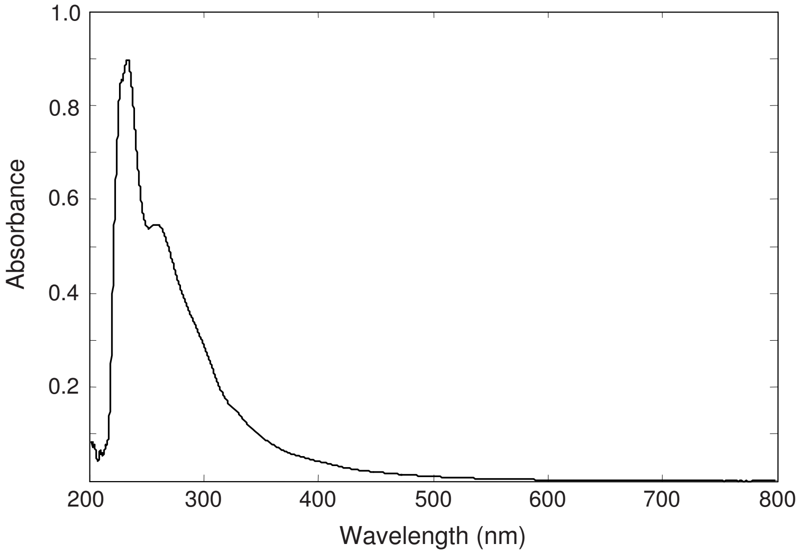


Fig. 3. Typical UV-Vis spectrum of a crude oil (ratio 1:4000 in CH_2Cl_2).

1.5.2. Infrared spectroscopy

Infrared (IR) spectroscopy spans the wavelength region from 2.5 to 25 µm (or 4000-400 cm^{-1}). It is the energy range that corresponds with the frequency of molecular vibrations. IR is a widespread and well accepted technique for molecular characterization. Intrinsically, IR spectra are very rich in information, since each molecule exhibits 3N-6 normal modes with N being the number of atoms in the molecule. The frequency of molecular vibrations is defined by the bond strength and the mass of the atoms. It increases with the bond strength and decreases with the atomic mass. Besides, the vibrational frequencies are sensitive to changes in the intra- and intermolecular interactions. The intensity of an IR absorption band depends on the change in the dipole moment during the vibration: a large dipole moment change gives rise to strong absorption and a small change to a weak band. Therefore, IR spectroscopy is particularly useful in determining 'polar' functional groups, such as C=O, N-H and O-H. However, less polar aliphatic, olephinic and aromatic C-H and C-C vibrations also exhibit characteristic IR bands, which make IR also a valuable technique for crude oil analysis. As an example, a typical IR spectrum of a crude oil is presented in Fig. 4.

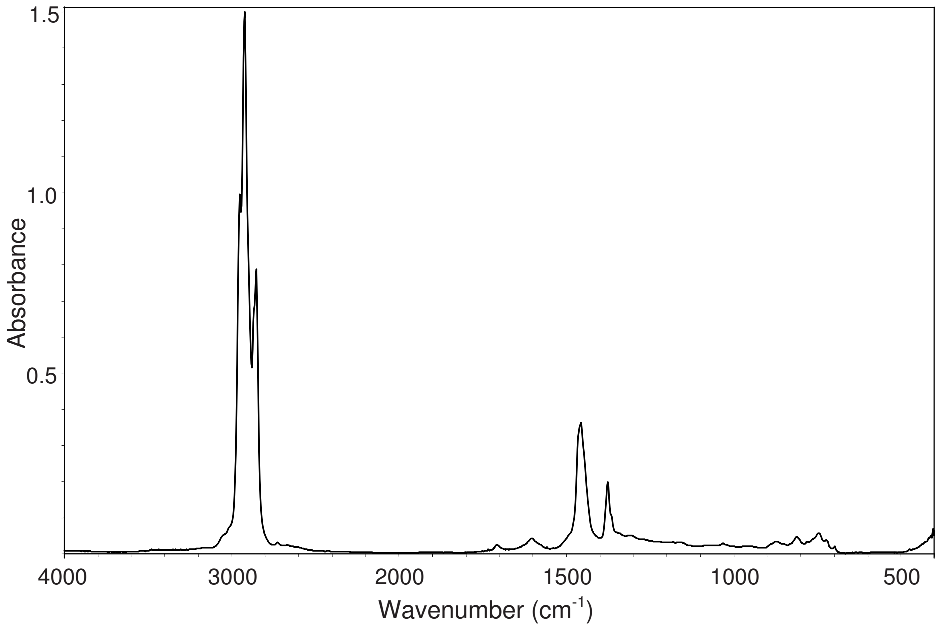


Fig. 4. Typical IR spectrum of a crude oil.

Numerous applications of IR spectroscopy in the analysis of crude oils and related products have been published over the years [19, 25-50]. Most of these studies deal with the characterization of crude oils, fuels, asphalts and bitumen [25, 29, 31, 34, 36, 40, 41]. Occasionally, an oxidation step is applied prior to the spectroscopic analysis in order to convert the relatively weak IR absorbing carbon-sulfur and sulfur-hydrogen vibrations into S-O and S=O, since the latter functional groups have a much higher intensity in the infrared [49]. IR spectroscopy is also used to quantify the degree of aromaticity and the hydrocarbon composition by straightforward integration of the corresponding peaks [37, 38, 44, 48]. Other applications are the assessment of crude oil quality [33] and the classification of crude oils and coal, based on specific IR spectral features [30]. Finally, the technique is often used to predict crude oil and bitumen properties from their IR spectra [32, 33, 37, 39, 46, 47, 50].

1.5.3. Near infrared spectroscopy

Near infrared (NIR) spectroscopy is carried out in the region 800 to 2500 nm (12500-4000 cm^{-1}). In this part of the electromagnetic spectrum, overtone and combination bands of fundamental vibrations are found. According to the selection rules of quantum mechanics, they are forbidden for harmonic vibrations, but for anharmonic vibrations like C-H and C=O stretching, they are allowed. The corresponding molar absorptivity is much smaller than for fundamental modes as observed in IR spectroscopy, which opens the possibility to use much longer path lengths. Moreover, since glass can be used as sample cell material, NIR analysis is very easy and straightforward. As illustrated by the typical NIR spectrum of a crude oil in Fig. 5, NIR absorption bands are usually very broad. As a consequence, it is difficult to assign specific features to chemical structures. For that reason, NIR is often used in combination with multivariate data analysis techniques to extract qualitative or quantitative information from the spectra.

Due to the easy sampling properties, the number of applications of NIR spectroscopy to crude oil analysis is considerable, particularly in combination with chemometrics. Similar to IR spectroscopy, it is used for the determination of physical and chemical properties of crude oils and crude oil fractions [53, 54] as well as for the prediction of parameters like viscosity and density [33, 55, 59, 63]. Furthermore, NIR spectroscopy is applied to classify the geographical origin of petroleum crudes [58] and the characterization of bitumen properties in order to determine the process ability of ores [61].

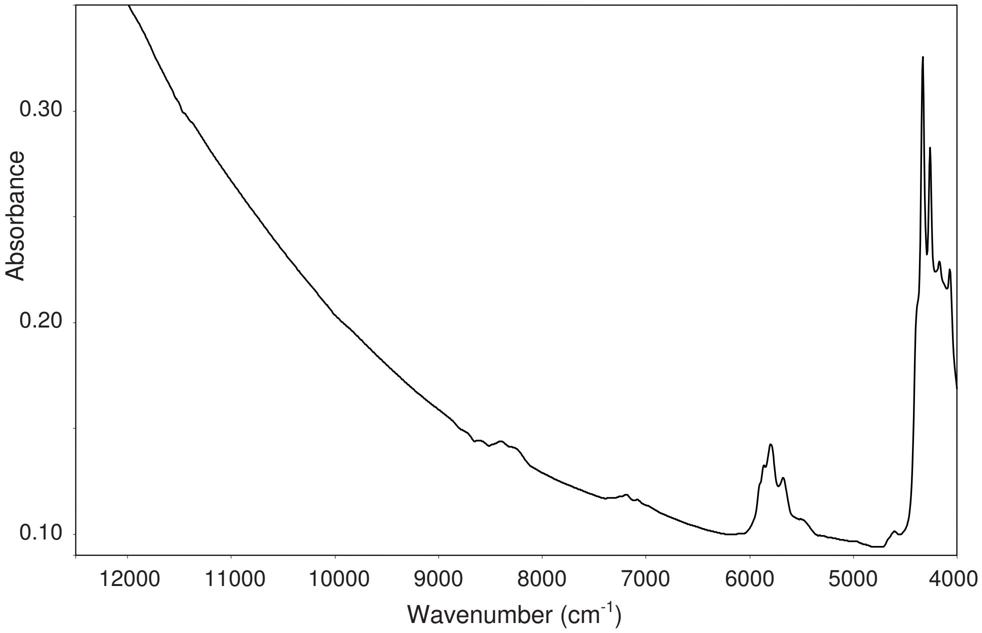


Fig. 5. Typical NIR spectrum of a crude oil.

1.5.4. Raman spectroscopy

Similar to IR and NIR spectroscopy, Raman spectroscopy is used to study molecular vibrations. However, different from the photon absorption mechanism of (N)IR spectroscopy, Raman is based on the phenomenon of inelastic scattering of radiation by a sample molecule. The light interacts with the electrons and bonds of the molecule and excites the molecule into a virtual state. In this physical process, a part of the photon energy is transferred to a molecule as a result of the transition of the molecule to an excited vibrational state. The partial energy loss of the photon that is associated with this transition is expressed relative to the excitation wavelength of the laser and is called the Raman shift. This shift, expressed in cm^{-1}, matches with the frequency of the corresponding molecular vibration and has exactly the same position in the Raman spectrum as in the IR spectrum provided the vibration is both Raman and IR active. The activity and hence the spectral intensity is defined by the selection rules. In IR, the intensity depends on the change in dipole moment during the vibration, whereas in Raman it relates to a change in polarizability. As a consequence, the more symmetrical functional groups, like C=C, dominate the Raman spectrum, whereas more polar groups, like C=O, cause the most IR intense bands. This difference makes both techniques complementary.

A drawback of Raman spectroscopy is its low scattering efficiency, which makes a laser source necessary to obtain a spectrum. Besides, the occurrence of fluorescence is a major nuisance of Raman spectroscopy. This phenomenon largely limits the application of Raman spectroscopy to crude oil analysis [64-71]. In order to reduce fluorescence, a pre-treatment with powdered charcoal and filtering is found to be effective but laborious [64]. Raman has been successfully used to determine the structural state of carbon in solid bitumen and kerogens [65, 66]. Another application is the analysis of distillation fractions of syncrude sweet blend to determine the aromatic content [68]. Lasers of various wavelengths were used to obtain useful Raman spectra and to calculate trends in the mono- and bicyclic aromatics content. Shoute *et al.* applied UV lasers to access oil sands-derived bitumen and other commercial petroleum products with Raman spectroscopy [69]. Only wavelengths below 240 nm, yielded fluorescence free Raman spectra which can be useful to determine the molecular composition of highly complex bitumen samples. Finally, Raman spectroscopy turned out to be a successful method to determine the asphaltene dimensions in Algerian crude oil [70, 71].

1.5.5. Nuclear magnetic resonance

Some elements, like hydrogen (^1H) and the ^{13}C isotope of carbon possess a nuclear spin. Due to this property, the nuclei of these elements behave as tiny magnets. It allows the detection of these elements in a molecule by applying a periodical external magnetic field. This technique is called Nuclear Magnetic Resonance (NMR) spectrometry and can be used to obtain information on molecular structures. NMR spectra reflect the energy shift of a specific atom compared to a reference compound. This shift is called chemical shift and is expressed in ppm units. In addition, the integrated peak intensities are proportional to the number of atoms that is present in the molecule. The environment of the nucleus, in terms of chemical bonding and neighboring atoms determines at which chemical shift an NMR peak is observed. Aromatic carbons, for instance, show up at a chemical shift of about 120 ppm, whereas aliphatic carbon atoms appear around 20 ppm in ^{13}C-NMR. In ^{1}H-NMR the aromatic hydrogens are found around 7 ppm and the aliphatic ones around 1 ppm. Typical ^{13}C-NMR and ^{1}H-NMR spectra of crude oil are presented in Fig. 6.

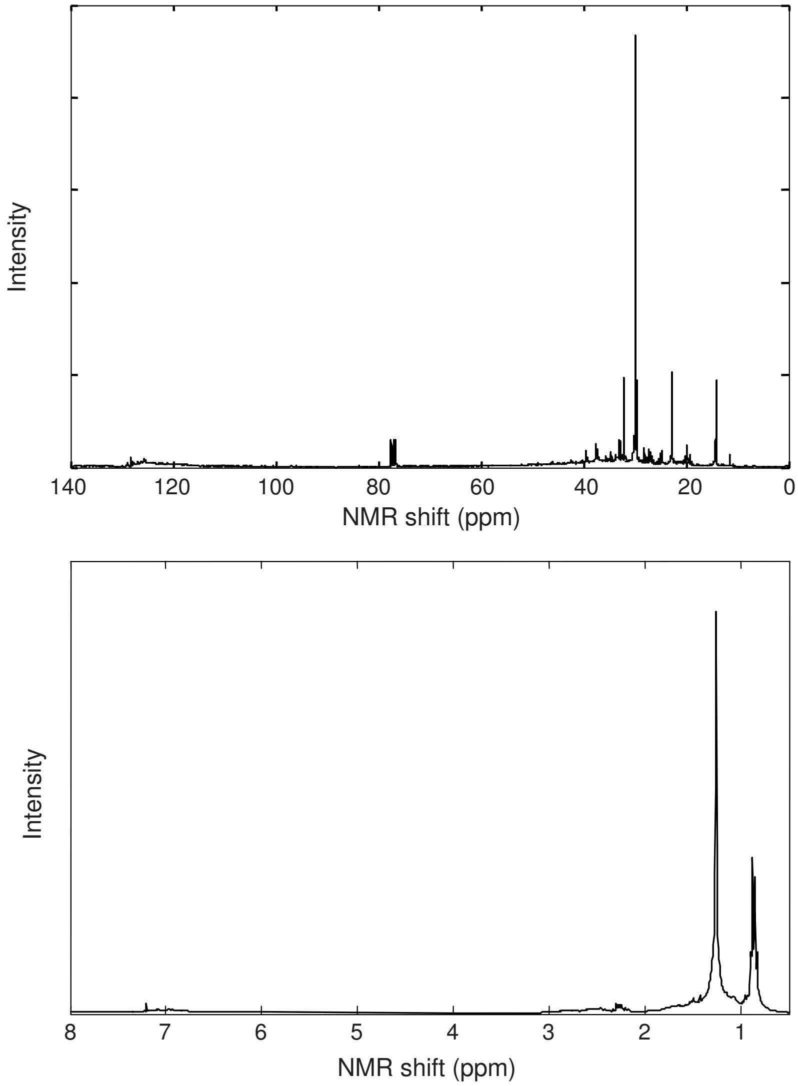


Fig. 6. Typical ^{13}C-NMR (top) and ^{1}H-NMR (bottom) spectra of a crude oil dissolved in CH_2Cl_2.

NMR spectrometry has been applied in several studies on the structural elucidation of molecules present in crude oils and crude oil fractions [72-88], such as asphaltenes [73], vacuum gas distillates [74] and bitumen [75]. Furthermore, ^{1}H-NMR has been used for the characterization of oil and bitumen [8-81] and to determine the viscosity of heavy oil and bitumen [85]. More recent examples include the structural characterization of fluid catalytic cracking feeds [86], the analysis of hydrocarbon mixtures and asphaltenes [87] and the characterization of soluble fractions of coal tar pitch [88].

1.5.6. Chemometrics or multivariate data analysis

Chemometrics is commonly defined as the application of statistical methods to chemical data in order to establish correlations between sample properties and complex analytical data. A variety of advanced mathematical data reduction and pattern recognition techniques is used for this purpose. The combination of chemometrics and spectroscopic data is well established. The most versatile techniques used in spectroscopy are Principal Component Analysis (PCA) and Partial Least Squares (PLS) regression. Both techniques can make use of the complete spectral region, which allows for extracting information from very complex spectra.

PCA is a data reduction technique and can be used qualitatively without any other information than the spectra. PCA decomposes a spectral data matrix **X** into a score matrix **T** and a loading matrix **P** according to the matrix formula **X**=**TP**. The formula is schematically depicted in Fig. 7. In this figure, matrix **X** represents *n* spectra, consisting of for instance *p* wavenumbers. The loading matrix **P** represents the *r* Principal Components (PC's) which are spectrum like patterns with *p* wavenumbers.

The score matrix **T** consists of the coordinates of the *n* original spectra on the new axes determined by the corresponding PC's. PCA will capture the maximum variance in the spectra and this will result in the first Principal Component PC1. The next largest variation is called PC2 and so on. These PC's will form the new axes for the dataset **X**. PCA will capture the most relevant information in the first few scores and noise will be represented by the higher PC's. If the original spectra consist of 2000 wavenumbers this means that the information is contained in a 2000 dimensional space. If the first 3 PC's capture more than 99 % of the variance, this reduces the space to only 3 dimensions while losing less than 1 % of the information.

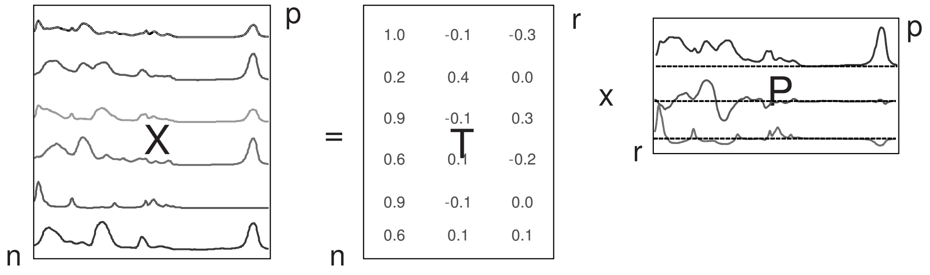


Fig. 7. Schematical representation of the PCA methodology.

PLS is a quantitative method and aims to build a regression model that can be used to predict a certain physical or chemical sample property, for instance, the concentration. Next to the spectral data set, PLS modeling requires a data matrix that contains the value of this property for each sample in the set. Alike PCA, PLS extracts the largest variation, but in addition, this spectral variation must be correlated to the parameter value as well. The PLS method is schematically depicted in Fig. 8. The main difference with PCA is the fact that the sample parameter matrix **C** is now involved and influences the decomposition of **X** through matrix **U**. The loading matrix **P** will now contain Latent Variables (LV's), which are comparable with PC's. During calibration the matrices **T**, **P**, **U** and **Q** are determined based on the known values of matrix **C**. Then spectra can be measured and the corresponding concentration can be predicted using the model parameters calculated in the calibration step, in which a regression vector is created. The dot product of this regression vector and a sample spectrum is a single value which corresponds to the modeled concentration. For a leave-one-out validated model, this vector is the average of all of the regression vectors calculated when each sample is left out of the model calculation. Plotting the regression vector can be very useful in determining the number of LV's to include in the model and in determining which spectral regions are most important in modeling the concentration.

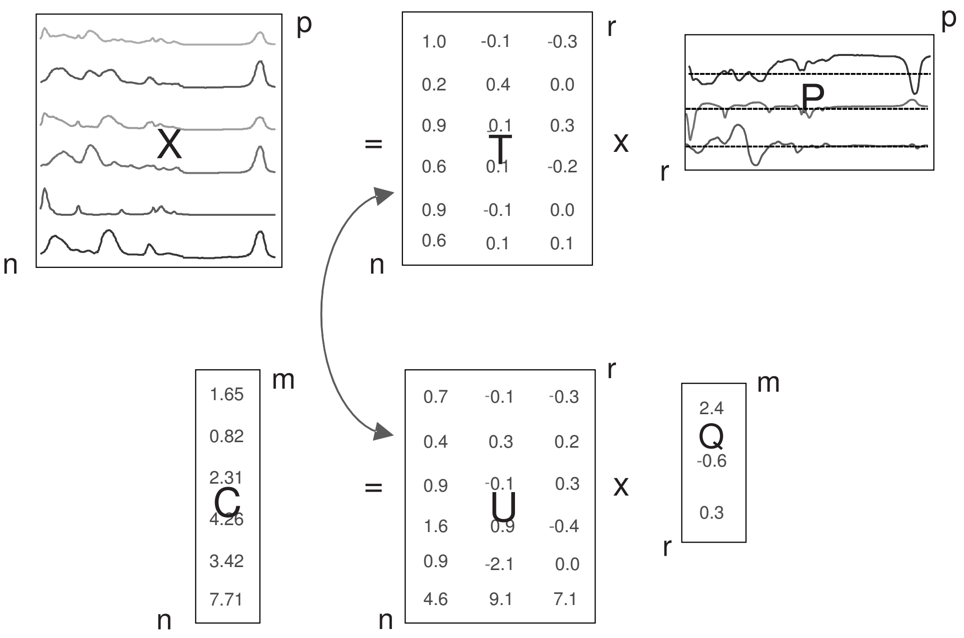


Fig. 8. Schematical representation of the PLS regression methodology.

PCA and PLS are often used, particularly in combination with vibrational spectroscopy. This is due to the fact that the information content of vibrational spectra is high, but very difficult to access by straightforward interpretation. Many examples of PCA and PLS as applied to the interpretation of IR [42, 43, 47, 89, 94-96] and NIR [55, 56, 58-60, 62, 90] spectra can be found in literature and patents. Teixeira *et al.*, for instance, applied multivariate calibration of IR data to detect alterations in Brazilian gasoline [94], while Honorato and co-workers used PCA to find the best calibration settings for simultaneous IR spectroscopic determination of gasoline properties [95]. PCA has also been used to compare the biodegradation level of crude oil by IR spectroscopy [96]. Other chemometrical applications to analyze IR data are the prediction of the source rock origin [91] and the determination of chemical-physical properties of gasoline [39]. For NIR spectroscopy, chemometrics has been used, for example, for the quantitative determination of physicochemical properties in the petrochemical industry [59] and to estimate crude oil properties from the spectra of residual fractions [60]. Other multivariate data analysis methods have been used in crude oil analysis as well, such as moment combined partial least squares (MC-PLS) regression to analyse petroleum and petrochemical products [92] and neural networks to determine correlations between chemical and rheological properties of asphalt [97].

1.6. Scope and outline of the PhD thesis

The aim of the research, described in this PhD thesis, was to explore the potential of various molecular spectroscopic techniques in combination with chemometric modeling techniques, to classify crude oils and blends in terms of a number of physicochemical properties that are relevant for oil industries. The main goal was to predict bitumen properties, in terms of so-called long residue (LR) and short residue (SR) properties. LR properties are defined as a series of physical and chemical properties of the residues obtained by flashing the oil at 370℃, while SR properties are described as physicochemical properties of the residues flashed at higher temperatures, *i.e.* 400-590℃ Maxwell-Bonnell. The LR and SR properties are used to determine the suitability of crude oils for bitumen manufacture and to predict bitumen properties as a function of the flash temperature directly from crude oil information. The underlying goal of this study was to link the more fundamental chemical composition of crude oils to the SR properties and thus to their bitumen quality. For this purpose six different spectroscopic techniques have been selected, namely mid IR, NIR, Raman, UV-Vis, ^{1}H-NMR and ^{13}C-NMR. The second subject of the research was to explore the potentials of spectroscopy as a tool for the sulfur speciation of crude oils. This study has been carried out with IR spectroscopy only. A clear objective of both studies has been to utilize state-of-the-art instrumentation and software packages in order to allow implementation of the results in crude oil production plants and refineries.

First, research has been carried out to determine the potentials and limitations of different spectroscopic techniques combined with chemometric modeling to classify crude oils in terms of a number of physicochemical properties. The results of this characterization study are described in **Chapter**

2. As spectroscopic methods, Raman, UV-Vis, IR, NIR and ^{1}H- and ^{13}C-NMR spectrometry have been investigated, while PLS regression and PCA have been used as chemometrical methods to build prediction models from the different spectra types. Models have been built for the LR properties in order to determine their suitability in bitumen application. Since modeling of IR spectra and, to a slightly smaller extent, NIR spectra turned out to give the best results, these types of data have been investigated in more detail and models were developed for the SR properties as well.

In order to determine the potentials of combining complementary spectral methods, the two most promising techniques as determined in Chapter 2, have been studied in **Chapter 3**. This chapter describes the results obtained from PLS modeling of IR and NIR spectra combined with ^{1}H- and ^{13}C-NMR data to predict long residue properties of crude oils and blends. Next, the effect of temperature treatment of crude oils and blends prior to recording the IR spectra has been studied. Two treatments have been investigated: 1) exposure of the crude oil for 24 h to 65 °C to reduce the contributions of volatile constituents, and 2) recording of the spectra as a function of the temperature over the range 20 to 65 °C with increments of 5 °C. The results of PLS-modeling of the corresponding spectra and the resulting effect on the accuracy of the prediction of LR and SR properties are presented in **Chapter 4**.

Chapter 5 summarizes the results of a feasibility study of PLS models that are based on mathematically mixed IR spectra of pure crude oils in order to predict LR properties of blends. If successful, this method allows fast screening of crude oils for potential application in a blend and eliminates the need for the physical preparation of such mixtures. **Chapter 6** describes the results of a study to the robustness of the prediction models. This was carried out by checking the effect of the spectral reproducibility on the results of PLS modeling and prediction. The same was done for spectra acquired from different spectrometers, equipped with different sampling accessories. As a spin-off of the preceding chapters, the PLS modeling of the IR spectra as a potential tool for sulfur speciation of crude oils has been studied. The results of that research are presented in **Chapter 7**.

Finally, **Chapter 8** summarizes the conclusions from the research described in this PhD thesis and finishes with some recommendations for future research.

References

1. H. Alboudwarej *et al.*, Oilfield Rev. Summer (2006) 34.
2. www.opec.org/library/FAQs/crude.htm.
3. www.oil-price.net/links.php?lang=en.
4. www.worldenergyforum.com/blog/1.
5. en.wikipedia.org/wiki/Petroleum.
6. J. Speight (Ed.), *The Chemistry and Technology of Petroleum*, Marcel Dekker, New York (1999) 215.
7. S. Zhao, L.S. Kotlyar, J.R. Woods, B.D. Sparks, J. Gao, J. Kung, K.H. Chung, Fuel 81 (2002) 737.
8. T. Gold, S. Soter, Scientif. Americ. 242 (1980) 154.
9. M.A. Lur'e, I.Z. Kurets, F.K. Shmidt, Chem. Technol. Fuels and Oils 39 (2003) 1.
10. D.R. McQueen (Ed.), *The Chemistry of Oil explained by Flood Geology*, Institute for Creation Research, Santee, California (1986).
11. A.G. Marshall, R.P. Rodgers, Acc. Chem. Res. 37 (2004) 53.
12. www.bycosin.se/Asph_more.htm.
13. www.astm.org.
14. www.americanenergyinstitutes.org.
15. K.A. Lintelmann, Anal. Chem. 67 (1995) 327R.
16. D. Felscher, Chem. Analityczna 22 (1977) 451.
17. J.G. Reynolds, E.L. Jones, J.A. Bennett, W.R. Biggs, Fuel Sci. Technol. Int. 7 (1989) 625.
18. O.C. Mullins, S. Mitra-Kirtley, Y. Zhu, Appl. Spectrosc. 46 (1992) 1405.
19. N. Pieri, J.P. Planche, J. Kister, Analysis 24 (1996) 113.
20. N. Pieri, J. Kister, L. Quoniam, J.P. Planche, L. Germanaud, Polycycl. Arom. Comp. 9 (1996) 29.
21. Y.-G. Bi, S-F. Yang, Fenxi Ceshi Xuebao 19 (2000) 41.
22. I.N. Evdokimov, N.Y. Eliseev, B.R. Akhmetov. J. Petrol. Science Engin. 37 (2003) 145.
23. I.N. Evdokimov, N.Y. Eliseev, B.R. Akhmetov, Fuel 85 (2006) 1465.
24. I.N. Evdokimov, A.P. Losev, Oil and Gas Business (2007) http://www.ogbus.ru/eng/authors/Evdokimov/Evdokimov_1e.pdf.
25. J. Knotnerus, Erdoel, Kohle, Erdgas, Petrochemie 23 (1970) 341.
26. J.M. Jacobson, M.R. Gray, Fuel 66 (1987) 753.
27. K.H. Michaelian, R.H. Hall, J.T. Bulmer, Can. J. Anal. Sci. Spectrosc. 49 (2004) 43.
28. E. Evans, B. Batts, N. Cant, Fuel 66 (1987) 327.

29. P.W. Yang, H.H. Mantsch, L.S. Kotlyar, J.R. Woods, Energ. Fuels 2 (1988) 26.
30. H. Ganz, W. Kalkreuth, Erdoel & Kohle, Erdgas, Petrochem. 43 (1990) 116.
31. E.A. Cloutis, AOSTRA J. Research 6 (1990) 17.
32. S. Kokot, N. Asker, Proc. 8th Int. Conf. Fourier Transform Spectrosc., SPIE- Int. Soc.Opt. Engin. 1575 (1992) 495.
33. S. Kokot, N. Asker, Proc. 6th Int. Conf. Near Infr. Spectrosc. 22 (1995) 1994.
34. N. Pieri, J. Kister, L. Quoniam, J.P. Planche, L. Germanaud, Polycycl. Arom. Comp. 9 (1996) 29.
35. H. Chung, M.-S. Ku, J.-S. Lee, Vibr. Spectrosc. 20 (1999) 155.
36. M.S. Akhlaq, J. Petr. Sci. Engin. 22 (1999) 229.
37. O.Y. Begak, A.M. Syroezhko, V.V. Fedorov, M.N. Malechkina, M.B. Nassonova, Russ. J. Appl. Chem. 75 (2002) 1173.
38. K.H. Michaelian, S. Zhang, R.H. Hall, J.T. Bulmer, Can. J. Anal. Sci. Spectrosc. 46 (2001) 10.
39. B. Pavoni, N. Raido, R. Plassa, S. Frignani, Ann. Chim. 94 (2004) 521.
40. W.I. Friesen, K.H. Michaelian, Y. Long, T. Dabros, Am. Chem. Soc., Div. Fuel Chem. 50 (2005) 228.
41. T.A. Adedosu, O.O. Sonibare, J. Appl. Sci. 5 (2005) 906.
42. H. Yuan, X. Chu, H. Li, Y. Xu, Fuel 85 (2006) 1720.
43. K.J. Johnson, R.E. Morris, S.L. Rose-Pehrsson, Energ. Fuels 20 (2006) 727.
44. K.H. Michaelian, R.H. Hall, K.I. Kenny, Spectrochim. Acta A 64 (2006) 703.
45. R.R. Coelho, I. Hovell, M.B. de Mello Monte, A. Middea, A. Lopes de Souza, Fuel Process. Technol. 87 (2006) 325.
46. Y. Hongfu, C. Xiaoli, L. Haoran, X. Yupeng, Fuel 85 (2006) 1720.
47. J.M. Brown, C.J. Chrostowsky, PCT Int. Appl. (2006) 2006023800.
48. L.A. Tomskaya, N.P. Makarova, V.D. Ryabov, Chem. Technol. Fuels and Oils 44 (2008) 280.
49. P. Doumenq, M. Guiliano, G. Mille, J. Kister, Anal. Chim. Acta 242 (1991) 137.
50. Z. Sikora, T. Wojdala, Nafta--Gaz 50 (1994) 298.
51. A. Espinosa, A. Martens, G. Ventron, C.D. Lambert, A. Pasquier, Eur. Pat. Appl. (1989) 305090.
52. D.A. LeFebre, L.M. Lane, US Patent (1995) 5430295.
53. S. Bages, B. Descales, D. Lambert, J.-R. Llinas, A. Martens, S. Osta, M. Sanchez, PCT Int. Appl. (1996) 9611399.
54. B. Descales, D. Lambert, J.-R. Llinas, A. Martens, S. Osta, M. Sanchez, S. Bages, U.S. Patent (2000) 5712797.
55. Z. Sikora, W. Salacki, Petrol. Coal 38 (1996) 65.
56. R. Van Doorn, F.H. Hooyman-Spaargaren, R.J. Neugebauer, L.M. Smeets, PCT Int. Appl. (1997) 9714953.
57. J. Kelly, D. Lambert, A. Martens, Eur. Pat. Appl. (1998) 0859236.

58. M. Blanco, S. Maspoch, I. Villarroya, X. Peralta, J.M. González, J. Torres, Appl. Spectr. 55 (2001) 834.
59. S. Macho, M.S. Larrechi, Trends Anal. Chem. 21 (2002) 799.
60. S. Satya, R.M. Roehner, M.D. Deo, F.V. Hanson, Energ. Fuels 21 (2007) 998.
61. R.J. Mikula, V.A. Munoz, N. Wang, B. Bjornson, D. Cox, B. Moisan, K. Wiwchar, J. Can. Petr. Techn. 42 (2003) 50.
62. H. Chung, Appl. Spectrosc. Rev. 42 (2007) 251.
63. M.V. Reboucas, E.C. Santos, F.S.V. Vieira, Vibr. Spectrosc. 44 (2007) 187.
64. M. Ahmadjian, C.W. Brown Anal. Chem. 48 (1976) 1257.
65. J. Jehlička, O. Urban, J. Pokorný, Spectrochim. Acta A 59 (2003) 2341.
66. M. He, X. Lu, S. Yao, Y. Liu, W. Fan, Dizhi Keji Qingbao 24 (2005) 67.
67. K.H. Michaelian, S.L. Zhang, R.H. Hall, J.T. Bulmer, Spectrochim. Acta A 57 (2001) 73.
68. K.H. Michaelian, R.H. Hall, J.T. Bulmer, Can. J. Anal. Sci. Spectrosc. 49 (2004) 43.
69. L.C.T. Shoute, K.J. Schmidt, R.H. Hall, M.A. Webb, S. Rifai, P. Abel, P.H. Arboleda, A. Savage, J.T. Bulmer, G.R. Loppnow, Appl. Spectrosc. 56 (2002) 1308.
70. Y. Bouhadda, D. Bormann, E.Y. Sheu, D. Bendedouch, A. Krallafa, M. Daaou, Fuel 86 (2007) 1855.
71. Y. Bouhadda, T. Fergoug, E.Y. Sheu, D. Bendedouch, A. Krallafa, D. Bormann, A. Boubguira, Fuel 87 (2008) 3481.
72. J.G. Speight, Fuel 49 (1970) 76.
73. J. Christopher, A.S. Sarpal, G.S, Kapur, A. Krishna, B.R. Tyagi, M.V. Jain, S.K. Jain, A.K. Bhatnagar, Fuel 75 (1996) 999.
74. F.Ali, Z.H. Khan, N. Ghaloum, Energ. Fuels 18 (2004) 1798.
75. A. Rafenomanantsoa, D. Nicole, P. Rubini, J.-C. Lauer, Energ. Fuels 8 (1994) 618.
76. C. Giavarini, C. Vecchi, Fuel 66 (1987) 868.
77. M.H.B. Hayes, M. Stacey, J. Standley, Fuel 51 (1972) 27.
78. M.U. Hasan, M.F. Ali, A. Bukhari, Fuel 62 (1983) 518.
79. M.U. Hasan, M.F. Ali, A. Bukhari, Fuel 64 (1985) 839.
80. J. Thiel, M.R. Gray, AOSTRA J. Research 4 (1988) 63.
81. C.H.Tsai, M.D. Deo, F.V. Hanson, A.G. Oblad, Fuel Sci. Technol. Int. 10 (1992) 1437.
82. J.M. Jacobson, M.R. Gray, Fuel 66 (1987) 753.
83. L. Michon, D. Martin, D., J.-P. Planche, B. Hanquet, Fuel 76 (1997) 9.
84. H.A. Akrami, M.F. Yardim, A. Akar, E. Ekinci, Fuel 76 (1997) 1389.
85. J. Bryan, K. Mirotchnik, A. Kantzas, J. Can. Petrol. Technol. 42 (2003) 29.
86. B. Behera, S.S. Ray, I.D. Singh, Fuel 87 (2008) 2322.
87. E. Durand, M. Clemancey, A.A. Quoineaud, J. Verstraete, D. Espinat, J.M. Lancellin, Energ. Fuels 22 (2008) 2604.
88. T.J. Morgan, A. George, D.B. Davis, A.A. Herod, R. Kandiyoti, Energ. Fuels 22 (2008) 1824.

89. H. Swierenga, F. Wulfert, O.E. de Noord, A.P. de Weijer, A.K. Smilde, L.M.C. Buydens, Anal. Chim. Acta 411 (2000) 121.
90. T. Zhang, Y. Liang, B. Li, H. Cui, F. Gong, Anal. Sci. 20 (2000) 717.
91. Q. Abbas, N. Dupuy, C. Rebufa, L.Vrielynck, J. Kister, A. Permanyer, Appl. Spectrosc. 60 (2006) 304.
92. H. Chung, S. Cho, Y. Toyoda, K. Nakano, M. Maeda, Analyst 131 (2006) 684.
93. W. Tuchler, A. Ecker, G.K. Schreferl, Petrol. Coal 42 (2000) 56.
94. L.S.G. Teixeira, F.S. Olieveira, H.C. dos Santos, P.W.L. Cordeiro, S.Q. Almeida, Fuel 87 (2008) 346.
95. F.A. Honorato, B. de Barros Neto, M.F. Pimentel, L. Stragevitsch, R.K.H. Galvao, Fuel 87 (2008) 3706.
96. G. Genov, E. Nodland, B. Barman Skaare, T. Barth, Org. Geochem. 39 (2008) 1229.
97. L. Michon, B. Hanquet, B. Diawara, D. Martin, J.-P. Planche, Energ. Fuels 11 (1997) 1188.

Chapter 2

Spectroscopic Characterization and Chemometric Classification of Crude Oils to Predict Bitumen Quality

Abstract

Research has been carried out to determine the feasibility of chemometric modeling of Raman, UV-Vis, ^{1}H-NMR, ^{13}C-NMR, IR and NIR spectra of crude oils to predict the long residue (LR) and short residue (SR) properties of these samples. The study has been carried out on 47 crude oils and 4 blends, representing a large variety in physical and chemical properties. Raman spectroscopy turned out to be not useful due to fluorescence and self absorption problems. For the remaining techniques, the spectra of 22 representative samples were selected by Principal Component Analysis (PCA) and used to build prediction models for 7 different LR properties. The results obtained for the Partial Least Squares (PLS) regression models based on IR and NIR spectra appeared to be significantly better than for those derived from the UV-Vis and NMR data. For that reason research was further focused on IR and NIR spectra. From the set of 51 samples, 28 were selected by PCA and used for calibration. The remaining 23 samples were used as test set to validate the PLS models. As a result, a novel method is described to predict LR properties as well as SR properties at different flashing temperatures based on the IR spectrum of a crude oil measured at room temperature. The resulting method has been patented (WO/2008/135411) and demonstrates that this integrated approach offers a fast and viable alternative for the currently applied elaborative ASTM (American Society for Testing and Materials) and IP (Institute of Petroleum) methods. In particular, IR spectra were found to be useful input for the prediction of different LR properties. Root mean square error of prediction values of the same order of magnitude as the reproducibility values of the ASTM methods were obtained for yield long on crude (YLC), density (D_{LR}), viscosity (V_{LR}), the sulfur content (S) and Pour Point (PP) , while the ability to predict the carbon residue (CR) was found to be useful for indicative purposes. The prediction of SR properties is also promising. Modeling of the IR spectra, and to a less extent, the NIR spectra as a function of the Atmospheric equivalent Flash Temperature (AFT) was successful for the prediction of the short residue properties density (D_{SR}) and viscosity (V_{SR}). Similar results were obtained from the models to predict SR properties as a function of the yield short on crude (YSC) values. Finally, it

was concluded that the applied protocol including sample pre-treatment and instrumental measurement is highly reproducible and instrument and accessory independent.

2.1. Introduction

Bitumen is a complex mixture of hydrocarbons and hydrocarbon derivatives, including aliphatic, olefinic and aromatic compounds. It is usually pictured as a colloidal solution of high molecular carbon-hydrogens (so-called asphaltenes) finely dispersed in a mixture of resins and oils [1, 2]. Bitumen is commonly known as an important component for road construction, but is also used for roofing, waterproofing adhesive and sealing applications. A substantial amount of bitumen is found in nature, *i.e.* buried in sand [3] but, more generally, it is obtained as a residue of crude oil after atmospheric and vacuum distillation in the refinery. As such, it is an important product of the oil industry and of considerable economical value. Therefore, oil companies are very careful to fully assess the bitumen from a crude oil to ensure it meets the market requirements, before proceeding to approving for normal refinery production.

At present a crude oil is assessed by laboratory preparation of actual bitumen samples, which are then passed through a range of further tests to confirm meeting the required quality. The assessment of these chemical and physical properties is carried out by standard methods developed by the ASTM [4] and the Energy Institute [5], formerly known as the Institute of Petroleum (IP). These methods are rather slow, elaborative and expensive, requiring large sample amounts of the crude to be sent to the laboratory. As a fast and viable alternative, spectroscopic methods, such as nuclear magnetic resonance (NMR) [6-17], Raman [18-21], infrared (IR) [10, 21-40], near infrared (NIR) [41-53] and fluorescence [54], have been proposed since the corresponding spectra reflect the complete molecular composition of the crude oil [55]. Particularly, when combined with chemometric modeling techniques [31, 34, 35, 39, 44, 45, 48-50, 56-60], it might be possible to classify crude oils in terms of suitability for bitumen manufacture from their spectra. Hence, it is a logical step to explore the potential of various molecular spectroscopic techniques in combination with chemometric modeling techniques to classify crude oils in terms of suitability for bitumen manufacture and to predict bitumen properties as a function of the flash temperature directly from crude oil information.

This approach should link the more fundamental chemical composition of crude oils to short residue properties and thus, to their bitumen quality. For this purpose six different spectroscopic techniques, i.e., Raman, UV-Vis, ^{1}H-NMR, ^{13}C-NMR, IR and NIR have been selected to predict bitumen properties in terms of so-called long residue (LR) and short residue (SR) properties. LR properties are commonly defined as a series of physical and chemical properties of the residues flashed at 370℃, while SR properties are described as physicochemical properties of the residues flashed at higher temperatures, *i.e.* 400-590℃ Maxwell-Bonnell. The approach of this study should link the more fundamental chemical composition of crude oils to SR properties and thus, to their bitumen quality. A novel method is described to predict SR properties of a crude oil at different flashing temperatures based on the IR spectrum measured at room temperature.

2.2. General methodology

2.*2.1. Experimental protocols*

As a first step, methods for sample storage, sample pre-treatment and sample preparation were developed in experimental protocols to assure the acquisition of reproducible, high quality spectra. Although the high viscosity of several of the crude oils would make it reasonable to perform the measurements at elevated temperatures, for practical reasons, it was chosen to carry out all measurements at room temperature (20℃). As some samples contain considerable light-ends that flash off at low temperatures, spectra were acquired in closed cells. Next, experimental and instrumental settings were optimized to obtain sufficient spectral quality in terms of signal to noise ratio for each of the techniques. The final settings have been summarized in protocols and applied throughout the study.

For IR, 22 samples have been analyzed 3 times with time gaps of a few weeks to optimize the experimental conditions and to develop the final protocol. All other spectra were measured in accordance with this protocol. In a separate experiment, the effect of evaporation of volatile compounds from the crude oil during IR-spectroscopic measurement was tested by comparison of spectra recorded with the closed sample cell, with spectra taken with an open cell as a function of time. Finally, measurements were carried out to study the effect of time during spectra acquisition with a closed cell to determine whether precipitation or a change in homogeneity occurred during spectrum acquisition.

PLS regression was applied to correlate the properties of long residues (LR) and short residues (SR) with the IR and NIR spectra of the crude oils. First, properties of residues flashed at 370℃ (i.e. long residues) have been modeled. The following LR properties were studied: the long residue yield-on-crude (YLC), density (D_{LR}), viscosity (V_{LR}), sulfur content (S), pour point (PP), asphaltenes (Asph) and carbon residue (CR). As input values for these properties, the data obtained with the following ASTM methods were used: D 2892 (YLC), D 4052 (D_{LR}), D 445 (V_{LR}), D2622 (S), D 97 (PP) and D 4530 (CR), while the values for Asph were determined by method IP 143.

In a second step, research was carried out to determine whether the short residue (bitumen) properties as a function of the Atmospheric equivalent Flash Temperature (AFT) could be related to the spectral data of the crude oils in order to predict bitumen quality. As a starting point for the prediction of the short residue properties, the parameters penetration (P), softening point or ring and ball (R&B), viscosity (V_{SR}) and density (D_{SR}) were considered as the main properties. As input values for the SR properties, the data obtained with the ASTM methods D 5 (P), D 36 (R&B), D 4052 (D_{SR}) and D 445 (V_{SR}) were used.

2.2.2. Materials

Spectra were acquired from a set of 47 crude oil samples and 4 blends. As can be seen from Table 1, the set covers a wide range of LR properties. SR properties were known for smaller sets of samples. To reduce evaporation of volatile constituents from the original samples, all crude oils were stored in the refrigerator at 3℃. Prior to analysis, samples were taken out of the refrigerator 8 h before measurement with caps closed to prevent condensation of moisture. Furthermore, samples were homogenized at least 1 h before measurement by shaking the sample every 10 min.

From the total set, the samples C1-C22 were selected to build prediction models for Raman, UV-Vis, ^{1}H-NMR, ^{13}C-NMR, IR and NIR spectra. After evaluation of the results obtained for the different techniques, samples C1-C28 were used for calibration purposes of the LR properties from the IR and NIR spectra and the remaining 23 samples V1-V23 for validation of the models obtained. Modeling to predict SR properties was carried out on smaller sets since these data were not known for all compounds.

Table 1. Crude oil samples used for calibration (C1-C28) and validation (V1-V23) with corresponding long residue (LR) properties. YLC: yield long-on-crude; D_{LR}: density; V_{LR}: viscosity; S: sulfur content; PP: pour point: Asph: asphaltenes content: CR: carbon residue.

	YLC (wt %)	D_{LR} (g/mol)	V_{LR} (-)	S (wt %)	PP (℃)	Asph (wt %)	CR (wt %)
C1	34.7	0.8889	29.5	0.72	42	1.2	3.9
C2	55.7	0.9507	37.2	4.75	30	9.5	14.0
C3	23.9	0.8839	28.9	1.08	30	1.0	3.3
C4	24.1	0.8827	28.8	1.11	36	0.7	3.3
C5	64.1	0.8665	30.8	0.14	48	0.1	4.4
C6	40.8	0.9003	29.1	0.70	36	1.6	6.4
C7	61.8	0.9504	37.9	1.17	15	1.2	9.9
C8	35.2	0.8962	34.6	0.56	30	0.6	4.1
C9	47.1	0.9188	33.3	0.84	3	0.2	4.7
C10	49.0	0.9488	36.8	3.22	21	5.8	12.6
C11	58.0	0.9273	36.3	2.19	24	3.2	9.6
C12	43.2	0.8660	29.2	0.25	42	0.2	4.4
C13	80.7	0.9820	41.8	3.05	45	8.0	14.0
C14	69.3	0.9086	36.2	0.99	48	0.2	9.7
C15	56.8	0.9800	39.3	5.42	27	12.2	17.7
C16	79.7	0.9842	41.9	3.14	39	8.0	14.2
C17	49.4	0.9178	33.8	2.05	18	2.2	7.7
C18	47.4	0.9236	33.5	2.15	27	2.7	8.0
C19	32.7	0.9583	34.5	4.87	27	2.0	10.7
C20	78.0	0.9429	39.8	0.30	30	4.5	12.6
C21	39.1	0.9313	35.0	0.43	30	0.4	6.0
C22	80.2	0.9958	44.1	3.88	48	12.0	16.8
C23	36.9	0.8865	30.9	0.81	39	1.3	6.3

C24	49.1	0.9473	36.4	3.14	18	6.3	12.4
C25	45.7	0.9332	33.2	3.23	6	3.6	10.4
C26	65.5	0.9771	41.9	4.31	39	15.4	19.2
C27	66.9	0.9344	35.7	1.47	9	2.0	8.7
C28	55.6	0.9573	38.1	4.47	21	8.4	15.3
V1	46.8	0.9278	32.7	2.85	39	5.3	11.0
V2	46.1	0.9299	33.3	3.17	39	3.2	9.8
V3	41.8	0.8970	31.5	0.42	39	0.1	5.0
V4	49.9	0.9208	33.0	1.54	36	6.0	10.8
V5	49.7	0.9444	35.3	4.30	9	4.2	11.7
V6	58.7	0.9510	37.5	3.39	33	9.0	14.1
V7	31.1	0.8822	31.9	0.46	36	0.2	4.3
V8	61.7	0.9470	36.4	4.26	18	3.5	15.0
V9	49.9	0.9471	35.9	3.76	21	5.6	12.2
V10	51.6	0.9136	34.0	2.82	48	4.5	9.1
V11	39.0	0.8969	30.1	0.48	36	3.3	6.8
V12	51.4	0.9457	35.7	4.12	18	5.4	12.3
V13	57.1	0.9546	38.3	2.52	15	8.6	13.8
V14	66.7	0.9696	43.3	2.50	54	15.3	19.3
V15	60.1	0.9682	38.6	5.09	18	9.9	16.1
V16	41.4	0.9080	31.3	1.13	27	2.2	6.4
V17	42.5	0.9325	33.6	3.09	30	2.2	8.3
V18	48.1	0.9229	32.8	2.03	24	2.7	7.7
V19	33.9	0.9533	35.1	4.55	24	2.1	10.4
V20	45.2	0.9382	34.9	2.88	15	5.2	11.3
V21	45.5	0.9376	35.3	2.76	21	5.6	10.9
V22	43.7	0.9424	35.5	3.05	15	4.7	11.4
V23	46.1	0.9392	35.4	2.91	21	5.4	11.8

2.2.3. Spectroscopic methods

Raman measurements were carried out on three different instruments: 1) a Kaiser RXN spectrometer equipped with a 70 mW 532 nm diode laser for excitation, a holographic grating for dispersion and a Peltier cooled Andor CCD camera for detection, 2) a Kaiser RXN spectrometer equipped with a 60 mW 785 nm diode laser a Peltier cooled Andor CCD camera for detection, and 3) a Perkin-Elmer System 2000 NIR FT-Raman with a 1064 nm laser line for excitation and a room temperature operated InGaAs detector. Spectra were recorded in glass vials (Spectra-Tech) at room temperature. Detector exposure time and scan accumulations were chosen, depending on the scattering properties of the samples.

For UV-Vis spectroscopy the crude oil samples were diluted in a ratio of 1:4000 in CH_2Cl_2 as the solvent in order to obtain sufficient transmission. Spectra were recorded with a Cary 50 spectrometer (Varian) equipped with a probe head that was connected to the instrument via optical fibers. The effective path length of the probe was 2 mm. Spectra were recorded from 1100 to 200 nm at a data point interval of 0.15 nm and at a scan rate of 360 nm/min.

^{1}H-NMR spectra were recorded at room temperature on a Bruker AC300 FT-NMR, set to ^{1}H 300 MHz using WinNMR 6.0 software for control and acquisition. The instrumental settings were as follows: SW: 4505.4 Hz; FW: 5700 Hz; SI and TD: 32K data points; O1: 6100; O2: not used; LB: 0.3 Hz and SR: 3367.25. For each sample, a fresh homogeneous solution was prepared of 50 mg crude oil in 0.5 ml $CDCl_3$. The solutions were spiked with tetramethylsilane (TMS). Solutions were analyzed in sealed 5 mm quartz NMR tubes with a spin speed of 20 Hz.

^{13}C-NMR spectra were recorded at room temperature on a Bruker AC300 FT-NMR, set to ^{13}C at 75 MHz using WinNMR 6.0 software for control and acquisition. The instrumental settings were as follows: SW: 17857 Hz; FW: 20000 Hz; SI and TD: 32K data points; O1: 6100; O2: 4600; LB: 2 Hz and SR: -1405. For each sample, fresh homogeneous solutions were prepared of 300 mg crude oil in 0.5 ml $CDCl_3$. The solutions were spiked with tetramethylsilane (TMS) and analyzed in sealed 5 mm quartz NMR tubes with a spin speed of 20 Hz.

IR measurements were carried out at room temperature on two different instruments: a Bruker Tensor-27 and a Perkin-Elmer 2000 Fourier transform spectrometer. Both instruments were equipped with a DTGS detector and the sample compartments were flushed with dry air to reduce interference of H_2O and CO_2. Spectra were recorded either with a horizontal ATR accessory (MIRacle^{TM}, Pike Technologies) with a ZnSe/Diamond crystal as the internal reflection element (Perkin-Elmer 2000) or with a horizontal ATR accessory (FastIR^{TM}, Harrick Scientific Products) with a ZnSe crystal as the internal reflection element (Tensor-27 instrument). The spectral resolution was 4 cm^{-1} for all spectra and 50 scans were accumulated with medium apodization for each spectrum. ATR-intensity correction was not applied.

NIR spectra were recorded at room temperature on two different instruments: a Bruker Tensor-27 and a Perkin-Elmer 2000 Fourier transform infrared spectrometer. Both instruments were equipped with a (NIR) DTGS detector and the sample compartments were flushed with dry air to reduce interference of H_2O and CO_2. The spectral resolution was 16 cm^{-1} and 100 scans were accumulated with medium apodization for each spectrum. Liquid samples were scanned in a fixed 0.1 mm cell equipped with CaF_2 windows. Highly viscous samples were analyzed in a demountable cell with KBr windows and a path length of approximately 0.1 mm.

2.2.4. Data analysis

Modeling was performed using the PLS Toolbox (Eigenvector Research, Inc.) for MatLab (The MathWorks, Inc.). Prior to modeling, the IR and NIR spectra were pre-processed by first taking the 1^{st} derivative (25 pt. Savitzky-Golay smoothing), followed by “Multiplicative Signal Correction” (MSC) and Mean Centering of the data. Pre-processing parameters were chosen based on prior knowledge for each spectroscopic technique combined with trial and error. For the IR spectra, the regions from 3200-2500 cm^{-1} and 1800-650 cm^{-1} were used as input, whereas for NIR the spectra from 7000 cm^{-1} to 4000 cm^{-1} (1429-2500 nm) were taken. The obtained root-mean-square error of cross validation values (RMSECV) were based on ”leave-one-out” (LOO) cross validation (CV). The number of latent variables (LV), used for a PLS model, was based on the minimum value of the RMSECV. In case the minimum was reached at only one LV in the PLS cross validation, the model was build with three LV’s in order to get a more realistic PLS model.

First, modeling for the LR properties was carried out on the spectra (IR and NIR) of the set of 51 samples (47 crudes, 4 blends). This set was divided into a group of 28 spectra for calibration and 23 spectra for validation. The calibration set was based on the results of a Principal Component Analysis (PCA) of the total collection without the 4 blends to obtain a representative subset. The Mahalanobis distance based on the first 6 scores was used to select 21 extreme samples with largest distance and 2 samples with the smallest distance to the center of the scores. The 5 additional samples of the calibration set were chosen randomly and the blends were added to the test set. Secondly, the 7 LR properties were modeled with PLS-1 with the described calibration and validation set of respectively 28 and 23 spectra (IR and NIR).

Next, PLS models (SIMPLS algorithm, PLS-1) were created from the IR spectra of 43 crude oils for which SR properties were known. PLS prediction models were made for the SR properties Penetration (P), ring-and-ball (R&B), density (D_{SR}) and viscosity (V_{SR}) at various flash temperatures. The corresponding data sets consisted of 33 (P), 37 (R&B), 40 (D_{SR}) and 38 (V_{SR}) IR spectra with 99, 119, 130 and 118 SR properties, respectively. Because of the smaller number of spectra for which SR properties were known in combination with the model complexity, LOO cross validation was applied instead of creating an independent validation set. In addition, modeling of the spectra was carried out after randomizing SR properties, as a test for the validity of the models obtained [61]. Finally, the SR properties were further determined at various AFT and this AFT-value was used as discriminator.

The process of the PCA as applied is schematically depicted in Fig. 1.

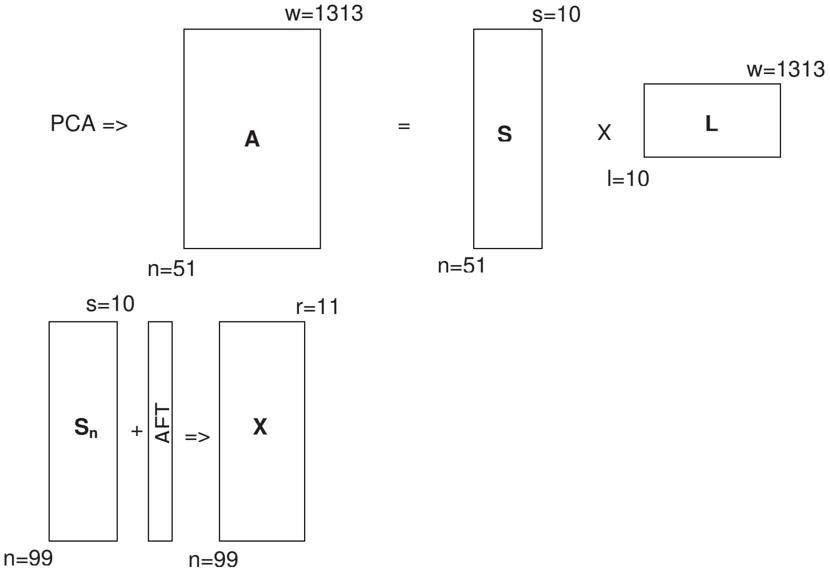


Fig. 1. The schematic representation of the creation of the **X**-matrix using PCA. **X** is used as input for the PLS modeling of the crude SR-property P at the available AFT values.

The data matrix **X** was created by PCA on the available IR spectra (**A**) after pre-processing (1st derivative and 25 pt. Savitzky-Golay smoothing, Multiplicative Signal Correction and Mean Centering). The first 10 scores of this PCA model (**S**) were used to describe the spectra for which the SR properties were available resulting in a score matrix $\mathbf{S_n}$. The discriminating AFT values were then added to $\mathbf{S_n}$ resulting in 11 variables in the new data matrix **X**, which were variance scaled for each crude. Next, the corresponding SR properties were used as the **Y** matrix in the PLS regression. The same procedure was followed using the "yield short on crude" (YSC) instead of the AFT values as additional input of the data matrix. PCA modeling was used to examine the differences between the different IR methods that were used.

2.3. Results and discussion

2.3.1. Raman spectroscopy

The Raman spectra of all compounds appeared not to be useful for modeling purposes. In all cases, excitation with both the 532 nm and the 785 nm laser line produced spectra with a very high fluorescence level, which completely obscured the much weaker Raman scattering. Similarly, heating effects and self-absorption prevented the acquisition of useful Raman spectra when using the 1064 nm laser line for excitation. UV laser excitation might be a way to reduce these interfering phenomena, but UV/Raman facilities are not available at the Debije Institute for Nanomaterials Science of Utrecht University. For that reason further studies to the potentials of this technique were not carried out.

2.3.2 UV-Vis spectroscopy

The UV-Vis spectra of the pure crude oils appeared to be too black to be analysed directly. For that reason dilute solutions were prepared. The suitability of hexane, acetone, CH_2Cl_2 and CCl_4 as the solvent were tested. Both CH_2Cl_2 and CCl_4 were found to virtually dissolve all constituents including asphaltenes. However, for environmental and safety reasons,

CH_2Cl_2 was selected for further application. A dilution factor of 1:4000 appeared to be useful to obtain spectra of sufficient signal to noise ratio. As an example of the quality of the spectra obtained, the data of three crudes are shown in Fig. 2. All compounds showed a highly similar pattern with two strong bands at around 250 and 300 nm. Since this is the absorption region of aromatic compounds, it may reflect information on the amount of asphaltenes in the crude oil.

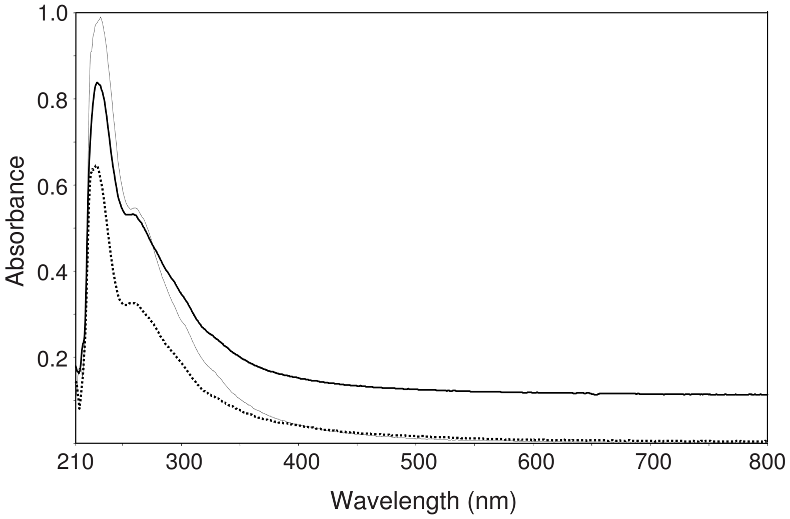


Fig. 2. UV-Vis spectra of 3 different crude oils (V2, V4 and C5) dissolved in CH_2Cl_2.

2.3.3. ^{1}H-NMR spectrometry

Characteristic ^{1}H-NMR spectra of 2 crude oils, more specifically sample V5 and V9, are shown in Fig. 3. Dilution factors of 5-20 times appeared to be necessary to obtain useful data. The intense shifts around 1 and 1.5 ppm are typical for saturated carbon-hydrogen groups, e.g., methyl and methylene groups from alkanes, whereas the much weaker peaks around 7 ppm reflects the presence of aromatic compounds [62]. These peaks dominate the spectra of all ^{1}H-NMR spectra. However, detailed analysis of the weaker peak pattern in exploded view revealed small differences between the different samples.

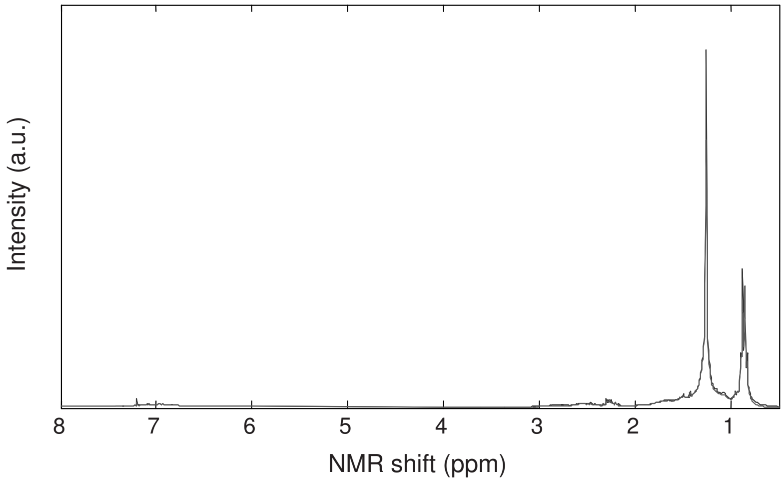


Fig. 3. ^{1}H-NMR spectra of 2 different crude oils (V5 and V9).

2.3.4. ^{13}C-NMR spectrometry

Characteristic ^{13}C-NMR spectra of the crude oils C1 and C5 are shown in Fig. 4. At first glance, the spectra seem very much alike but closer examination reveals some significant differences. As a consequence, ^{13}C spectra seem more promising than the corresponding ^{1}H spectra. A complicating factor of both types of NMR spectra with respect to chemometrical modeling is the limited data point reproducibility. This problem is inherent to the calibration of NMR spectrometers and negatively affects modeling.

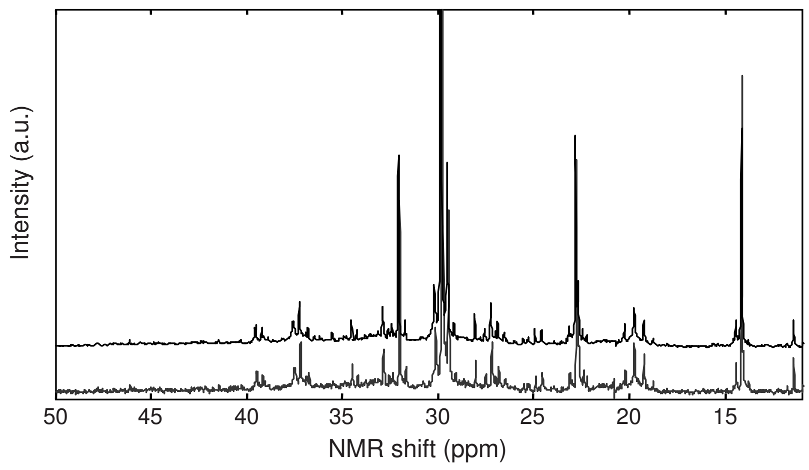


Fig. 4. ^{13}C-NMR spectra of crude oils C5 (upper trace) and C8 (lower trace). Offset of the baseline has been applied to allow for better comparison.

2.3.5. Infrared spectroscopy

As illustrated by the IR spectra of sample C1 in Fig. 5, crude oils and blends are characterized by strong absorptions around 2900 and 1400 cm^{-1}, originating from the C-H stretching and bending vibrations, respectively. Furthermore, the spectra show bands in the range 900-600 cm^{-1} due to aromatic ring deformation and out-of-plane vibrations and long alkyl chain CH_2 rocking modes. However, as shown in Fig. 6, the IR spectra of crude oils, recorded in an open cell as a function of time, may show considerable changes, particularly in the fingerprint region as result of the evaporation of volatile constituents during measurement. For that reason, only closed sample cell data were used for further modeling experiments. Under these conditions the spectra were found to be highly reproducible, as can be seen in Fig. 5.

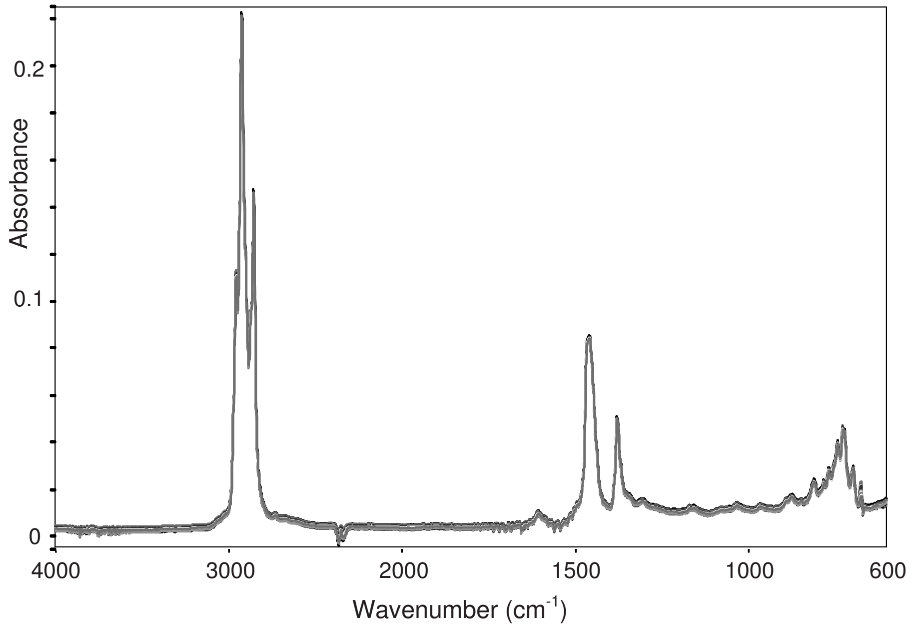


Fig. 5. Overlay of 16 subsequently recorded IR absorbance spectra of sample C1 in a closed ATR sample cell recorded over a 1 h time interval.

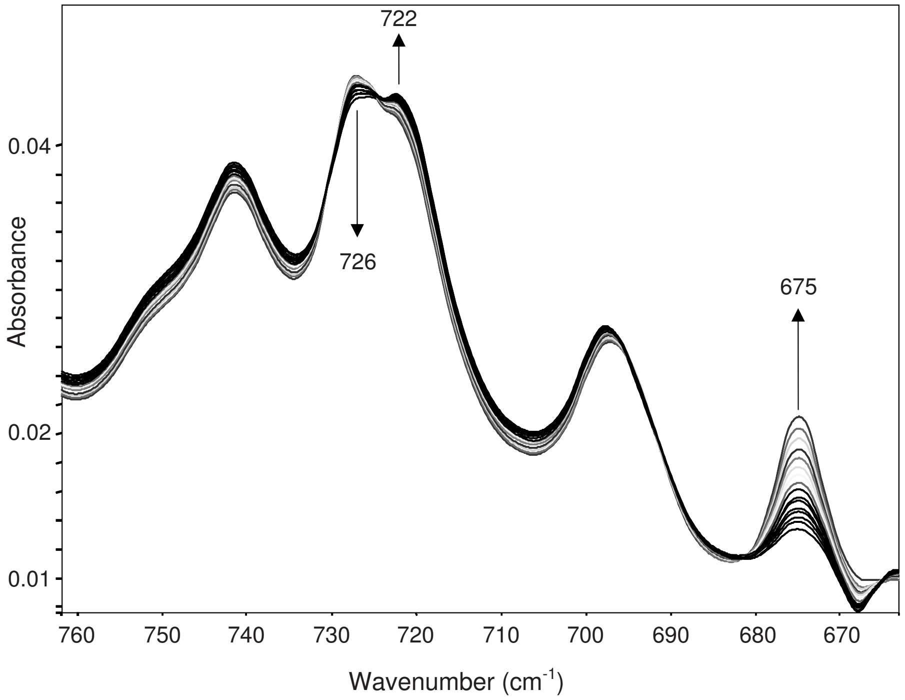


Fig. 6. Effect of evaporation of volatile components on the ATR-IR spectra of sample C1 when using an open sample cell.

2.3.6. Near infrared spectroscopy

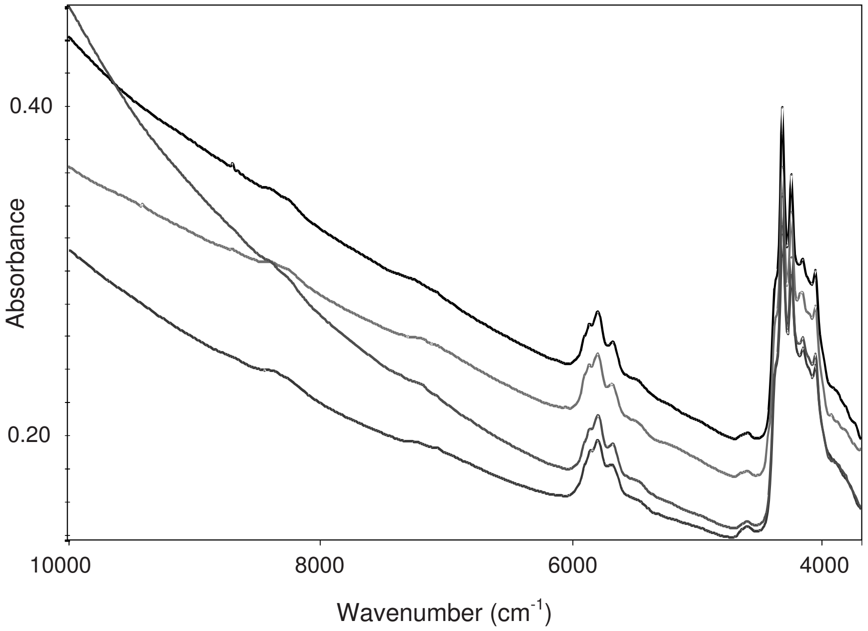


Fig. 7. NIR spectra of 4 different crude oil samples (C9, C12, C13 and C22).

As shown in Fig. 7, the NIR spectra of the crude oils are characterized by absorption bands related to overtones/combinations of –C-H stretching and bending vibrations around 4100 cm^{-1} (2439 nm) and 5500 cm^{-1} (1818 nm). Also weaker bands can be observed at 7600 cm^{-1} (1316 nm) and 8300 cm^{-1} (1205 nm). Furthermore, considerable baseline differences are present, which can be attributed to different scattering effects and electronic transitions at higher wavenumbers. In accordance with literature [40], the effect was found to be the strongest for the dark and viscous heavy oils, such as C2, C10, C15, C22, V14 and V15, which usually contain relatively high amounts of aromatic compounds including the solid asphaltenes. It should be noted that the highly viscous samples require the application of a demountable cell of which the path length cannot be adjusted with high accuracy, thus limiting a fair mutual comparison. This problem was partially circumvented in the modeling phase by normalizing the spectra on the strongest band present, i.e., around 4100 cm^{-1} (2439 nm) or by applying MSC.

2.3.7. Data analysis of long residue properties

The calibration sets of 22 UV-Vis, ^{1}H NMR, ^{13}C NMR, IR and NIR spectra were used to build PLS models for each of the 7 LR properties. The results are summarized in Tables 2-6, showing the mean value for each of the properties, the RMSECV based on the PLS model that was built, and the number of latent variables (LV's) used. For all models, the RMSECV is based on LOO cross validation. All spectra were pre-processed before calibration was carried out. Pre-processing parameters were chosen based on prior knowledge of the spectroscopic technique combined with trial and error. The number of LV's or principal components, used for a PLS model, was based on the minimum value of the RMSECV. In case the minimum was reached at only one LV in the PLS cross validation, the model was built with three LV's in order to get a more realistic PLS model.

Table 2. Results of a PLS-regression of the 22 UV-Vis spectra versus the LR properties.

LR property	mean value	RMSECV	used LV's
YLC (wt %)	49.7	6.7	2
D_{LR} (g/ml)	0.926	0.023	2
V_{LR} (-)	34.4	2.8	3
S (wt %)	2.08	0.57	2
PP (℃)	27.3	12.8	2
Asph (wt %)	3.4	2.5	2
CR (wt %)	9.2	1.7	2

Table 3. Results of a PLS-regression of the 22 ^{1}H NMR spectra versus the LR properties.

LR property	mean value	RMSECV	used LV's
YLC (wt %)	49.7	11.7	4
D_{LR} (g/ml)	0.926	0.025	3
V_{LR} (-)	34.4	2.5	3
S (wt %)	2.08	1.58	4
PP (℃)	27.3	11.1	4
Asph (wt %)	3.4	3.4	3
CR (wt %)	9.2	3.3	3

Table 4. Results of a PLS-regression of the 22 ^{13}C NMR spectra versus the LR properties.

LR property	mean value	RMSECV	used LV's
YLC (wt %)	49.7	7.9	3
D_{LR} (g/ml)	0.926	0.020	4
V_{LR} (-)	34.4	2.5	2
S (wt %)	2.08	1.09	3
PP (°C)	27.3	9.2	4
Asph (wt %)	3.4	3.4	3
CR (wt %)	9.2	2.6	5

Table 5. Results of a PLS-regression of the 44 (2 replicates) ATR-IR spectra versus the LR properties.

LR property	mean value	RMSECV	used LV’s
YLC (wt %)	49.7	2.1	6
D_{LR} (g/ml)	0.926	0.006	5
V_{LR} (-)	34.4	0.9	6
S (wt %)	2.08	0.29	6
PP (°C)	27.3	6.0	6
Asph (wt %)	3.4	1.9	6
CR (wt %)	9.2	1.1	5

Table 6. Results of a PLS-regression of the 22 NIR spectra versus the LR properties.

LR property	mean value	RMSECV	used LV's
YLC (wt %)	49.7	2.5	5
D_{LR} (g/ml)	0.926	0.009	4
V_{LR} (-)	34.4	1.3	3
S (wt %)	2.08	0.64	4
PP (°C)	27.3	11.3	3
Asph (wt %)	3.4	2.5	2
CR (wt %)	9.2	1.2	3

As appears from Tables 2-6, the results of modeling based on IR and NIR spectra are in all cases better than those built on UV-Vis and ^{1}H- and ^{13}C-NMR spectra. Compared to IR, the results of modeling for the NIR spectra were slightly worse as expressed by the larger RMSECV values. This is attributed to the stronger baseline effects in the NIR spectra. From a theoretical point of view, it can be argued that NIR spectra contain at maximum the same information as the IR spectra. For that reason better models than obtained from IR are not very likely.

The results from the UV-Vis data indicate that the usefulness of this technique for predicting LR properties is limited. At first glance the results from NMR are somewhat disappointing, but as previously indicated this is due to data acquisition inaccuracies. This is a well-known problem in NMR and tools to correct for these inaccuracies are currently being developed in literature [63]. Therefore, it should be emphasized that such enhanced pre-processing procedures will most likely considerably improve the results. As such, this offers potentials for further research as soon as this problem has been solved. As a consequence of these observations and considerations, it was decided to focus further research on the applicability of the IR and NIR spectra using an extended calibration set of 28 spectra and an independent set of 23 test spectra for validation, selected by PCA of the total set of 51. The results obtained for the IR spectra are summarized in Table 7, showing the mean value of the larger data set for each of the LR properties, the RMSECV based on the model that was built, the number of LV's used and the predictions in terms of the root-means-square-error-of-prediction (RMSEP) values. The latter have also been depicted as prediction plots in Fig. 8.

Table 7. Results of PLS modeling and prediction of the LR properties of crude oils based on IR. Models are based on 28 spectra for calibration and 23 for validation. Mean values and reproducibility values originate from ASTM and IP methods.

LR-property	Mean value	Reproducibility(+/-)	RMSECV	LV's	RMSEP
YLC (wt %)	52.5	1.2	3.7	9	2.8
D_{LR} (g/ml)	0.930	0.0005	0.012	8	0.007
V_{LR} (-)	35.3	-	2.0	6	1.3
S (wt %)	2.16	0.09*S (S>0.9%)	0.56	8	0.24
PP (°C)	29.6	9	9.2	5	10.9
Asph (wt %)	4.1	0.2*Asph	2.2	7	2.5
CR (wt %)	9.7	0.046*(3+CR)	2.1	6	1.6

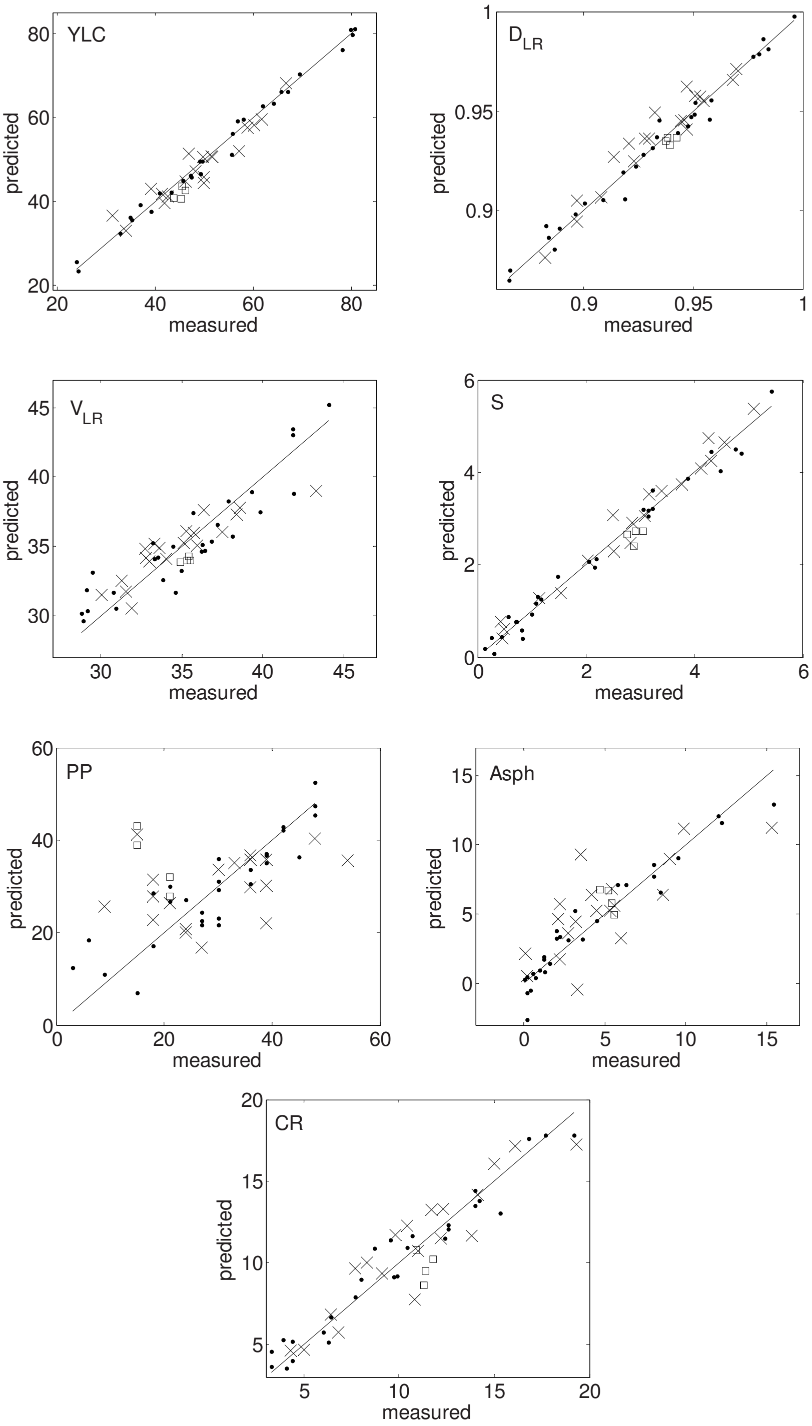


Fig. 8. Prediction plots of PLS modeling of the LR properties based on the IR spectra of crudes oils. (•) calibration spectra, (X) validation spectra of which (□) blends.

As appears, the plots obtained for the properties S, PP, Asph and CR are somewhat scattered. For PP, the corresponding RMSEP-value is relatively large, but of the same order of magnitude as the reproducibility of the ASTM method. The RMSEP values for S, Asph and CR are larger than the reproducibility of the physicochemical methods, which implies that these models are more useful for indicative purposes. However, the models obtained for the properties YLC, D_{LR} and V_{LR} perform very well. The high reliability of the predictions of these properties is reflected in the small RMSEP-values, particularly, regarding the small number of latent variables that was used in these models. The RMSEP values are not as good as the reproducibility values of the corresponding ASTM methods, but within reasonable limits. The RMSEP values are even smaller than the corresponding RMSECV's. This result and the fact that the properties of the test set are close to the average values confirm the validity of the models. Besides, the more extreme samples were selected for the calibration set. It follows that the RMSEP-values are a better indication of the performance of the models than the RMSECV values.

The same pre-processing and modeling procedure as for IR was applied to build PLS models for the different LR properties from the calibration set of 28 NIR spectra. The results are presented in Table 8. The independent set of 23 NIR test spectra was used for validation of the models and the RMSEP-values obtained are added to Table 8. The corresponding prediction plots are shown in Fig. 9.

Table 8. Results of PLS modeling and prediction of the LR properties of crude oils based on NIR spectroscopy. Models are based on 28 spectra for calibration and 23 for validation. Mean values and reproducibility values obtained by ASTM and IP methods.

LR property	Mean value	Reproducibility(+/-)	RMSECV	LV's	RMSEP
YLC (wt %)	52.5	1.2	3.0	5	2.4
D_{LR} (g/ml)	0.930	0.0005	0.017	4	0.008
V_{LR} (-)	35.3	-	2.2	4	1.6
S (wt %)	2.16	0.09*S (S>0.9%)	0.76	7	0.48
PP (°C)	29.6	9	8.7	6	11.6
Asph (wt %)	4.1	0.2*Asph	3.0	4	2.9
CR (wt %)	9.7	0.046*(3+CR)	2.8	4	2.2

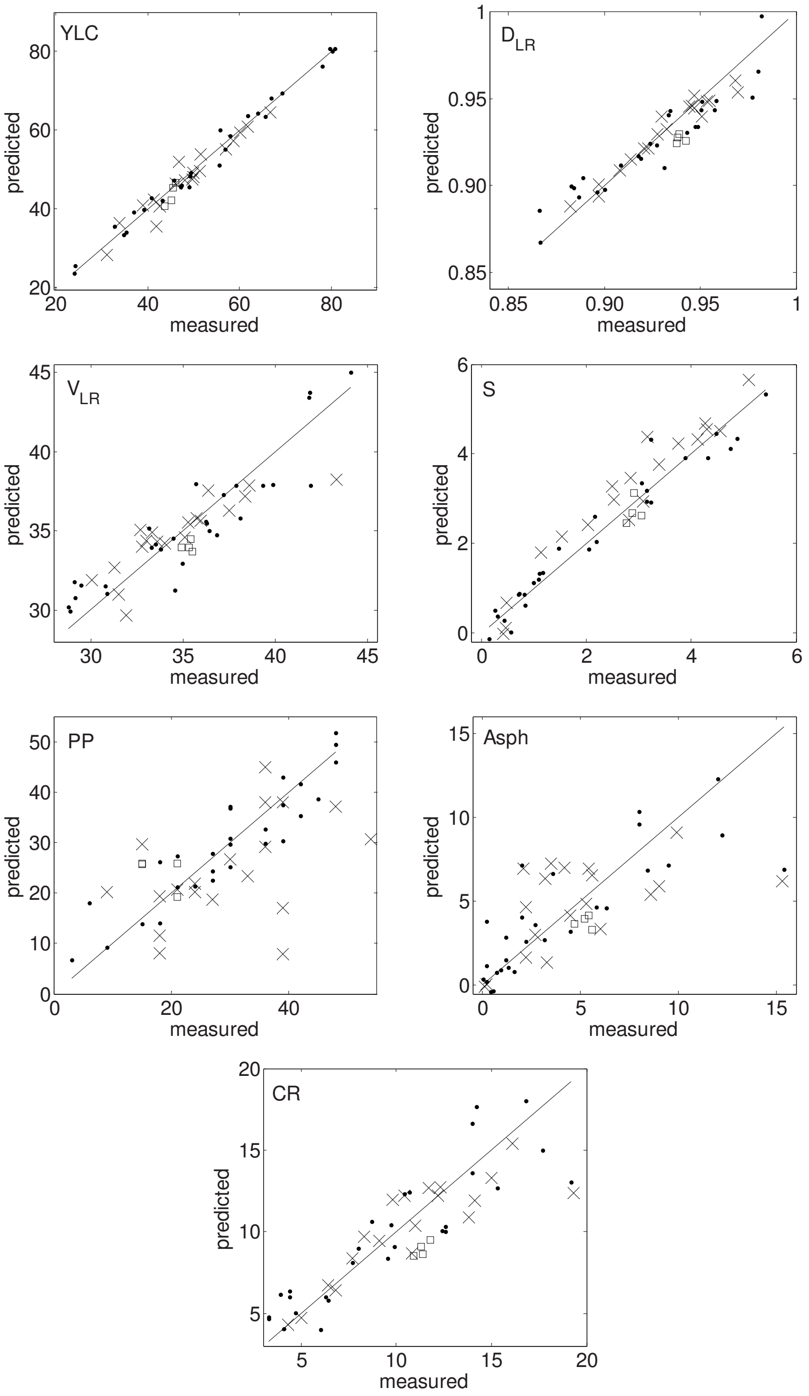


Fig. 9. Prediction plots of PLS modeling of the LR properties based on the NIR spectra of crudes oils. (•) calibration spectra, (X) validation spectra of which (□) blends.

Similar to IR, the properties YLC, D_{LR} and V_{LR} are predicted very well. Furthermore, the RMSEP values are again smaller than the corresponding RMSECV's and fairly close to the reproducibility values of the ASTM-methods. The latter is also valid for the property PP, while the models for S, Asph and CR are more useful for rough, though very fast estimates.

Comparison of the results obtained for IR and NIR demonstrate that in general the IR models perform better than the ones based on the NIR spectra for all properties. In most cases, the RMSEP-values are better and the plots less scattered. At first glance, one could attribute the lower performance of the NIR models to the strong baseline differences at higher wavenumbers in the NIR spectra. However, as indicated by Mullins *et al.* [40], these effects are partly due to differences in electronic transitions and become most apparent for the heavy crudes, since these contain relatively high amounts of aromatic compounds including asphaltenes. For this reason, one might expect a better performance of the NIR model to predict the property Asph, but this is not observed. Furthermore, it should be noted that the NIR spectra have been recorded on an extended range IR instrument. Possibly a dedicated high quality NIR instrument would have provided a better performance of the models. For that reason, it is not possible to judge from this study, whether IR is a better technique to predict LR and SR properties than NIR. However, the practical limitation that such an NIR spectrometer was not available lead to the decision to focus further research on IR spectroscopy.

Finally, the use of a reduced IR spectral region was explored. However, the best results were obtained when using all parts of the spectrum that contain absorption bands, which implies the range 3200-2500 and 1800-670 cm^{-1}. This region was used in all further modeling experiments on IR data.

2.3.8. Short residue properties as function of the flash temperature

In a next step, the IR data were combined with the Atmospheric equivalent Flash Temperature (AFT) values in order to predict the SR properties P, R&B, D_{SR} and V_{SR} from the IR spectra at different AFT's. First, a PCA was applied to the 51 available crude oil IR spectra after MSC and MC using the regions 3200-2500 and 1800-670 cm^{-1}. The first ten scores were used to describe each spectrum and the AFT value was added as the 11th variable and variance scaling was applied. The resulting data set consists of 33 unique crude spectra for P, 37 for R&B, 40 for D_{SR} and 38 for V_{SR} combined with 99, 119, 130 and 118 SR properties for the different AFT values. Subsequently, a PLS regression was performed against the available SR property values. The results, expressed as RMSECV-values and the corresponding number of LV's used, are presented in Table 9. In addition, the plots of the values predicted by the PLS model versus the measured SR properties are presented in Fig. 10. As appears, the RMSECV values for D_{SR} and V_{SR} are small, offering a good alternative for the corresponding ASTM methods, whereas Log P and R&B are not predicted very well.

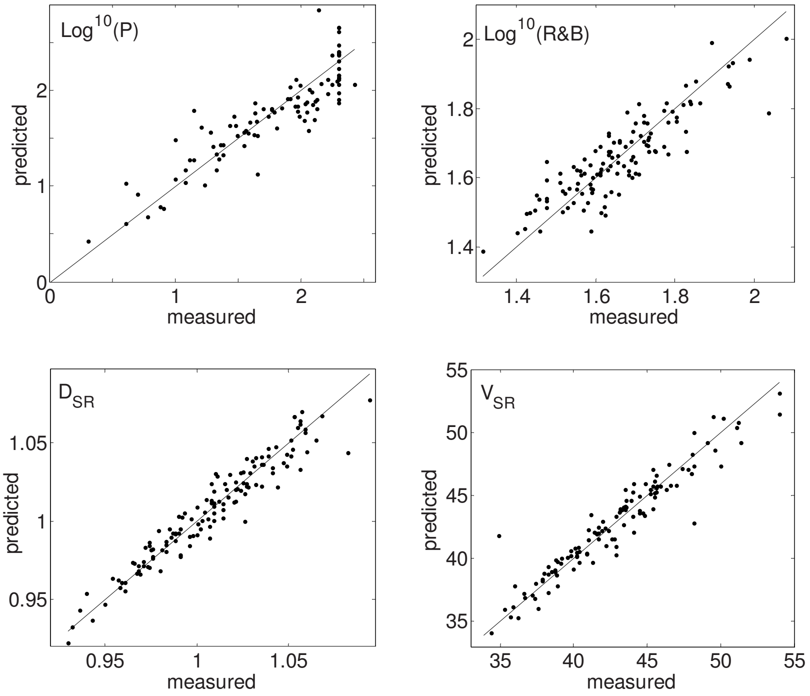


Fig. 10. PLS-correlation plot obtained from IR spectra for the SR properties $\log^{10}(P)$, $\log^{10}(R\&B)$, D_{SR} and V_{SR} using AFT values as additional input.

Table 9. Results of PLS modeling the SR properties at different flash temperatures (AFT input) of crude oils based on IR spectroscopy. Mean values and reproducibility values obtained by ASTM methods.

SR property (AFT)	Mean Value	Reproducibility(+/-)	RMSECV	LV's
Log^{10} P (0.1 mm)	1.73	0.01-0.5	0.32	3
Log^{10} R&B (°C)	1.66	0.04	0.09	3
D_{SR} (g/ml)	1.007	0.005	0.015	3
V_{SR} (-)	42.7	-	1.8	3

The results of the prediction models with the YSC values as input are presented in Table 10 and Fig. 11. They illustrate that the performance of the prediction models is better for all SR properties, except for a slight improvement from 0.09 to 0.08 for R&B, when AFT is used as discriminator.

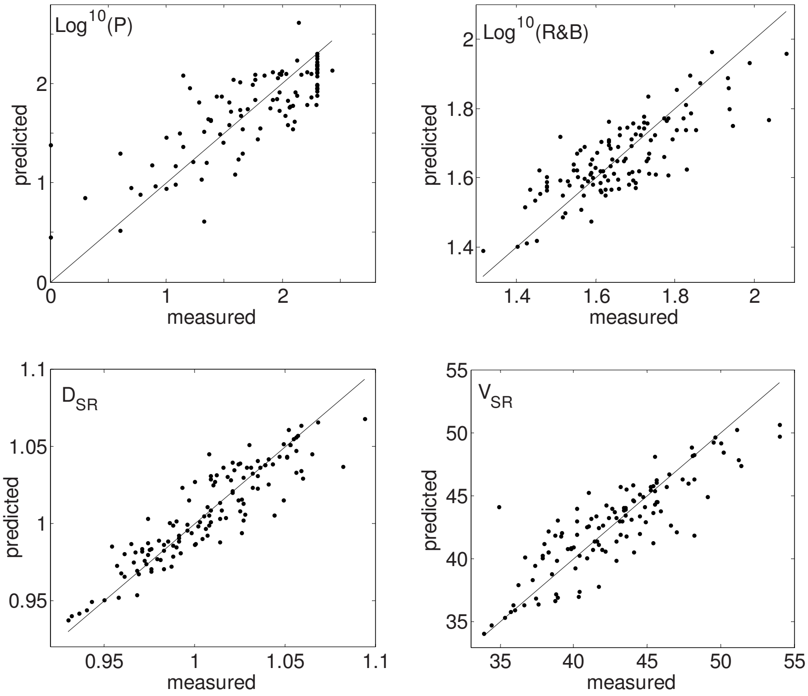


Fig. 11. PLS-correlation plot obtained from IR spectra for the SR properties $\log^{10}$(P), $\log^{10}$(R&B), D_{SR} and V_{SR} using YSC values as additional input.

Table 10. Results of PLS modeling the SR properties at different flash temperatures (YSC input) of crude oils based on IR spectroscopy. Mean values and reproducibility values obtained by ASTM methods.

SR property (YSC)	Mean Value	Reproducibility(+/-)	RMSECV	LV's
Log^{10} P (0.1 mm)	1.73	0.01-0.5	0.39	3
Log^{10} R&B (°C)	1.66	0.04	0.08	3
D_{SR} (g/ml)	1.007	0.005	0.020	3
V_{SR} (-)	42.7	-	2.1	3

2.4. Conclusions

Raman and UV-Vis spectroscopy appear not to be useful to predict LR properties from the corresponding crude oil spectra. Hence, it is neither useful to study the potential of predicting SR properties. For Raman, this is due to fluorescence and self-absorption phenomena, which prevent the acquisition of useful data. With respect to UV-Vis spectroscopy, the limited value can be attributed to the small amount of aromatic compounds and the inherent large bandwidth of electronic transition bands.

The results obtained from ^{1}H- and ^{13}C-NMR data are disappointing regarding the commonly acknowledged high information content of the NMR spectra. This is largely due to the lack of reproducibility of the data, inherent to the limited spectral acquisition and processing facilities of the applied NMR spectrometer. Data point shifting and calibrating might be a way to solve this problem, but this turned out to be very time-consuming. Pre-processing is another option, but quite complex for the available NMR data. Software to correct for this type of data point shifting is currently being developed commercially and as a result improvement of the results of prediction can be expected.

IR and NIR spectroscopy combined with PLS modeling have the highest potential to predict LR and SR properties as a function of the flash temperature. Both techniques offer a very fast alternative for the time-consuming ASTM and IP methods, currently used to obtain this type of information. In accordance with vibrational spectroscopy theory, IR and NIR techniques score best on the same properties as a result of redundancy in the information content. However, the results obtained from the IR models are better than for NIR in all cases. Even though pronounced electronic absorption profiles in the NIR spectra of heavy crude oils are present, significant better predictions for the property Asph are not observed. Possibly, a dedicated NIR instrument might improve the predictive power of the NIR models.

As demonstrated by the prediction plots and RMSEP-values obtained from the set of 23 validation spectra, the prediction of the LR properties YLC, D_{LR} and V_{LR} from the IR data is quite reliable, while the inaccuracy in the prediction of PP is similar to that of the corresponding ASTM method. It follows that IR spectra of crude oils can be used to predict these LR properties with reasonable accuracy. The results to predict SR properties from the IR spectra of the crude oils are also promising. Modeling as a function of the Atmospheric equivalent Flash Temperature (AFT) was particularly successful for the density (D_{SR}) and the viscosity (V_{SR}). It is expected that a higher accuracy and reproducibility of the input property values on the one hand, and further optimization of the models by deliberate wavelength selections on the other hand, will lead to even better results. Other potential options for improvement are the application of other pre-processing methods and outlier removal. The performance of the models to predict SR properties decreases when the yield short on crude (YSC) is used as input instead of the AFT. The results of modeling the IR and NIR spectra recorded at different instruments and with different sample cells show that the applied analysis protocol including sample pre-treatment is well chosen and leads to highly reproducible results.

References

1. D.E. Wetmore, C.K. Hancock, R.N. Traxler, Anal. Chem. 38 (1966) 225.
2. A.N. Ratov, A.P. Krylov, Neftekhimiya 36 (1996) 195.
3. N. Berkowitzt, J.G. Speight, Fuel 54 (1975) 138.
4. www.astm.org.
5. www.americanenergyinstitutes.org.
6. J.G. Speight, Fuel 49 (1970) 76.
7. M.H.B. Hayes, M. Stacey, J. Standley, Fuel 51 (1972) 27.

8. M.U. Hasan, M.F. Ali, A. Bukhari, Fuel 62 (1983) 518.
9. M.U. Hasan, M.F. Ali, A. Bukhari, Fuel 64 (1985) 839.
10. J.M. Jacobson, M.R. Gray, Fuel 66 (1987) 753.
11. C. Giavarini, C. Vecchi, Fuel 66 (1987) 868.
12. J. Thiel, M.R. Gray, AOSTRA J. Research 4 (1988) 63.
13. C.H. Tsai, M.D. Deo, F.V. Hanson, A.G. Oblad, Fuel Sci. Technol. Int. 10 (1992) 1437.
14. L. Michon, D. Martin, J.-P. Planche, B. Hanquet, Fuel 76 (1997) 9.
15. H.A. Akrami, M.F. Yardim, A. Akar, E. Ekinci, Fuel 76 (1997) 1389.
16. J. Bryan, K. Mirotchnik, A. Kantzas, J. Can. Petrol. Technol. 42 (2003) 29.
17. F. Ali, Z.H. Khan, N. Ghaloum, Energ. Fuels 18 (2004) 1798.
18. K.H. Michaelian, S.L. Zhang, R.H. Hall, J.T. Bulmer, Spectrochim. Acta A 57 (2001) 73.
19. L.C.T. Shoute, K.J. Schmidt, R.H. Hall, M.A. Webb, S. Rifai, P. Abel, P.H. Arboleda, A. Savage, J.T. Bulmer, G.R. Loppnow, Appl. Spectrosc. 56 (2002) 1308.
20. J. Jehlička, O. Urban, J. Pokorný, Spectrochim. Acta A 59 (2003) 2341.
21. K.H. Michaelian, R.H. Hall, J.T. Bulmer, Can. J. Anal. Sci. Spectrosc. 49 (2004) 43.
22. E. Evans, B. Batts, N. Cant, Fuel 66 (1987) 327.
23. P.W. Yang, H.H. Mantsch, L.S. Kotlyar, J.R. Woods, Energ. Fuels 2 (1988) 26.
24. H. Ganz, W. Kalkreuth, Erdoel & Kohle, Erdgas, Petrochem. 43 (1990) 116.
25. E.A. Cloutis, AOSTRA J. Research 6 (1990) 17.
26. S. Kokot, N. Asker, Proc. 8th Int. Conf. Fourier Transform Spectrosc., SPIE- Int. Soc.Opt. Engin. 1575 (1992) 495.
27. S. Kokot, N. Asker, Proc. 6th Int. Conf. Near Infr. Spectrosc. 22 (1995) 1994.
28. N. Pieri, J. Kister, L. Quoniam, J.P. Planche, L. Germanaud, Polycycl. Arom. Comp. 9 (1996) 29.
29. H. Chung, M.-S. Ku, J.-S. Lee, Vibr. Spectrosc. 20 (1999) 155.
30. K.H. Michaelian, S. Zhang, R.H. Hall, J.T. Bulmer, Can. J. Anal. Sci. Spectrosc. 46 (2001) 10.
31. B. Pavoni, N. Raido, R. Plassa, S. Frignani, Ann. Chim. 94 (2004) 521.
32. W.I. Friesen, K.H. Michaelian, Y. Long, T. Dabros, Am. Chem. Soc., Div. Fuel Chem. 50 (2005) 228.
33. T.A. Adedosu, O.O. Sonibare, J. Appl. Sci. 5 (2005) 906.
34. H. Yuan, X. Chu, H. Li, Y. Xu, Fuel 85 (2006) 1720.
35. K.J. Johnson, R.E. Morris, S.L. Rose-Pehrsson, Energ. Fuels 20 (2006) 727.
36. K.H. Michaelian, R.H. Hall, K.I. Kenny, Spectrochim. Acta A 64 (2006) 703.
37. R.R. Coelho, I. Hovell, M.B. de Mello Monte, A. Middea, A. Lopes de Souza, Fuel Process. Techn. 87 (2006) 325.
38. Y. Hongfu, C. Xiaoli, L. Haoran, X. Yupeng, Fuel 85 (2006) 1720.
39. J.M. Brown, C.J. Chrostowsky, PCT Int. Appl. (2006) 2006023800.
40. O.C. Mullins, S. Mitra-Kirtley, Y. Zhu, Appl. Spectrosc. 46 (1992) 1405.

41. A. Espinosa, A. Martens, G. Ventron, C.D. Lambert, A. Pasquier, Eur. Pat. Appl. (1989) 305090.
42. D.A. LeFebre, L.M. Lane, US Patent (1995) 5430295.
43. S. Bages, B. Descales, D. Lambert, J.-R. Llinas, A. Martens, S. Osta, M. Sanchez, PCT Int. Appl. (1996) 9611399.
44. Z. Sikora, W. Salacki, Petrol. Coal 38 (1996) 65.
45. R. Van Doorn, F.H. Hooyman-Spaargaren, R.J. Neugebauer, L.M. Smeets, PCT Int. Appl. (1997) 9714953.
46. J. Kelly, D. Lambert, A. Martens, PCT Int. Appl. (1998) 9836274.
47. B. Descales, D. Lambert, J.-R. Llinas, A. Martens, S. Osta, M. Sanchez, S. Bages, Sylvie. U.S. patent (2000) 5712797.
48. M. Blanco, S. Maspoch, I. Villarroya, X. Peralta, J.M. González, J. Torres, Appl. Spectr. 55 (2001) 834.
49. S. Macho, M.S. Larrechi, Trends Anal. Chem. 21 (2002) 799.
50. S. Satya, R.M. Roehner, M.D. Deo, F.V. Hanson, Energ. Fuels 21 (2007) 998.
51. R.J. Mikula, V.A. Munoz, N. Wang, B. Bjornson, D. Cox, B. Moisan, K. Wiwchar, J. Can. Petr. Techn. 42 (2003) 50.
52. H. Chung, Appl. Spectrosc. Rev. 42 (2007) 251.
53. M.V. Reboucas, E.C. Santos, F.S.V. Vieira, Vibr. Spectrosc. 44 (2007) 187.
54. O. Abbas, C. Rebufa, N. Dupuy, A. Permanyer, J. Koster, D.A. Azevedo, Fuel 85 (2006) 2653.
55. K.A. Lintelmann, Anal. Chem. 67 (1995) 327R.
56. T. Zhang, Y. Liang, B. Li, H. Cui, F. Gong, Anal. Sci. 20 (2000) 717.
57. H. Swierenga, F. Wulfert, O.E. de Noord, A.P. de Weijer, A.K. Smilde, L.M.C. Buydens, Anal. Chim. Acta 411 (2000) 121.
58. Q. Abbas, N. Dupuy, C. Rebufa, L.Vrielynck, J. Kister, A. Permanyer, Appl. Spectrosc. 60 (2006) 304.
59. H. Chung, S. Cho, Y. Toyoda, K. Nakano, M. Maeda, Analyst 131 (2006) 684.
60. W. Tuchler, A, Ecker, G. Kunar Schreferl, Petrol. Coal 42 (2000) 56.
61. H. van der Voet, Chemom. Intell. Lab. Syst. 25 (1994) 313.
62. J.B. Lambert, E.P. Mazzola (Eds), *Nuclear Magnetic Resonance*, Pearson Education Inc., New Jersey (2004).
63. J. Forshed, R.J.O. Torgrip, K. Magnus Åberg, B. Karlberg, J. Lindberg, S.P. Jacobsson, J. Pharm. Bio. Anal. 38 (2005) 824.

Chapter 3

Prediction of Long Residue Properties from Combined Infrared, ^{1}H-NMR and ^{13}C-NMR Spectra

Abstract

Research has been carried out to determine the potential of partial least squares (PLS) modeling of IR spectra of crude oils combined with the corresponding ^{1}H- and ^{13}C-NMR data, to predict the long residue (LR) properties of these substances. The study elaborates further on the results described in Chapter 2. PLS-modeling has been carried out for 7 different LR properties, *i.e.*, yield-long-on-crude (YLC), density (D_{LR}), viscosity (V_{LR}), sulfur content (S), pour point (PP), asphaltenes (Asph) and carbon residue (CR). The research work was based on the spectra of 48 crude oil samples of which 28 were used to build the PLS-models and the remaining 20 for validation. For each property, PLS-modeling was carried out on single type IR, ^{13}C-NMR and ^{1}H-NMR spectra and on 3 sets of merged spectra, i.e. IR + ^{1}H-NMR, IR + ^{13}C-NMR and IR + ^{1}H-NMR + ^{13}C-NMR. The merged spectra were created by considering the NMR data as a scaled extension of the IR spectral region. In addition, PLS modeling of coupled spectra was performed after a Principal Component Analysis (PCA) of the IR, ^{13}C-NMR and ^{1}H-NMR calibration sets. For these models, the 10 most relevant PCA scores of each set were concatenated and scaled prior to PLS-modeling. The validation results of the individual IR models, expressed as root-mean-square-error-of-prediction (RMSEP) values, turned out to be slightly better than those obtained for the models using single input ^{13}C-NMR or ^{1}H-NMR data. For the models based on IR spectra combined with NMR data, a significant improvement of the RMSEP-values was not observed neither for the models based on merged spectra, nor for those based on the PCA-scores. It implies, that the commonly accepted complementary character of NMR and IR is, at least for the crude oil and bitumen samples under study, not reflected in the results of PLS-modeling. In view of these results, it can be concluded that IR spectroscopy is preferred over NMR spectrometry for the prediction of LR properties of crude oils at site.

3.1. Introduction

The characterization of crude oils in terms of physical and chemical properties is of major importance in petrochemical industries [1-2]. A low density, for instance, simplifies transportation and points to a high gasoline yield, while high concentrations of sulfur and heavy metals have environmental implications. Crude oils are also analyzed to determine their potential for bitumen application. Although this type of bitumen is a residual substance obtained after distillation, it is widely used for *e.g.* road pavement, roofing and sealing and therefore representing a substantial economical value. This explains the demand to determine the bitumen suitability of a crude oil prior to distillation as well.

Current standard methods to characterize crude oils and bitumen are mainly based on a series of elaborate and time-consuming tests developed by the ASTM [3] and the IP [4] institutes. As a much faster alternative, molecular spectroscopic techniques have been proposed, with the main argument that these methods theoretically reflect the complete qualitative and quantitative molecular composition of a sample [5]. Next to Raman spectroscopy [6-10], the majority of the studies to characterize crude oils involve IR [9, 11-30] and NIR) spectroscopy [17, 19, 30-44]. Occasionally, NMR spectrometry has been used [11, 45-55]. Besides, it has been shown that multivariate data analysis of the spectra by means of methods, such as Partial Least Squares (PLS) regression, Principal Component Analysis (PCA) and artificial neural networks (ANN) can be of considerable help to enhance the reliability of the characterization and classification [21, 25, 29, 35, 36, 39, 40, 42, 56-59].

In Chapter 2, the potential of direct characterization and chemometrical classification of crude oils from their Raman, UV-Vis, NMR, IR and NIR spectrum has been explored. The aim of that study was to link the more fundamental chemical composition of these substances to so-called long residue (LR) and short residue (SR) properties. Hence, it would be possible to classify crude oils in terms of suitability for bitumen manufacture and to predict bitumen properties as a function of the flash temperature directly from the crude oil spectra. The results revealed that IR spectroscopy in particular can be used as a fast and viable alternative for the currently applied elaborative ASTM and IP laboratory tests.

However, next to IR spectroscopy, NMR is commonly acknowledged as a very powerful tool for structural analysis. Besides, the information from NMR spectra is supposed to be highly complementary to that from vibrational spectra. It is, therefore, a logical next step to explore improvement of the current IR prediction models by adding ^{1}H- and ^{13}C-NMR data to the IR spectra as input for modeling. In order to take full advantage of the complementary character of both techniques, this can be done as stand-alone IR, ^{1}H-NMR and ^{13}C-NMR methods, but also in an integrated manner by concatenating the IR and NMR spectra to single input data for PLS-modeling. For that reason, we decided to perform such a study on both the separate data sets of IR, ^{1}H- and ^{13}C-NMR spectra of the same 48 compounds, and on 3 combinations of the spectra, i.e., IR + ^{1}H-NMR, IR + ^{13}C-NMR and IR + ^{1}H-NMR + ^{13}C-NMR. It will be shown that the predictive value of the models based on IR spectra perform slightly better than the ones based on NMR data as well as on the combinations of IR and NMR spectra.

3.2. Experimental

3.2.1. Crude oils, properties and treatment

All crude oil samples have been stored and prepared according to the protocol as described in Chapter 2. All measurements have been carried out at 20°C in closed cells. Under these conditions the spectra were found to be highly reproducible [30].

Partial Least Squares (PLS) regression has been performed to correlate each of the LR properties with their IR, ^{1}H-NMR and ^{13}C-NMR spectra or the combinations of these data as described above. The following LR properties have been studied; long residue yield-on-crude (YLC), density (D_{LR}), viscosity (V_{LR}), sulfur contents (S), pour point (PP), asphaltenes (Asph) and carbon residue (CR).

From the total number of 51 samples as listed with their corresponding LR properties in Table 1 of Chapter 2, 48 crude oils have been used in this study. The blends V21-V23 were excluded as no NMR data were available for these compounds. The set of 48 samples was divided into a subset of 28 calibration compounds (compounds C1-C28) for PLS-modeling and 20 samples for validation (compounds V1-V20) as result of a PCA of the IR spectra. For each LR-property, PLS prediction models were built from the individual IR, ^{13}C-NMR and ^{1}H-NMR calibration sets and from the 3 calibration sets of combined spectra, i.e., IR + ^{1}H-NMR, IR + ^{13}C-NMR and IR + ^{1}H-NMR + ^{13}C-NMR. Prior to concatenation of the different spectra, scaling and normalization was carried out. Besides, a 1st derivative was applied to the IR data mainly to eliminate baseline effects. Merging of the IR and NMR spectra was carried out by concatenating the ^{1}H- and/or ^{13}C-NMR to the IR data points. The regions 3200-2500 and 1800-670 cm^{-1} were used for the IR data (950 data points). For the ^{1}H-NMR the regions 0.5-4 and 6-8 ppm (1100 data points) and for the ^{13}C-NMR the regions 10-50 and 115-170 ppm were used (1900 data points). This resulted in an X/Y input matrix of arbitrary units for each sample. In order to minimize the effect of scaling and hence obtain better prediction models, PLS modeling was also carried out on the combined scores from a PCA of the individual IR, ^{13}C-NMR and ^{1}H-NMR calibration sets. The variance captured by the first 10 scores was 99.4, 97.3 and 99.5 % respectively. For each spectroscopic technique, these 10 most relevant scores were merged and autoscaled resulting in input matrices of 20 scores for the combination IR + ^{1}H-NMR and IR + ^{13}C-NMR, and 30 scores for the combination IR + ^{1}H-NMR + ^{13}C-NMR.

Finally, all models obtained were tested by submitting the individual or merged validation data for prediction. The results for each of the LR-properties, expressed as root-mean-square-error-of-prediction (RMSEP) values, have been compared and evaluated.

3.2.2. IR and NMR spectroscopy

IR measurements were carried out at room temperature on a Bruker Tensor 27 FTIR spectrometer. IR data were recorded with a deuterated triglycine sulfate (DTGS) detector. The sample compartment of the instrument was flushed with dry air to reduce interference of H_2O and CO_2. IR spectra were recorded with a horizontal ATR accessory (FastIR^{TM}, Harrick) equipped with a heatable base plate and ZnSe as the optical element at 20°C. A cover plate was used to prevent sample evaporation during measurement. The optical resolution of the IR spectra was 4 cm^{-1} with apodization set on 'medium' and 50 scans were accumulated for each spectrum.

^{13}C-NMR-spectra were recorded at room temperature on a Bruker AC300 FT-NMR instrument, set to ^{13}C at 75 MHz using WinNMR 6.0 software for control and acquisition. The instrumental settings were as follows: SW: 17857 Hz; FW: 20000 Hz; SI and TD: 32K data points; O1: 6100; O2: 4600; LB: 2 Hz and SR: -1405. For each sample, fresh homogeneous solutions were prepared of 300 mg crude oil in 0.5 ml $CDCl_3$. The solutions were spiked with tetramethylsilane (TMS) and analyzed in sealed 5 mm quartz NMR tubes with a spin speed of 20 Hz.

^{1}H-NMR spectra were recorded at room temperature on a Bruker AC300 FT-NMR instrument, set to ^{1}H at 300 MHz using WinNMR 6.0 software for control and acquisition. The instrumental settings were as follows: SW: 4505.4 Hz; FW: 5700 Hz; SI and TD: 32K data points; O1: 6100; O2: not used; LB: 0.3 Hz and SR: 3367.25. For each sample, a fresh homogeneous solution was prepared of 50 mg crude oil in 0.5 ml $CDCl_3$. The solutions were spiked with tetramethylsilane (TMS). Solutions were analyzed in sealed 5 mm quartz NMR tubes with a spin speed of 20 Hz.

Prior to modeling the ^{1}H- and ^{13}C-NMR spectra, a data point shift correction was applied to compensate for the shifts observed in the NMR spectra [60-62]. The latter is a commonly known problem that is attributed to variations in the sample matrix and to instrumental instabilities [62], hampering the building of statistical models with NMR data [63, 64]. The measured ^{13}C-NMR data were shift corrected by dividing the spectra from 10 to 115 ppm into 9 different ppm regions as indicated in Table 1. Next, a peak in each region was used to determine the average shift value and this value was used to shift the data points in that particular region. The same procedure was applied to the ^{1}H-NMR spectra, using the 5 different regions as listed in Table 1.

Table 1. Regions and peak positions (ppm) used for the shift correction of the ^{1}H- and ^{13}C-NMR spectra.

	Region (ppm)			Peak position (ppm)	
^{1}H-NMR	0.70	1.00		0.87	0.91
	1.00	1.35		1.20	1.30
	1.40	1.45		1.41	1.44
	2.32	2.38		2.33	2.36
	6.50	8.00		7.20	7.30
^{13}C-NMR	10.00	13.00		11.30	11.60
	13.00	18.00		14.00	14.50
	18.50	24.20		22.70	23.00
	24.00	25.80		24.10	24.70
	26.20	27.80		26.00	26.85
	27.70	28.50		27.80	28.50
	29.20	31.00		29.20	30.00
	31.00	50.00		31.90	32.20
	75.00	80.00		76.90	77.10

3.2.3. Chemometrics

PLS modeling was carried out as in Chapter 2 with the PLS Toolbox (Eigenvector Research, Inc.) for MatLab (The MathWorks, Inc.). Prior to modeling, the IR spectra were pre-processed by first taking the 1st derivative (25 pt. Savitzky-Golay smoothing), followed by Multiplicative Signal Correction (MSC) and Mean Centering (MC) of the data. The ^{1}H- and ^{13}C-NMR spectra were pre-processed by a 25 pt. Savitzky-Golay smoothing followed by MSC and MC. Pre-processing parameters were chosen based on prior knowledge for each spectroscopic technique combined with trial and error. Wide spectral regions were used as input, since it was concluded from a previous study [30] that this approach provided the best models. The obtained error values expressed as Root Mean Square Error of Cross Validation (RMSECV) values were based on ”leave one out” (LOO) cross validation (CV). The number of latent variables (LV), used for a PLS model, was related to the minimum value of the RMSECV-value but limited to 10 to reduce the chances of overfitting.

PCA of the IR, ^{1}H-NMR and ^{13}C-NMR spectral data sets was also carried out using the MatLab PLS Toolbox package. For PLS modeling of the PCA scores, the 10 most relevant scores from the IR, ^{1}H-NMR and ^{13}C-NMR calibration sets were joined and autoscaled. In all cases, the models with the lowest RMSECV (LOO) were used to predict the LR properties of the samples in the validation sets.

3.3. Results and discussion

3.3.1. Models based on separate IR, ^{1}H- and ^{13}C-NMR spectral data sets

As described in Chapter 2 and illustrated in Fig. 1, the IR spectra of the 48 crude oils are very similar. All IR spectra are dominated by characteristic strong aliphatic C-H stretching and deformation bands around 2900 and 1450 cm^{-1}, respectively, while the band of medium intensity at around 1375 cm^{-1} originates from the symmetric bending vibrations of methyl groups attached to carbon atoms. Furthermore, a complex band pattern in the region 1200-650 cm^{-1} is present in all spectra, which can be assigned to a mixture of aliphatic and aromatic skeletal vibrations. Also, the $(CH_2)_{n>4}$ in phase rocking at 720 cm^{-1}, characteristic for longer aliphatic chains, is observed in all IR spectra. More detailed examination of the fingerprint region, however, reveals significant differences, when comparing the spectra of light crudes with those of the more dark and viscous heavy ones.

As illustrated in Fig. 2, heavy crudes exhibit in general stronger and broader bands in the C-H out of plane region 900-700 cm^{-1}, whereas the light samples show sharper peaks and two additional bands at around 675 and 695 cm^{-1}. The latter bands can be assigned to the sextant ring bending or puckering mode of relatively small aromatic molecules, such as benzenes and xylenes [65]. This type of compounds is known to be a considerable fraction (~ 5% w/w) of light crudes [66]. In contrast, heavy crude oils contain large amounts of so-called asphaltenes, a group of compounds that mainly consists of complex polycyclic aromatic molecules of 4 to 5 fused aromatic rings, together with aliphatic bridges and heteroatoms.

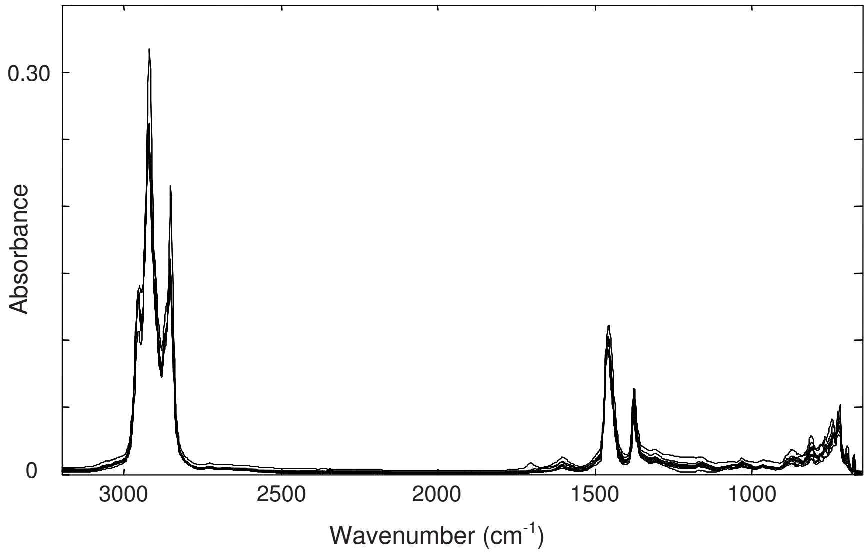


Fig. 1. Overlay of the IR spectra of 10 representative crude oil samples.

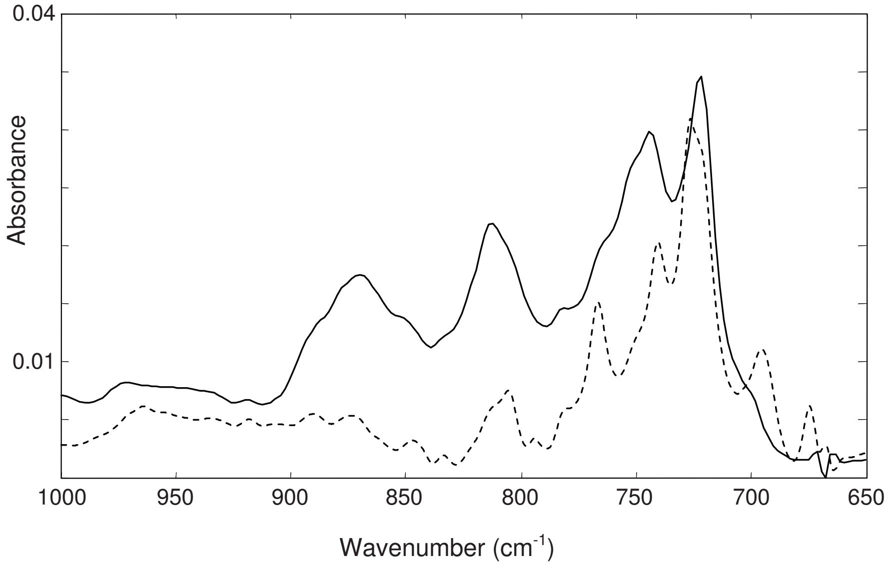


Fig. 2. The aromatic fingerprint region (1200-650 cm^{-1}) of a heavy crude (sample C22, solid line) and a light crude oil (sample C3, dashed line).

Not only the broad band observed around 875 cm^{-1} can be assigned to these compounds, but also the slightly elevated baseline between 1300 and 1000 cm^{-1} (Fig. 1). Finally, heavy crudes show, in general, a minor contribution at

1705 cm^{-1}, which is indicative for the presence of C=O and hence of a small fraction of oxidized compounds.

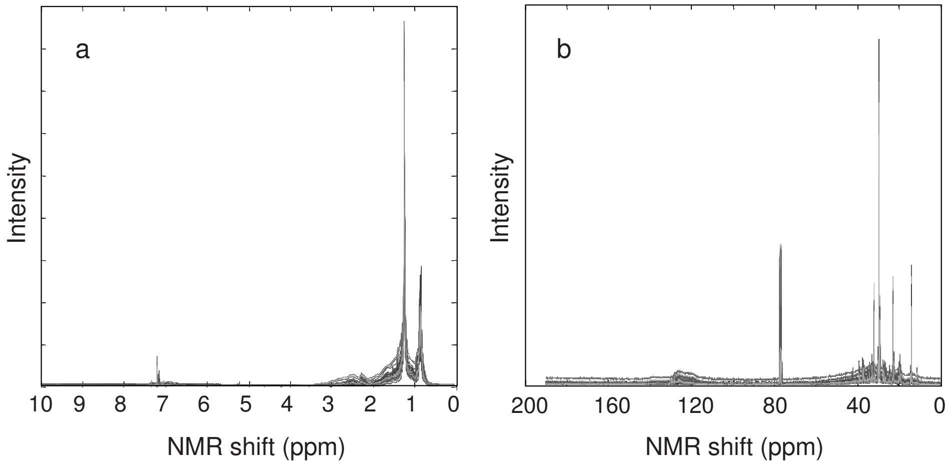


Fig. 3. The ^{1}H-NMR (a) and ^{13}C-NMR (b) spectra of 48 crude oil samples.

Table 2. RMSEP values of PLS models to predict LR properties from IR, ^{1}H-NMR or ^{13}C-NMR spectra. Models are based on calibration spectra C1-28. RMSEP values are obtained from validation spectra V1-20. Mean values and reproducibility obtained from physical tests have been included.

Input		LR-property						
		YLC (wt %)	D_{LR} (g/mL)	V_{LR} (-)	S (wt %)	PP (°C)	Asph (wt %)	CR (wt %)
	Mean	52.5	0.930	35.3	2.16	29.6	4.1	9.7
	Reproduc.	1.2	0.0005	-	0.09 xS	9	0.2x Asph	0.046x (3+CR)
IR	RMSEP	2.8	0.007	1.3	0.24	10.9	2.5	1.6
^{1}H-NMR	RMSEP	6.9	0.014	2.1	0.66	9.2	3.4	2.9
^{13}C-NMR	RMSEP	5.4	0.019	2.4	0.99	11.7	3.7	3.5

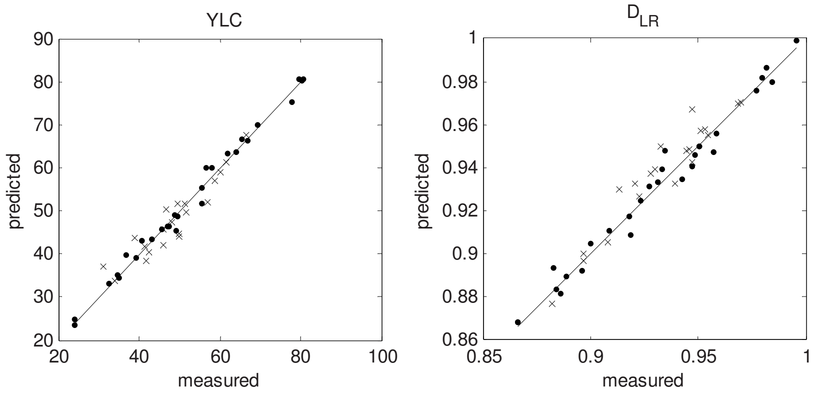
Likewise, the MSC corrected ^{1}H-NMR and ^{13}C-NMR spectra, presented in Figure 3, are mutually very much the same. In accordance with literature [43, 4, 51], small differences in the ^{1}H-NMR spectra are observed in the ratio of aliphatic protons (δ = 0-4 ppm) and aromatic protons (δ = 6-9 ppm). Similar information can be extracted from the ^{13}C-NMR spectra where the aromatic and olephinic ^{13}C chemical shifts cause a complex pattern at δ = 118-170 ppm, while the aliphatic carbons show up at δ = 0-70 ppm [49, 51]. The validation results for the different spectroscopic methods, expressed as RMSEP values and obtained by submitting the independent test set of 20 crude oil spectra to the different PLS models, have been summarized in Table 2.

As appears from Table 2, the RMSEP-values for IR, ^{1}H-NMR and ^{13}C-NMR are of the same order of magnitude as the reproducibility values, determined by the chemical/physical methods (except for D_{LR}). Besides, the 3 spectroscopic techniques perform the best on the same properties. In particular, the models for YLC, D_{LR}, V_{LR}, S and PP exhibit a reliability level similar to that of the corresponding ASTM and IP methods. To illustrate this, the correlation plots of the properties YLC and D_{LR} for the 3 spectroscopic methods are presented in Fig. 4.

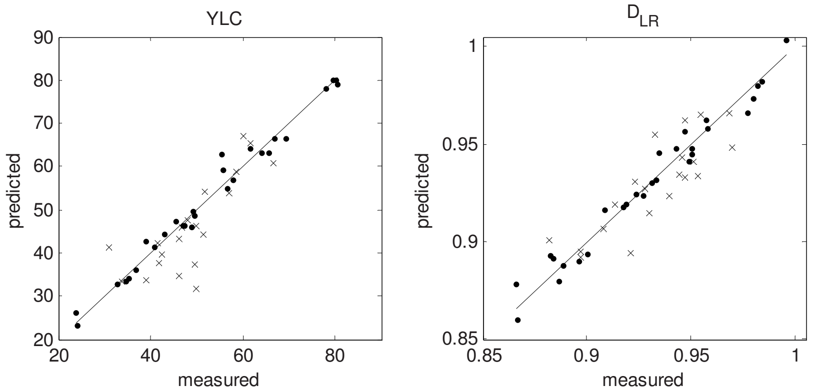
Table 2 also shows that the scores for the properties Asph and CR are less good, but the corresponding models are still useful to obtain a fair indication of the property-values. Further comparison of the RMSEP values reveals that the models built from IR spectra perform significantly better than the ones based on ^{1}H-NMR and ^{13}C-NMR data, except for PP. It should be noted, however, that the IR spectra have been recorded from pure crude oils. In contrast, NMR data acquisition required the samples to be dissolved in $CDCl_3$. This will not only lead to a change in the intermolecular interactions of the oil constituents, but also to differences in concentration between heavy and light crude oil samples. Possibly, this results in a decreased performance of the NMR models.

The RMSEP-values obtained for the IR models appear to be practically identical to the ones reported in Chapter 2, even though in the present study only 48 samples have been used instead of 51. The latter is due to the fact that the NMR data of 3 blends were not available.

IR



^{1}H-NMR



^{13}C-NMR

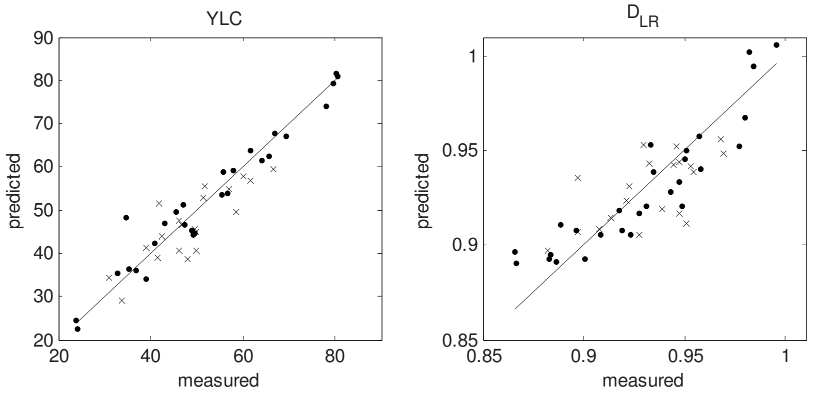


Figure 4. Correlation plots of the PLS models to predict the LR-properties YLC and D_{LR} of crude oils. Models are based on the IR, ^{1}H-NMR and ^{13}C-NMR spectra of 28 calibration samples (•) and 20 validation samples (x).

3.3.2. Models based on merged IR, ^{1}H- and ^{13}C-NMR spectral data

The results of PLS modeling the 3 data sets of merged IR and NMR spectra, expressed as RMSEP-values, are presented in Table 3. In addition, the PLS correlation plots of the properties YLC and D_{LR} are presented in Fig. 5 to illustrate the performance of the models and to allow comparison with the results of modeling single type spectra. It appears that the combination of IR and NMR spectra results in improved RMSECV values, but the RMSEP values are worse. This is particularly reflected when comparing the plots of Fig. 4 and 5. In Fig. 5, the calibration points (•) almost coincide with the calibration line, whereas the validation samples (x) show larger deviations compared to Fig. 4. It indicates that for the data sets of combined IR and NMR spectra, the selection criterion of taking the number of LV’s that corresponds with the minimal RMSECV (LOO) value might result in overfitting of the data. This was confirmed by taking less LV's. As expected, the RMSEP values were not better, but the calibration models turned out to be more realistic. Comparison of the results with those obtained from the single type spectra (Table 2) shows that the RMSEP values of the combined data are of the same order of magnitude as the ones obtained for the single type IR, ^{1}H-NMR and ^{13}C-NMR models for all properties. Apparently, the commonly accepted complementary character of IR and NMR spectroscopic data is, at least for crude oil samples, not reflected in a significant improvement of the results of prediction of crude oil LR properties. The same conclusion, that a combination of two complementary spectra of the same compound is not necessarily more powerful than a single spectroscopic technique has also been reported for the combination of IR and mass spectra [67], for IR and Raman spectra [68] and NIR and MIR spectra [69, 70].

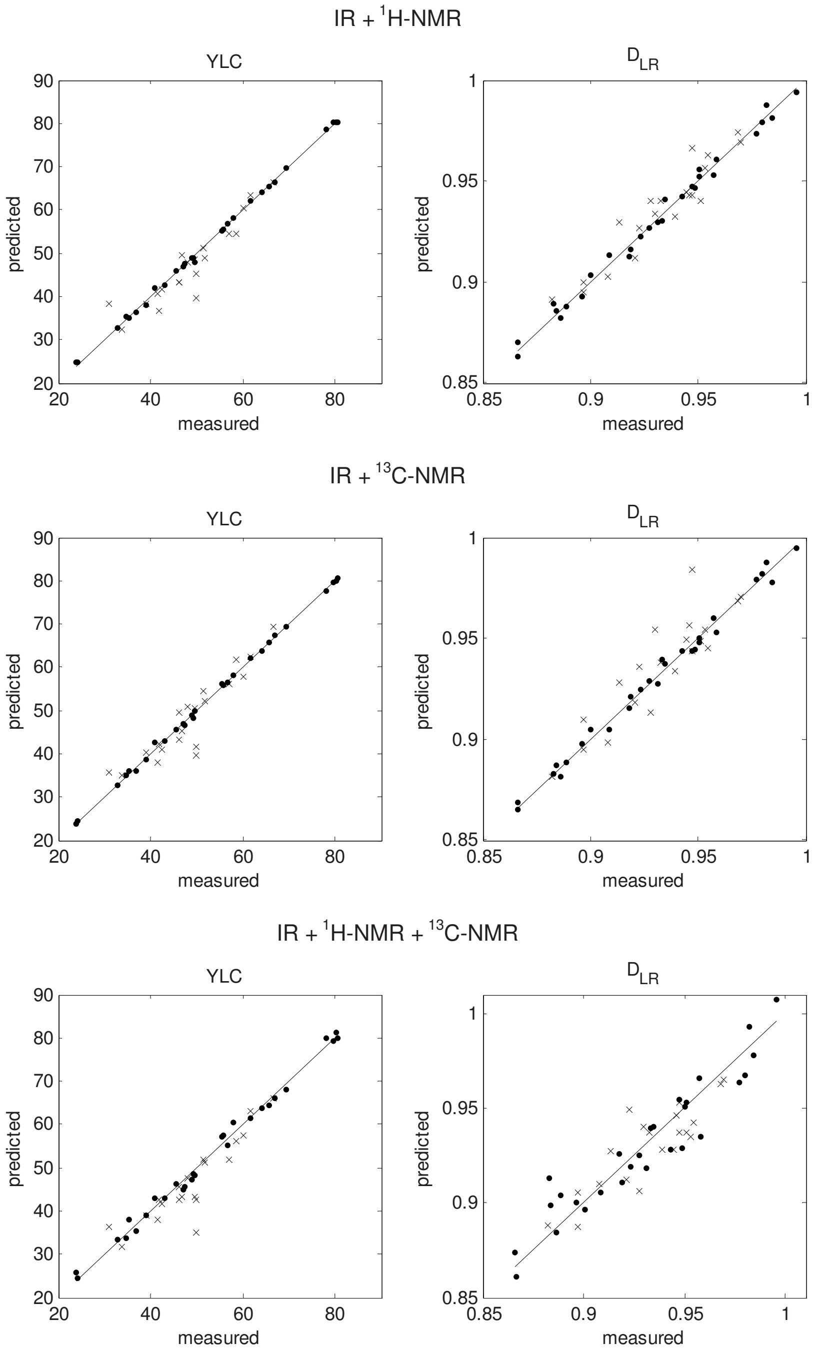


Fig. 5. Correlation plots of the PLS models to predict the LR-properties YLC and D_{LR} of crude oils. Models are based on the merged IR + ^{1}H-NMR, IR + ^{13}C-NMR and IR + ^{1}H-NMR + ^{13}C-NMR spectra of 28 calibration samples (•) and 20 validation samples (x).

Table 3. Results of PLS models to predict the LR properties of crude oils V1-V20 from merged IR, ^{1}H-NMR and/or ^{13}C-NMR spectra. Models were built from the spectra of samples C1-C28.

Input		LR-property						
		YLC (wt %)	D_{LR} (g/mL)	V_{LR} (-)	S (wt %)	PP (°C)	Asph (wt %)	CR (wt %)
IR+^{1}H-NMR	RMSEP	3.7	0.008	1.3	0.54	13.0	3.1	2.2
IR+^{13}C-NMR	RMSEP	3.7	0.013	1.8	0.44	11.6	3.0	2.6
IR+^{1}H+^{13}C-NMR	RMSEP	4.7	0.012	1.9	0.65	10.3	3.1	2.8

In our opinion, a plausible explanation for the redundancy in the information of the crude oil spectra relates to the fact that these compounds consist of a huge amount of components of strongly related structures, *i.e.* aliphatic and aromatic carbon hydrogen compounds. As a consequence, the peak patterns in the IR and NMR spectra are not only extremely complex, but also very much representing the same type of information, i.e., C-H and C-C vibrations versus C and H chemical shifts. This strongly reduces the complementary and hence discriminating potential of IR and ^{1}H-NMR and ^{13}C-NMR. Besides, the fact that crude oils largely consist of the same type of hydrocarbon compounds also causes a large redundancy in spectral information. In other words, the vibrational C-H bands that dominate the IR spectra correlate with both the ^{1}H and ^{13}C shifts and therefore hardly provide different or additional information.

The results of PLS modeling the joined 10 best scores from PCA of the separate IR, ^{1}H-NMR and ^{13}C-NMR calibration sets are presented in Table 4. The obtained RMSEP-values show that modeling PCA scores instead of spectra is only slightly favourable if ^{13}C-NMR data are used. For that reason, we conclude that modeling the IR and NMR spectral information after PCA data reduction does not provide a significant improvement in the results of predicting LR properties, as compared to the models based on straightforward merging the original spectra.

Table 4. Results of the PLS models to predict the LR properties of crude oils V1-V20 from the joined PCA scores of a PCA of the IR, ^{1}H-NMR or ^{13}C-NMR spectra of the crude oils C1-28.

Input data (10 scores)		LR-property						
		YLC (wt %)	D_{LR} (g/mL)	V_{LR} (-)	S (wt %)	PP (°C)	Asph (wt %)	CR (wt %)
IR+^{1}H-NMR	RMSEP	3.8	0.009	1.5	1.00	10.5	2.8	2.0
IR+^{13}C-NMR	RMSEP	3.8	0.010	1.7	0.63	13.0	2.7	2.2
IR+^{1}H+^{13}C-NMR	RMSEP	3.7	0.010	1.6	0.89	11.2	2.7	2.2

3.4. Conclusions

PLS modeling of the IR, ^{1}H-NMR and ^{13}C-NMR spectra to predict the LR properties of crude oils offers a valuable and fast alternative for the comprehensive laboratory methods that are currently used for this purpose. Different from what was expected, models based on IR, ^{1}H-NMR or ^{13}C-NMR spectral data score about the same on the different properties. The same conclusion is drawn from merging IR spectra with ^{13}C-NMR and/or ^{1}H-NMR data as input for modeling. It demonstrates that the combination of complementary spectroscopic techniques does not automatically lead to better results. For all data sets, correlations at a level that is competitive with the reproducibility of the currently applied ASTM and IP methods have been obtained for the properties YLC, V_{LR}, D_{LR} and PP. The scores for S, Asph and CR are somewhat lower but still useful for indicative prediction. Regarding the slightly better results for IR spectroscopy and taking into account the absence of sample preparation and the straightforward way of data acquisition, IR is preferred over NMR for the prediction of LR properties of crude oils at site.

References

1. C.T. Mansfield, B.N. Barman, J.V. Thomas, A.K. Mehrotra, J.M. McCann, Anal. Chem. 71 (1999) 81R.
2. H.N. Giles (Ed.), *Significance of Tests for Petroleum Products*, 7th Edition (2003), ASTM International, 51.
3. www.astm.org.
4. www.americanenergyinstitutes.org.
5. K.A. Lintelmann, Anal. Chem. 67 (1995) 327R.

6. K.H. Michaelian, S.L. Zhang, R.H. Hall, J.T. Bulmer, Spectrochim. Acta A 57 (2001) 73.
7. L.C.T. Shoute, K.J. Schmidt, R.H. Hall, M.A. Webb, S. Rifai, P. Abel, P.H. Arboleda, A. Savage, J.T. Bulmer, G.R. Loppnow, Appl. Spectrosc. 56 (2002) 1308.
8. J. Jehlička, O. Urban, J. Pokorný, Spectrochim. Acta A 59 (2003) 2341.
9. K.H. Michaelian, R.H. Hall, J.T. Bulmer, Can. J. Anal. Sci. Spectrosc. 49 (2004) 43.
10. M. He, X. Lu, S. Yao, Y. Liu, W. Fan, Dizhi Keji Qingbao 24 (2005) 67.
11. E. Evans, B. Batts, N. Cant, Fuel 66 (1987) 327.
12. J.M. Jacobson, M.R. Gray, Fuel 66 (1987) 749.
13. P.W. Yang, H.H. Mantsch, L.S. Kotlyar, J.R. Woods, Energ. Fuels 2 (1988) 26.
14. H. Ganz, W. Kalkreuth, Erdoel & Kohle, Erdgas, Petrochem. 43 (1990) 116.
15. E.A. Cloutis, AOSTRA J. Research 6 (1990) 17.
16. S. Kokot, N. Asker, Proc. 8th Int. Conf. Fourier Transform Spectrosc., SPIE- Int. Soc.Opt. Engin. 1575 (1992) 495.
17. S. Kokot, N. Asker, Proc. 6th Int. Conf. Near Infr. Spectrosc. 22 (1995) 1994.
18. N. Pieri, J. Kister, L. Quoniam, J.P. Planche, L. Germanaud, Polycycl. Arom. Comp. 9 (1996) 29.
19. H. Chung, M.-S. Ku, J.-S. Lee, Vibr. Spectrosc. 20 (1999) 155.
20. K.H. Michaelian, S. Zhang, R.H. Hall, J.T. Bulmer, Can. J. Anal. Sci. Spectrosc. 46 (2001) 10.
21. B. Pavoni, N. Raido, R. Plassa, S. Frignani, Ann. Chim. 94 (2004) 521.
22. W.I. Friesen, K.H. Michaelian, Y. Long, T. Dabros, Am. Chem. Soc., Div. Fuel Chem. 50 (2005) 228.
23. T.A. Adedosu, O.O. Sonibare, J. Appl. Sci. 5 (2005) 906.
24. H. Yuan, X. Chu, H. Li, Y. Xu, Fuel 85 (2006) 1720.
25. K.J. Johnson, R.E. Morris, S.L. Rose-Pehrsson, Energ. Fuels 20 (2006) 727.
26. K.H. Michaelian, R.H. Hall, K.I. Kenny, Spectrochim. Acta, A 64 (2006) 703.
27. R.R. Coelho, I. Hovell, M.B. de Mello Monte, A. Middea, A. Lopes de Souza, Fuel Process. Techn. 87 (2006) 325.
28. Y. Hongfu, C. Xiaoli, L. Haoran, X. Yupeng, Fuel 85 (2006) 1720.
29. J.M. Brown, C.J. Chrostowsky, PCT Int. Appl. (2006) 2006023800.
30. P. de Peinder, T. Visser, D.D. Petrauskas, F. Salvatori, F. Singelenberg, F. Soulimani, B.M. Weckhuysen, Appl. Spectrosc. 62 (2008) 414.
31. P. de Peinder, T. Visser, D.D. Petrauskas, F. Salvatori, F. Singelenberg, B.M. Weckhuysen, PCT Int. Appl. (2008) 135411.
32. A. Espinosa, A. Martens, G. Ventron, C.D. Lambert, A. Pasquier, Eur. Pat. Appl. (1989) 305090.
33. D.A. LeFebre, L.M. Lane, US patent (1995) 5430295.
34. S. Bages, B. Descales, D. Lambert, J.-R. Llinas, A. Martens, S. Osta, M. Sanchez, PCT Int. Appl. (1996) 9611399.
35. Z. Sikora, W. Salacki, Petrol. Coal 38 (1996) 65.

36. R. Van Doorn, F.H. Hooyman-Spaargaren, R.J. Neugebauer, L.M. Smeets, PCT Int. Appl. (1997) 9714953.
37. J. Kelly, D. Lambert, A. Martens, PCT Int. Appl. (1998) 9836274.
38. B. Descales, D. Lambert, J.-R. Llinas, A. Martens, S. Osta, M. Sanchez, S. Bages, US Patent (2000) 5712797.
39. M. Blanco, S. Maspoch, I. Villarroya, X. Peralta, J.M. González, J. Torres, Appl. Spectr. 55 (2001) 834.
40. S. Macho, M.S. Larrechi, Trends Anal. Chem. 21 (2002) 799.
41. R.J. Mikula, V.A. Munoz, N. Wang, B. Bjornson, D. Cox, B. Moisan, K. Wiwchar, J. Can. Petr. Techn. 42 (2003) 50.
42. S. Satya, R.M. Roehner, M.D. Deo, F.V. Hanson, Energ. Fuels 21 (2007) 998.
43. H. Chung, Appl. Spectrosc. Rev. 42 (2007) 251.
44. M.V. Reboucas, E.C. Santos, F.S.V. Vieira, Vibr. Spectrosc. 44 (2007) 187.
45. J.G. Speight, Fuel 49 (1970) 76.
46. M.H.B. Hayes, M. Stacey, J. Standley, Fuel 51 (1972) 27.
47. M.U. Hasan, M.F. Ali, A. Bukhari, Fuel 62 (1983) 518.
48. M.U. Hasan, M.F. Ali, A. Bukhari, Fuel 64 (1985) 839.
49. J.M. Jacobson, M.R. Gray, Fuel 66 (1987) 753.
50. C. Giavarini, C. Vecchi, Fuel 66 (1987) 868.
51. J. Thiel, M.R. Gray, AOSTRA J. Research 4 (1988) 63.
52. C.H. Tsai, M.D. Deo, F.V. Hanson, A.G. Oblad, Fuel Sci. Technol. Int. 10 (1992) 1437.
53. L. Michon, D. Martin, J.-P. Planche, B. Hanquet, Fuel 76 (1997) 9.
54. J. Bryan, K. Mirotchnik, A. Kantzas, J. Can. Petrol. Technol. 42 (2003) 29.
55. F. Ali, Z. H. Khan, N. Ghaloum, Energ. Fuels 18 (2004) 1798.
56. T. Zhang, Y. Liang, B. Li, H. Cui, F. Gong, Anal. Sci. 20 (2000) 717.
57. H. Swierenga, F. Wulfert, O.E. de Noord, A.P. de Weijer, A.K. Smilde, L.M.C. Buydens, Anal. Chim. Acta 411 (2000) 121.
58. Q. Abbas, N. Dupuy, C. Rebufa, L.Vrielynck, J. Kister, A. Permanyer, Appl. Spectrosc. 60 (2006) 304.
59. H. Chung, S. Cho, Y. Toyoda, K. Nakano, M. Maeda, Analyst 131 (2006) 684.
60. R. Siuda, G. Balcerowska and D. Aberdam, Chemometrics Intell. Lab. Systems 40 (1998) 193.
61. A.M. Weljie, J. Newton, P. Mercier, E. Carlson, C.M. Slupsky, Anal. Chem. 78 (2006) 4430.
62. J. Forshed, R.J.O. Torgrip, K. Magnus Åberg, B. Karlberg, J. Lindberg, S.P. Jacobsson, J. Pharm. Bio. Anal. 38 (2005) 824.
63. M. Defernez, I. J. Colquhoun, Phytochem. 62 (2003) 1009.
64. S. Halouska, R. Powers, J. Magn Reson. 178 (2006) 88.

65. D. Lin-Vien, N.B. Colthup, W.G. Fateley, J.G. Graselli, (Eds), *Handbook of Infrared and Raman Characteristic Frequencies of Organic Molecules*, Academic Press Inc., New York (1991).
66. Statoil (2005). Commercial assays of crude oils from different regions in the North Sea. Oslo: Statoil (http://www.statoilhydro.com/en/ouroperations/tradingproducts/crudeoil/crudeoilassays/pages/ crudeoilassay.aspx).
67. V. Schoonjans, D.L. Massart, J. Pharm. Bio. Anal. 26 (2001) 225.
68. P.J. de Groot, H. Swierenga, G.J. Postma, W.J. Melssen, L.M.C. Buydens, Appl. Spectrosc. 57 (2003) 642.
69. L.P. Brás, S.A. Bernardino, J.A. Lopes, J.C. Menezes, Chemom. Intell. Lab. Syst. 75 (2005) 91.
70. R. Karoul, A.M. Mouazen, É. Dufour, L. Pillonel, E. Schaller, D. Picque, J. De Baerdemaeker, J.-O. Bosset, Eur. Food Res. Technol. 223 (2006) 44.

Chapter 4

Temperature Effects on the Performance of the IR Prediction Models

Abstract

Research has been carried out to the effect of the temperature on the results of predicting long residue (LR) and short residue (SR) properties of crude oils from their IR spectra. Two effects on the results of partial least squares (PLS) modeling of the spectra have been studied: 1) the exposure of the crude oils to 65℃ for 24 h prior to recording the IR spectra, and 2) recording the spectra of the 65℃ pre-exposed samples at 10 different temperatures between 20 and 65℃. It is concluded that pre-exposure of crude oils to 65℃ to reduce the amount of volatile constituents does not result in a significant improvement in the results of predicting the LR and SR properties as compared to the models based on the IR spectra of the untreated materials. Modeling of the spectra recorded as a function of the temperature between 20 and 65℃ revealed a linear relation with the temperature, which is attributed to a decrease of the density with increasing temperature. Pre-processing of the spectra by mean centering and a subtraction step showed that the root mean square error of prediction (RMSEP) values improved for the LR-properties density ($D_{LR)}$, viscosity (V_{LR}), sulfur (S) and carbon residue (CR), but worsened for the LR properties yield long on crude (YLC), Pour point (PP) and asphaltenes (Asph). The effect on the results of predicting the SR properties was found to be negligibly small.

4.1. Introduction

In the Chapters 2 and 3, the results have been described of research to the potential of molecular spectroscopic methods combined with partial least squares (PLS) regression to predict the quality of crude oils for bitumen application from a single spectrum. Aim of these studies was to determine the value of such an integrated approach as a fast alternative for the elaborate and time-consuming laboratory tests which are currently used in oil industries to determine the so-called long residue (LR) and short residue (SR) properties of crude oils. Six standard spectroscopic techniques that can be applied on a routine basis have been investigated, i.e., Raman, UV-Visible, IR, NIR and ^{1}H- and ^{13}C-NMR spectroscopy. The best results to predict LR and SR properties were obtained from models based on IR and NIR spectra, although in accordance with theory, the performance of both techniques was much the same due to the redundancy in spectral information [1, 2].

The underlying idea to explore molecular spectroscopic methods to determine LR and SR properties was that the spectral data obtained with these techniques contain information on structures and interactions on a molecular scale. This level of detail is important since all molecules in a crude oil will affect each other by a mechanism of attractive and repulsive intra- and intermolecular interactions. Obviously, the temperature plays an important role in this process. For that reason, data acquisition in the previous studies has been carried out at a carefully controlled temperature of 20°C. Besides, a change of the temperature of a crude oil affects the inter- and intramolecular interactions and hence (some of) its physicochemical properties [3]. Crude oil features that are known to be temperature sensitive are, for instance, the gel-point, pour point and viscosity, parameters which are highly important in pipeline transportation [4-8]. For that reason, heating experiments are often carried out on crude oils to obtain a thermal fingerprint or to determine the amount of paraffins and waxes [9, 10]. Furthermore, the temperature is known to affect adhesion phenomena between a crude oil and brine, injected brine and rock surfaces that hampers the extraction of the oils from soil [11] or promotes its sorption [12].

For obvious reasons, a change in the inter- and intramolecular interactions of a crude oil as result of a temperature increase will also affect its spectral characteristics [13, 14]. Such differences might be useful to extract additional information on crude oil LR and SR properties, thus offering a way to improve the prediction models. Therefore, it was decided to carry out modeling of IR spectra recorded over a temperature range from 20°C to 65°C. Additionally, it was chosen to apply a pretreatment of the samples by exposing the crude oils to 65°C for 24 h in open air to reduce the spectral contributions of the volatile constituents thus coming closer to the characteristics of a long residue fraction. The reason is that knowledge of LR and SR properties is particularly important for bitumen applications, where the heavy components prevail over the volatile ones. A pretreatment at elevated temperature will reduce the influence of the lighter fractions and, as a consequence, might lead to better bitumen prediction models. This chapter describes the results of this study.

4.2. Experimental

4.2.1. Materials

Modeling experiments have been carried out on the same collection of samples as used in the previous and which are listed in Table 1 of Chapter 2. For the prediction of the LR properties, the samples C1-C28 were used for calibration and V1-V20 for validation. 43 samples of the total set were used to build prediction models for the SR properties, since these data were only known for a smaller number of compounds. For that reason, a division into a calibration and a validation set was not made for this part of the study. All compounds have been stored and prepared as described in Chapter 2. The effect of an extra heat pre-treatment on the results of predicting the LR and SR properties was studied first. For this purpose, 20 mL aliquots of each sample were brought into an open 100 mL beaker glass. After precisely determining the starting weight, the temperature was increased to 65℃ at a ramp of 1℃/min to reduce the amount of volatile species in the sample. After 24 h at 65℃, the sample was covered with a watch glass and allowed to cool down to ambient temperature (20℃). Finally, the weight loss was determined and then the spectrum of the residue was recorded at 20℃.

Next, the effect of modeling the IR spectra of crude oils, recorded at different temperatures, for the LR and SR properties was carried out. Hereto, the spectra of the crude oils, previously exposed to 65℃ for 24 h, were recorded over a temperature range from 20℃ up to 65℃ with steps of 5℃, resulting in 10 spectra per sample. After each step, the temperature of the sample was kept constant for 2 min to allow scanning at the specified temperature.

The following LR properties were studied; the long residue yield-on-crude (YLC), density (D_{LR}), viscosity (V_{LR}), sulfur content (S), pour point (PP), asphaltenes (Asph) and carbon residue (CR). As input values for these properties, the data obtained with the following ASTM methods were used: D

2892 (YLC), D 4052 (D_{LR}), D 445 (V_{LR}), D2622 (S), D 97 (PP) and D 4530 (CR), while the values for Asph were determined by method IP 143. As SR properties, the penetration (P), softening point or ring and ball (R&B), viscosity (V_{SR}) and density (D_{SR}) were considered as the main properties. As input values for the SR properties, the data obtained with the ASTM methods D 5 (P), D 36 (R&B), D 4052 (D_{SR}) and D 445 (V_{SR}) were used.

4.2.2. Spectroscopic data acquisition

IR measurements were carried out on a Bruker Tensor-27 FTIR spectrometer equipped with a deuterated triglycine sulfate (DTGS) detector. Samples were analyzed with a horizontal ATR accessory (FastIR^{TM}, Harrick) with a heating cover plate and ZnSe as the optical element. The cover plate was used to achieve homogeneous heating and to prevent sample evaporation during measurement. Data acquisition was performed by co-adding 50 scans/spectrum at a spectral resolution of 4 cm^{-1} with apodization set to 'medium'. The sample compartment of the instrument was flushed with dry air to reduce interference of H_2O.

4.2.3. Chemometrics

Since the temperature variation results in a data cube with the crude oil sample, wavenumber and temperature as the three dimensions, PARAFAC (parallel factor) modeling was tested [15-17]. However, it appeared that the complexity of the crude oil spectra prevented a straightforward application of this multi-way method. Instead of using the combination of the PARAFAC scores with MLR (multiple linear regression) for quantification [18] it was decided to unfold the multi-way array to a matrix and perform PCA combined with PLS.

Partial Least Squares (PLS) modeling was carried out with the PLS Toolbox (Eigenvector Research, Inc.) for MatLab (The MathWorks, Inc.). Prior to modeling, the IR spectra were pre-processed by first taking the 1st derivative (25 pt. Savitzky-Golay smoothing), followed by Multiplicative Signal Correction (MSC) and Mean Centering (MC) of the data. Pre-processing parameters were chosen based on prior knowledge for each spectroscopic technique combined with trial and error. The spectral region 3200-670 cm^{-1} was used as input, since it was concluded from the previous studies that this interval led to the best results. The obtained error values (Root Mean Square Error of Cross Validation (RMSECV) of the prediction models were based on ”leave one out” (LOO) cross validation (CV). The number of latent variables (LV), used for a PLS model, was based on the minimum value of the RMSECV.

The study on the effect of the 65℃ temperature pre-treatment on the results of modeling was carried out on the same set of 28 spectra for calibration and 20 spectra for validation as used in Chapter 3. The study on PLS-modeling of the spectra recorded as a function of the temperature for the LR and SR properties was initially performed on the same calibration and validation sets. However, the modeling results indicated that sample C22, the heaviest crude of the collection, was an outlier. Visual inspection of the corresponding spectra revealed extreme water adsorption at the ATR crystal during measurement. Repeating the experiment for this crude oil showed the same anomalous behaviour and for that reason it was removed from the calibration set.

Next, modeling of the remaining 27 spectra for the 7 different LR properties was performed in two steps. First, for each crude oil, the mean spectrum of the 10 spectra recorded between 20 and 65℃ was calculated and subtracted from the variable temperature spectra. This mean spectrum was subtracted from each of the 10 individual spectra in order to exhibit no more than the spectral changes due to temperature change. Next, principal component analysis (PCA) was performed on all 470 (47x10) pre-processed spectra (1^{st} derivative, 25 pt. Savitzky-Golay smoothing, followed by MSC and MC). From this PCA, the scores 1-3 of the 3 spectra recorded at 20, 40 and 60℃ were added to the PCA scores 1-10 of the spectrum of the same crude but without pre-treatment and recorded at 20℃. This resulted in 19 data points (variables) for each sample on which PLS-modeling was applied: 10 representing information on the crude oils recorded without pre-treatment at 20℃, and 9 (3 x 3) for the variation due to the temperature. It should be noted that the first 3 scores from the 20, 40 and 60℃ spectral data were taken deliberately, since more scores appeared to deteriorate the RMSECV values for all LR properties. Finally, the 7 prediction models obtained were tested by submitting the 20 spectra of the validation set. For each model, the results have been expressed as root-mean-square-error-of-prediction (RMSEP) value.

Modeling to predict the SR properties P, R&B, D_{SR} and V_{SR} by PLS was performed on the IR spectra of 43 crude oils for which SR properties were known at one or more AFT-values. Due to the limited number of spectra, LOO cross validation has been applied instead of creating independent calibration and validation sets. The same procedure as for modeling for the LR properties was used, but the available AFT values were added as input, resulting in 20 data points (variables) that were PLS modeled versus the SR properties. The spectral data matrix X was created by a PCA on the available IR spectra of the untreated samples after pre-processing (baseline correction, MSC and MC). The first 10 scores of this PCA model were used to describe each spectrum. In the next step the same 9 scores (3x3) that were used for the PLS modeling of the LR properties were added. The AFT values, if available, were then added to the scores of the spectra (10) and the temperature series (9) resulting in 20 variables, which were variance scaled for each crude oil. PLS prediction models were made for the SR properties P, R&B, D_{SR} and V_{SR} at various flash temperatures. The ASTM method to determine the P-property is not able to determine values higher than 200. For these compounds with a higher P value, these were set to 200 to include as much data as possible. The different data sets consisted of 36 (P), 40 (R&B), 43 (D_{SR}) and 41 (V_{SR}) IR spectra with 103, 123, 134 and 122 properties respectively. The corresponding SR properties were used as the Y matrix in the PLS regression. The same procedure was followed using the yield-short-on-crude (YSC) property instead of the AFT values as additional input of the data matrix.

4.3. Results and discussion

The weight loss of the different samples as result of exposure to a temperature of 65℃ during 24 h in open air is depicted in Fig. 1. The weight losses vary from 43% for the light crude oils (*e.g.* sample C3) to 1% for the very heavy and highly viscous samples (*e.g.* sample C22).

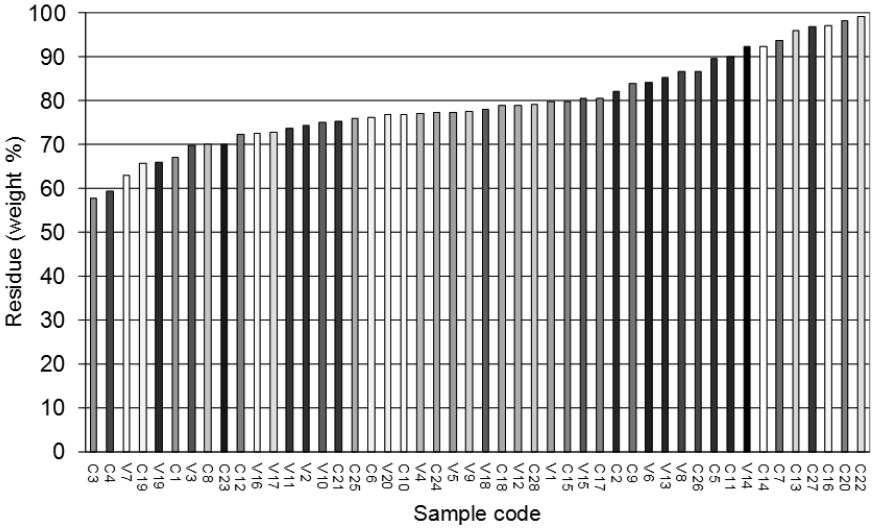


Fig. 1. Weight loss of crude oils as result of exposure to 65℃ during 24 h.

An effect on the IR spectra of the heavy crude oils is virtually absent. However, for the light samples, changes in the band pattern are clearly visible, particularly in the region of the C-H stretching vibrations around 3000 cm^{-1}. This is illustrated in Fig. 2 by the spectra of sample C2 before and after exposure to 65℃. As can be seen, the bands at around 2955 and 2870 cm^{-1}, which can be assigned to CH_3 anti-symmetric and symmetric stretching vibrations, respectively, have been decreased in intensity compared to the CH_2 stretching bands at 2925 and 2850 cm^{-1}. This effect can be largely attributed to the relatively higher volatility of short and branched aliphatic compounds, compared to the longer paraffin-like species. In addition a difference is observed in the region of the out-of-plane vibrations of aromatic compounds below 800 cm^{-1}.

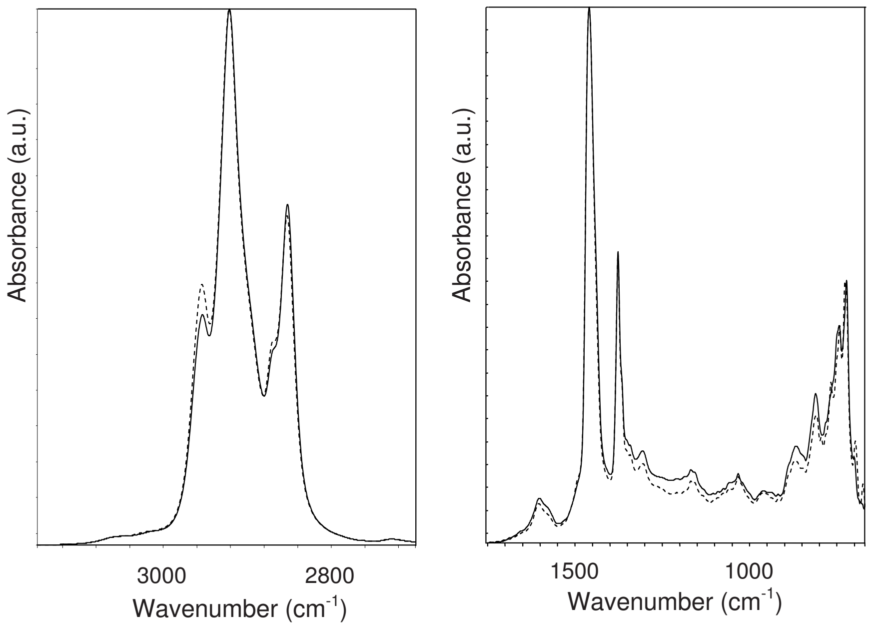


Fig. 2. IR spectra of sample C2 before (….) and after (—) exposure to 65°C (24h).

4.3.1. The effect of exposure to 65°C on the prediction of the LR properties

The results of modeling the IR spectra recorded at 20°C after exposure to 65°C for 24 h for the LR properties are shown in Table 1. The corresponding prediction plots are presented in Fig. 3. For reasons of comparison, the results of modeling the IR spectra of the same 28+20 crude oils without the 65°C temperature pre-treatment as previously reported in Chapter 3, have been included in Table 1. Comparison of the RMSEP-values shows no significant improvement for the samples as a result of the exposure to 65°C. Therefore, it is concluded that a reduction of the volatile constituents prior to recording the IR spectra, does not lead to better results when predicting LR properties. For the YLC property the prediction even is significantly worse. This is probably due to the fact that before the pre-treatment a spectrum contains information on the complete crude whereas after the heating the volatile compounds are removed. Measuring the ratio of the volatiles and the long residue will be advantageous for the prediction of the yield of long residue on a crude oil.

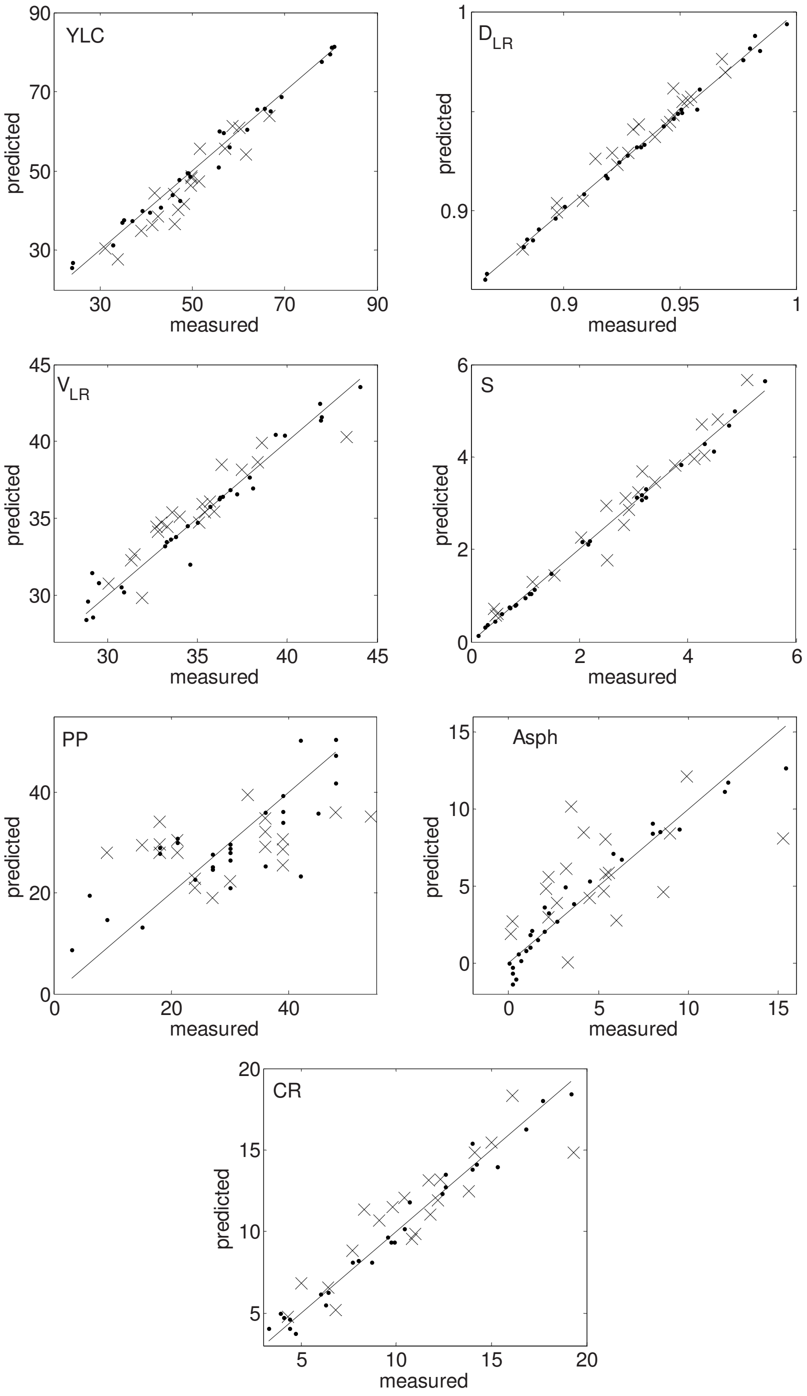


Fig. 3. PLS prediction plots for the LR properties of crude oils from spectra recorded at 20°C after exposure to 65°C for 24 h. Calibration spectra (•) and validation spectra (x).

Table 1. PLS modeling results of the LR properties based on the IR spectra of 28 calibration and 20 validation samples before and after exposure to 65°C for 24 h.

LR property	Mean value	With 65°C pretreatment			Without 65°C pretreatment		
		RMSECV	LV's	RMSEP	RMSECV	LV's	RMSEP
YLC (wt %)	52.5	5.9	7	4.5	3.7	9	2.8
D_{LR} (g/ml)	0.930	0.005	7	0.007	0.012	8	0.007
V_{LR} (-)	35.3	1.3	6	1.4	2.0	6	1.3
S (wt %)	2.16	0.29	9	0.32	0.56	8	0.24
PP (°C)	29.6	10.1	4	10.7	9.2	5	10.9
Asph (wt %)	4.1	2.2	6	3.2	2.2	7	2.5
CR (wt %)	9.7	1.4	6	1.7	2.1	6	1.6

4.3.2. The effect of exposure to 65°C on the prediction of the SR properties

Similar to predicting the SR properties of the untreated crude oils, PLS modeling was carried out on the IR spectra of the samples, pre-exposed to 65°C, using the AFT and the YSC values as input for the X-matrix. Also, LOO cross validation was applied instead of independent calibration and validation sets. The results of prediction of the SR models based on the IR spectra of the 43 available crude oils when using the AFT values as input are shown in Table 2. The corresponding prediction plots are presented in Fig. 4. Again, for reasons of comparison, the results obtained from the SR prediction models of the untreated samples have been included in Table 2. Note that the latter data originate from Chapter 2, since the prediction of SR properties from combined IR and NMR data have not been studied in Chapter 3. Due to the smaller data set and the applied cross validation instead of using separate calibration and validation data sets, only RMSECV-values are given. Comparison of the RMSECV-values in Table 2 reveals no significant improvement as result of the high temperature pre-treatment. The SR properties D_{SR} and V_{SR} are predicted well, while the Log P and Log R&B perform reasonably well. Similar results have been acquired from the SR models when using the YSC values as input for the X-matrix. This is shown in Table 3 and the corresponding prediction plots in Fig. 5. RMSECV-values obtained for the untreated samples have been included in Table 3. As for using the AFT-values as additional input, there is hardly any difference compared to the results obtained for the non pre-treated samples. The scores for the SR properties D_{SR} and V_{SR} are better than for Log P and Log R&B, while the overall performance when using AFT as extra input is better than when YSC input is applied. Closer examination of the log P plots in Fig. 4 and 5 revealed a distinctive pattern at the measured Log P-values of 2.3. This is due to the fact that P-values exceeding 200 units have been automatically set to 200 by the physicochemical method used. Obviously, using these data in the Y-matrix has a negative effect on the accuracy of the prediction model for this property. However, leaving out these values from modeling would improve the accuracy of the predictions of the P-values smaller than 200, but it also would reduce the general applicability of the model.

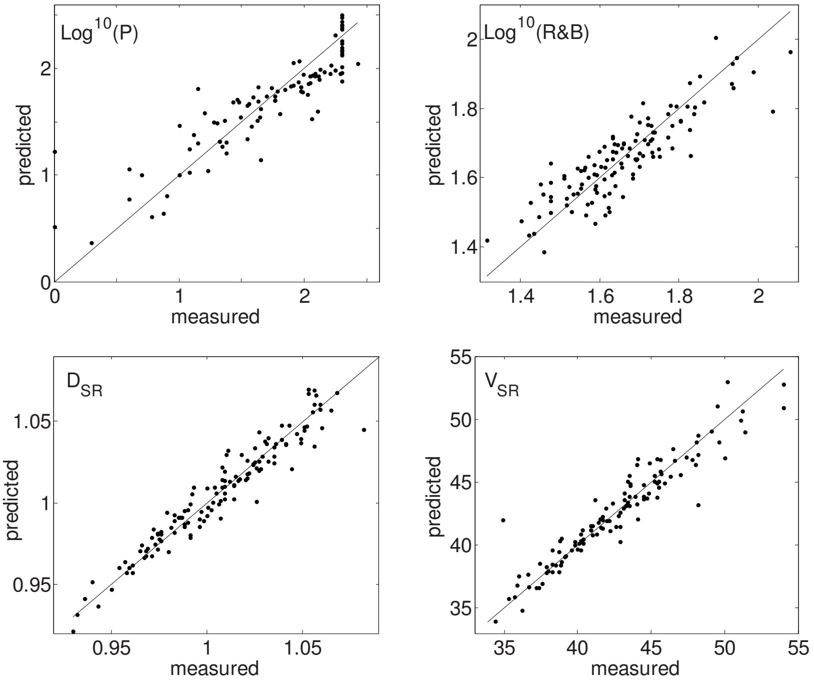


Fig. 4. Prediction plots of PLS modeling of the SR properties based on the IR spectra of crude oils pre-exposed to 65℃, using AFT values as additional input.

Table 2. Results of PLS-modeling to predict SR properties of crude oils based on IR spectra of crude oils before and after pre-exposure to 65℃, using AFT values as additional input.

SR-property	Mean value	With 65℃ pretreatment		Without 65℃ pretreatment	
		RMSECV	LV's	RMSECV	LV's
Log^{10} P (0.1 mm)	1.73	0.32	3	0.32	3
Log^{10} R&B (℃)	1.66	0.08	3	0.09	3
D_{SR} (g/ml)	1.008	0.013	3	0.015	3
V_{SR} (-)	42.7	1.7	3	1.8	3

In order to improve the applicability and reliability of the models, additional physicochemical analysis was carried out for the 8 compounds with P values above 200 compounds to obtain more accurate input data. The same was done for 6 samples with assumed inaccurate values for R&B. The results of modeling the IR spectra of the untreated samples before and after correction of these P and R&B values, using AFT values as additional input are listed in Table 4. The corresponding prediction and LV plots 1-3 are shown in Fig. 6.

Table 3. Results of PLS-modeling to predict SR properties of crude oils based on IR spectra of crude oils before and after pre-exposure to 65℃, using YSC values as additional input.

SR-property	Mean value	With 65℃ pretreatment		Without 65℃ pretreatment	
		RMSECV	LV's	RMSECV	LV's
Log^{10} P (0.1 mm)	1.73	0.41	3	0.39	3
Log^{10} R&B (℃)	1.66	0.10	3	0.08	3
D_{SR} (g/ml)	1.008	0.017	3	0.020	3
V_{SR} (-)	42.7	2.6	3	2.1	3

Table 4. Results of PLS modeling the SR properties Log P and Log R&B using the AFT-values as additional input before and after correction of P and R&B reference values.

	Before correction			After correction		
SR-property	Mean value	RMSECV	LV's	Mean value	RMSECV	LV's
Log^{10} P (0.1 mm)	1.73	0.32	3	1.83	0.37	3
Log^{10} R&B (℃)	1.66	0.09	3	1.65	0.10	3

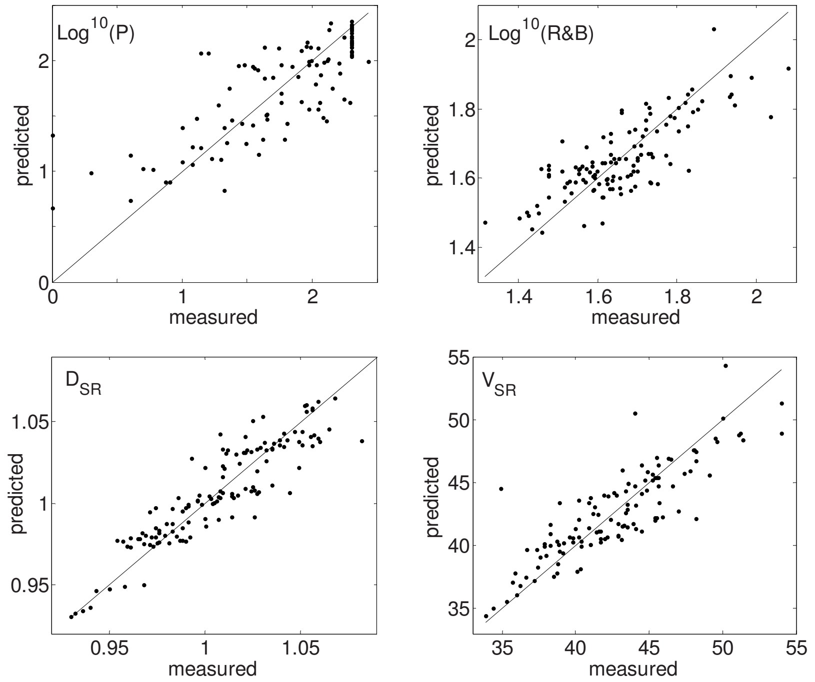


Fig. 5. PLS prediction plots of the SR properties based on the IR spectra of crude oils pre-exposed to 65 °C, using YSC values as additional input.

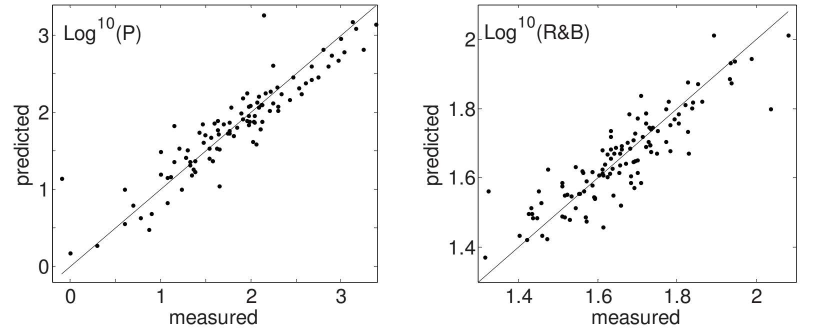


Fig. 6. Results of modeling the SR property Log P and Log R&B as function of AFT with improved reference data.

Obviously, as can be seen from Table 4, modeling with the corrected P and R&B values as input resulted in minor changes of the mean property values and the RMSECV-values. It could be argued that the P model as well as the R&B model has slightly decreased. However, the range of the P model has increased significantly from a maximum of 2.4 to 3.4, which corresponds with

P values of 269 and 2440 and this extension could be very valuable for softer crudes.

4.3.3. The effect of modeling spectra recorded at 20-65 ℃ on the prediction of the LR properties

As a check to determine leakage or sample change during the measurements at elevated temperature, after each cycle, the spectra recorded prior to heating in the ATR-cell and after cooling back to 20 ℃ were compared for each sample. As illustrated in Fig. 7 for sample V2, the spectra are virtually identical. Besides, leakage was not observed in the accessory.

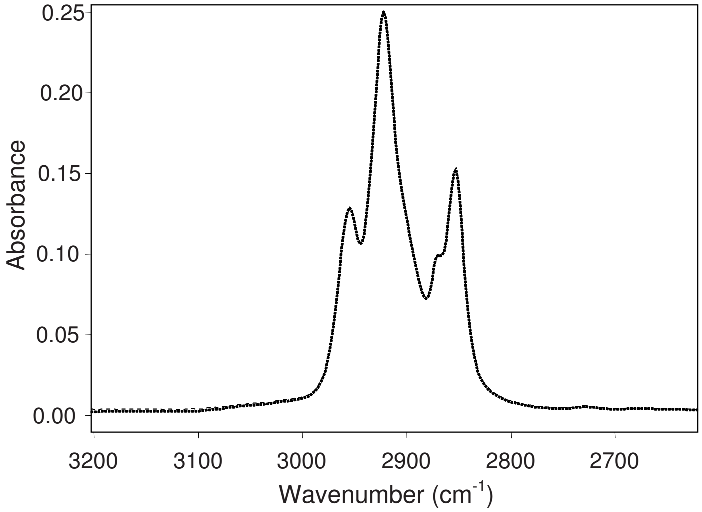


Fig. 7. Overlay of the C-H stretching region of the pre-exposed sample V2 recorded in the ATR-cell, prior to heating to 65 ℃ (dotted line) and after cooling down to ambient temperature (plain line).

The variable temperature spectra of calibration sample C22 are presented in Fig. 8 and the extreme water adsorption at the ATR crystal surface during measurement is evident. The two broader peaks at 3400 and 1640 cm^{-1} belong to OH stretching and bending vibrations of water, respectively. These peaks are absent in the variable temperature spectra of all other crude oils. It illustrates the fact that the sample is considered to be an outlier and why it has been removed from the calibration set.

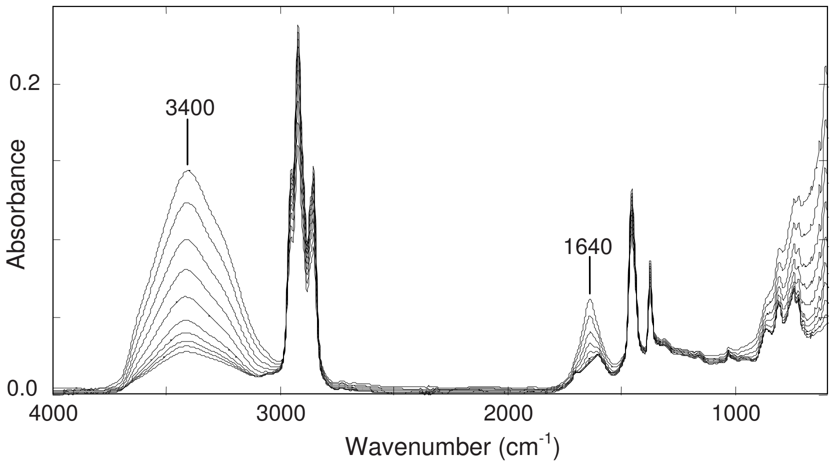


Fig. 8. Variable temperature spectra (20-65℃) of sample C22, pre-treated at 65℃.

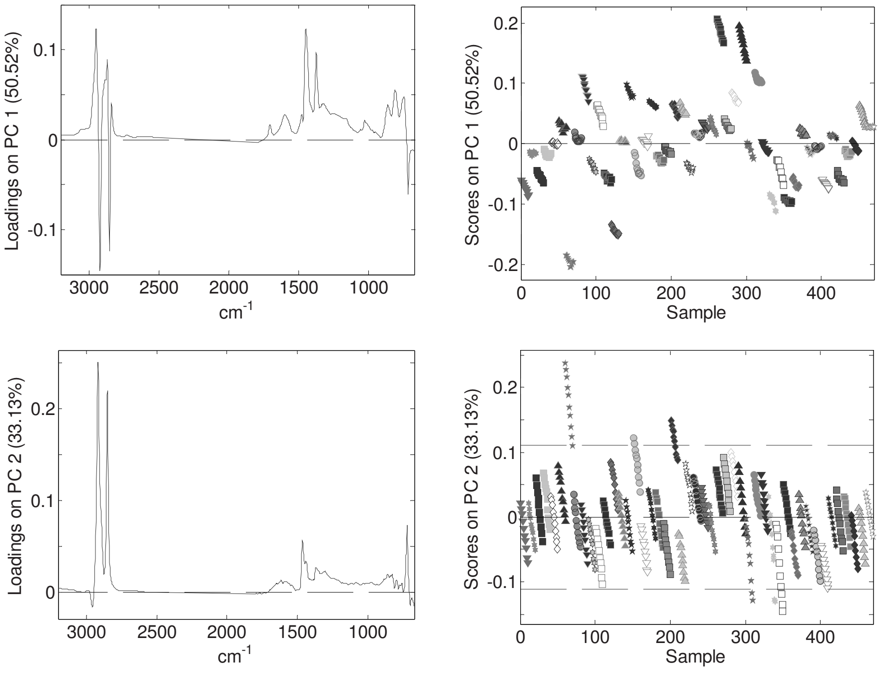


Fig. 9. PCA of spectra recorded at 10 different temperatures for 47 crude oils: loadings and corresponding scores for PC1 (top) and PC2 (bottom).

The results of modeling the spectra of the remaining 47 crude oils recorded at 10 different temperatures between 20 and 65℃ for the different LR properties revealed an almost linear relationship with the temperature for all samples. A PCA of the dataset of the 470 spectra after MC gave a rather constant decrease of the score values with increasing temperature. This is illustrated in Fig. 9, showing the scores for PC1 and PC2 for the 47 samples at 10 different temperatures. Similar trends were observed for PC3 and PC4.

Next, in order to study the spectral differences as a function of temperature and its potential for LR prediction purposes, the spectra recorded at the 10 different temperatures were mean centred by subtracting the mean spectrum of the 10 temperature spectra for each sample. This step ensures that the differences represented by the spectra are due to the temperature change and not to the overall crude composition. The results of this pre-processing are shown in Fig. 10.

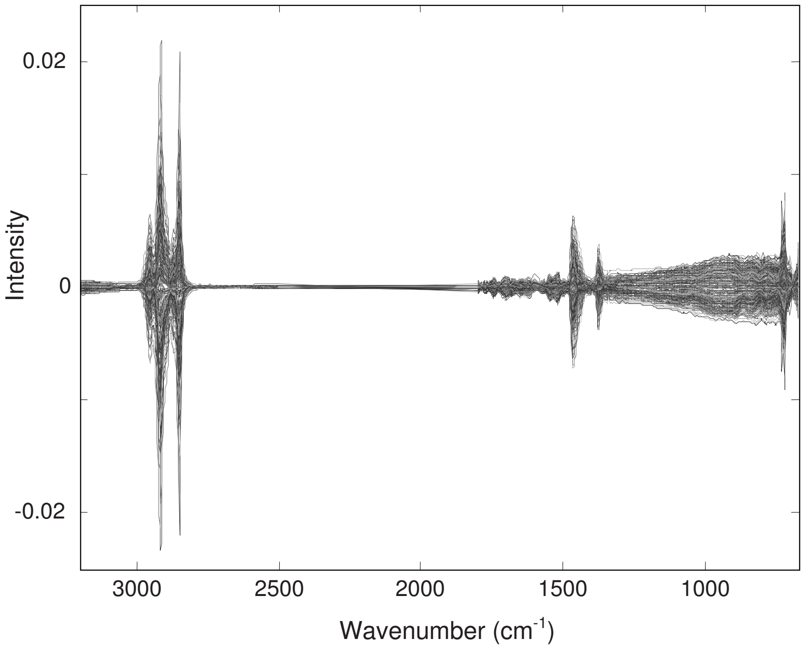


Fig. 10. Mean centered difference spectra of 47 crude oils recorded in the temperature range 20-65℃. Spectra were obtained by subtraction of the mean spectrum from the spectra recorded at 10 different temperatures.

The peaks pointing up correspond to the spectra recorded at low temperature and the peaks pointing down to the high temperature data.

A PCA of the temperature MC corrected data provided the results illustrated in Fig. 11, showing the loading of PC1 describing 93.5% of the variation and the corresponding scores plot on PC1. Obviously, this observation can be largely attributed to a decrease in the density of the samples, proportional to the increase of the temperature. It indicates that the predictive power of such models will be limited.

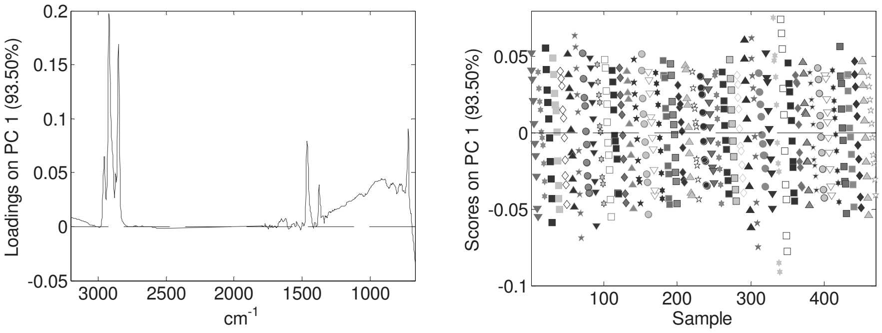


Fig. 11. Results of mean centred difference spectra of 47 crude oils recorded in the temperature range 20-65℃. The spectra were obtained by subtraction of the mean spectrum from the spectra recorded at 10 different temperatures.

However, closer examination reveals that different samples exhibit slightly different patterns, suggesting that some discriminative and hence predictive potential might be present. This is not confirmed by the RMSEP values obtained after PLS-modeling of the original spectra (*i.e.* from the samples not exposed to 65℃) combined with the spectra recorded at 20, 40 and 60℃ after data reduction. On the other hand it appears that the decreased number of used LV's, implicating more robust models, for D_{LR} and V_{LR} leads to comparable RMSEP values. These results have been summarized in Table 5 together with the prediction values of the models obtained for the 28 calibration and the 20 validation spectra of the untreated samples as previously presented in Chapter 2.

Table 5. Results of modeling spectra recorded at 20, 40 and 60℃ for the LR properties of 27 65℃ pretreated samples, compared to the results of the models built from the spectra of 28 untreated samples. The RMSEP is determined by using the 20 validation samples.

	Variable temperature data (27+20)				Original spectra (28+20)			
	Mean	RMSECV	LV's	RMSEP	Mean	RMSECV	LV's	RMSEP
YLC (wt %)	51.5	4.1	4	4.5	52.5	3.7	9	2.8
D_{LR} (g/ml)	0.928	0.015	3	0.008	0.930	0.012	8	0.007
V_{LR} (-)	35.0	2.1	3	1.4	35.3	2	6	1.3
S (wt %)	2.09	0.72	3	0.46	2.16	0.56	8	0.24
PP (℃)	28.9	11.5	4	10.9	29.6	9.2	5	10.9
Asph (wt %)	3.8	2.7	4	3.0	4.1	2.2	7	2.5
CR (wt %)	9.4	2.1	6	2.5	9.7	2.1	6	1.6

4.3.4. The effect of modeling spectra recorded at 20-65℃ on the prediction of the SR properties

The first 10 scores of the PCA of 47 untreated crude spectra were combined with the 9 scores obtained by PCA of the spectra recorded at 10 different temperatures (20-65℃) and the AFT values. Modeling results of these 20 variables against the different SR properties are shown in Table 6. The corresponding correlation plots are presented in Fig. 12. The results obtained for the spectra recorded at 20℃ without pre-exposure to 65℃ have been included in Table 6. Obviously, the same samples were used for both series of prediction models.

Table 6. Results of modeling spectra recorded at 20, 40 and 60℃ for the SR properties, using the AFT-value as additional input. Modeling based on corrected data for P and R&B; sample C22 excluded. Results for modeling of the spectra of the original samples have been added for comparison.

	Variable temperature data			Original spectra		
SR-property	Mean value	RMSECV	LV’s	Mean value	RMSECV	LV’s
Log^{10} P (0.1 mm)	1.86	0.33	3	1.83	0.37	3
Log^{10} R&B (℃)	1.65	0.09	3	1.65	0.10	3
D_{SR} (g/ml)	1.006	0.013	3	1.008	0.015	3
V_{SR} (-)	42.5	1.6	3	42.7	1.8	3

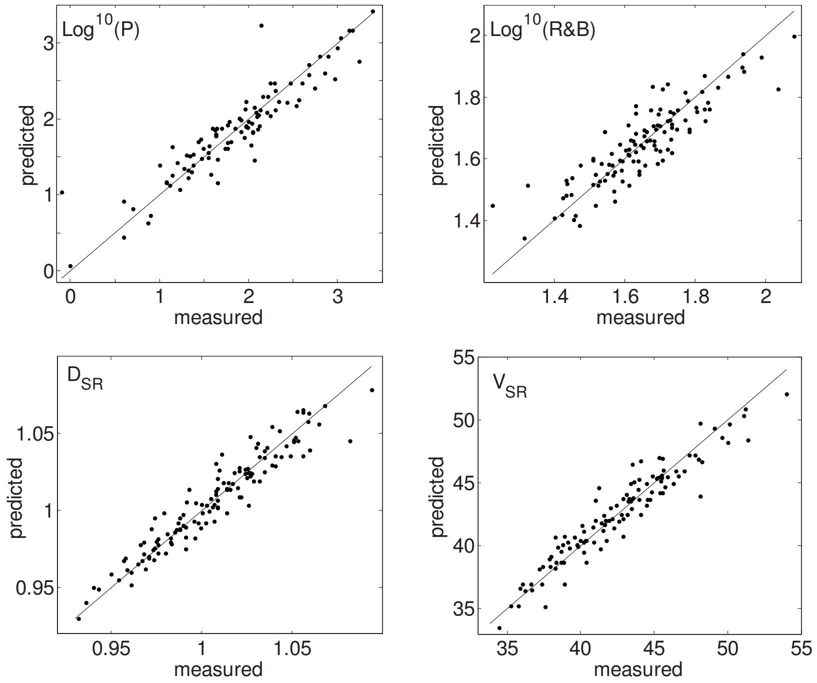


Fig. 12. Correlation plots of modeling the SR properties Log P, Log R&B, D_{SR} and V_{SR} as function of AFT using the IR data obtained at 20, 40 and 60°C.

As can be seen, slightly smaller RMSECV values were obtained when modeling the spectra recorded at different temperatures. However, model validation by an independent test set is necessary to determine whether these differences are really significant.

It should be noted that, until this study, no more than 3 LV's have been used in the PLS models in order to prevent overfitting. However, when using PLS models with a maximum of 5 LV's the results appeared to be significantly improved as is shown in Table 7. Obviously, these models should be validated before they are used in practice since the data set is rather limited and hence no external validation has been performed by using an independent test set. Besides, several data points observed in the correlation plots require closer examination of the physicochemical data to figure out whether these points should be considered as reliable or as outliers.

Table 7. Results of modeling spectra recorded at 20, 40 and 60°C for the SR properties using AFT as input and a maximum of either 3 or 5 LV's for the PLS models. Modeling based on corrected data for P and R&B; sample C22 excluded.

SR-property	Mean value	Variable temperature data		Variable temperature data	
		RMSECV	LV's	RMSECV	LV's
Log^{10} P (0.1 mm)	1.86	0.33	3	0.30	5
Log^{10} R&B (°C)	1.65	0.09	3	0.08	5
D_{SR} (g/ml)	1.006	0.013	3	0.010	5
V_{SR} (-)	42.5	1.6	3	1.3	5

4.4. Conclusions

The exposure of crude oils to 65°C during 24 h does not result in a significant improvement in the results of predicting the LR and SR properties. The performance of the different models was about the same as for the models based on the IR spectra of the untreated materials. Modeling of the differences between spectra of the same sample recorded at different temperatures revealed a linear relation with the temperature. Most likely this is the result of a decrease of the density with an increase of the temperature.

Modeling the IR spectra including the spectra recorded at 20, 40 and 65°C after mean centering for temperature and a PCA analysis produced virtually the same prediction errors. However, a more detailed analysis showed that the number of used LV's decreased for all LR-properties, which implies more robust and simpler models. A more detailed exploration by means of external model validation is necessary to determine the significance of these differences and to asses this topic.

References

1. P. de Peinder, D.D. Petrauskas, F. Singelenberg, F. Salvatori, T. Visser, F. Soulimani, B.M. Weckhuysen, Appl. Spectrosc. 62 (2008) 414.
2. P. de Peinder, T. Visser, D.D. Petrauskas, F. Salvatori, F. Soulimani, B.M. Weckhuysen, Vibr. Spectrosc. (2009) (in press).
3. J.D. McLean, P.K. Kilpatrick, J. Colloid. Interfac. Sci. 189 (1997) 242.
4. J.S. Buckley, Y. Liu, J. Petrol. Sc. Engin. 20 (1980) 155.
5. J.D. McLean, P.K. Kilpatrick, J. Colloid. Interf. Sc. 189 (1997) 242.
6. R.N. Singh, A.K. George, S. Arafin, J. Phys. D: Appl. Phys. 39 (2006) 1220.

7. C. Li, Q. Yang, M. Lin, J. Petrol. Sci.Engin. 65 (2009) 167.
8. J. Ding, J. Zhang, H. Li, F. Zhang, X. Yang, Energ. Fuels 20 (2006) 2531.
9. J.M. Létoffé, P. Claudy, M.V. Kok, M. Garcin, J.L. Volle, Fuel 74 (1995) 810.
10. R. Venkatesan, P. Singh, H.S. Fogler, J. Soc. Petrol. Sci. Engin. 7 (2002) 349.
11. J.S. Buckley, N.R. Morrow in *Physical Chemistry of Colloids and Interfaces in Oil Production*, H. Toulhoat and J. Lecourtier (Eds), Technip, Paris (1992).
12. K.M. Oghenejoboh, M.F.H. Abowie, Y.T. Puyate, Pollut. Res. 27 (2008) 7.
13. R. Roehner, F.V. Henson, US Patent 6841779 B1 20050111 (2005).
14. L. Andrade I.A. Farhat, K. Aeberhardt, R. Bro, S.B. Engelsen, Appl. Spectrosc. 63 (2009) 141.
15. R. Bro, Chemom. Intell. Lab. Syst. 38 (1997) 149.
16. A.K. Smilde, Chemom. Intell. Lab. Syst. 5 (1992) 143.
17. S.P. Gurden, J.A. Westerhuis, R. Bro, A.K. Smilde, Chemom. Intell. Lab. Syst. 59 (2001) 121.
18. A.C. Peinado, F. van den Berg, M. Blanco, R. Bro, Chemom. Intell. Lab. Syst. 83 (2006) 75.

Chapter 5

Prediction of Long Residue Properties of Potential Blends from Mathematically Mixed IR Spectra of Pure Crude Oils

Abstract

Research has been carried out to determine the feasibility of partial least squares regression (PLS) models to predict the long residue (LR) properties of potential blends from IR spectra that have been created by linearly co-adding the IR spectra of crude oils. It is found that the PLS prediction models developed for 7 different LR properties (i.e., yield long on crude (YLC), density (D_{LR}), viscosity (V_{LR}), sulfur content (S), pour point (PP), asphaltenes (Asph) and carbon residue (CR)), enabled to predict the LR properties of 16 blends in two ways. The first predictions were carried out on the IR spectra recorded from the physically prepared blend samples. Next, IR spectra were submitted to the PLS models, that were created mathematically by linearly co-adding the IR spectra of the corresponding crude oils in the appropriate weight ratio. Minor differences in the real and artificial blend spectra have been observed which have been assigned to non-linear effects. However, pre-processing of the spectra, by subsequently taking the 1st derivative, multiplicative signal correction (MSC) and mean centring (MC), resulted in predicted LR-property values of the two parallel sets that are largely the same. It implies that mimicking blend spectra by mathematically mixing the IR spectra of crude oils is a valuable, fast, clean and cheap alternative for the elaborate preparation and testing methods of real blends currently used in the laboratory. Besides, the method can be used as a rapid screening tool for large series of potential blends.

5.1. Introduction

Blending of crude oils is a commonly applied method in petrochemical industries to optimize unit throughput and meet product specification [1-8]. Extra heavy crude oils with low viscosity are, for instance, blended with light fractions to facilitate transport and improve the process ability. Another example is mixing of lower grade crude oils with higher grade materials to obtain a product with specific chemical and physical properties, such as bitumen. To determine mixing ratio, most companies use models or basic reference sets based on laboratory test data [7-11]. This is, however, largely a trial and error process of physically blending followed by analysis until the target specifications have been reached. For that reason, a method to predict the properties of a potential blend, without the need for elaborate preparation and testing, could be highly profitable.

Numerous studies have been undertaken to characterize and classify petrochemical compounds. Many of these apply spectroscopic methods as the discriminating tool, such as IR [12-19], NIR [20-27], Raman [28-30], UV-Visible [31, 32], MS [33] and NMR spectrometry [13, 33-39]. In addition, chemometric methods have been applied to spectroscopic data to develop models that enable determination or prediction of specific oil properties [16, 20, 23, 25, 26, 40-49]. Specific studies on the characterization of blends have been undertaken as well [50-52] of which Brown *et al.* [50] patented a method on IR analyzing an unknown crude oil as a blend of known components in order to predicts it properties.

In Chapter 2, we reported on the potential of IR characterization and chemometric classification of crude oils and blends to determine the properties of the long residue (LR) and short residues (SR). The aim was to develop prediction models to classify crude oils in terms of suitability for bitumen manufacture and to predict LR and SR bitumen properties as a function of the flash temperature directly from crude oil IR spectra. It has been demonstrated that such models can replace the elaborate and time-consuming ASTM (American Society for Testing and Materials) and IP (Institute of Petroleum) laboratory tests that are currently applied for characterization.

Since blends are mixtures of crude oils in a specific weight ratio, it might be possible to mimic the IR spectra of blends by linearly co-adding the spectra of the underlying crude oils in the same ratio. Next, the PLS prediction models could be applied to predict the LR and SR properties of these blends from the corresponding, artificially created IR spectra. If valid, this approach would eliminate the elaborate preparation of real blends in the laboratory, thus offering a ‘clean’ and fast testing and screening tool for large series of potential blends. In order to determine the validity of this approach, research was carried out on the prediction of the LR properties of 16 blends from IR spectra that were mathematically created from the corresponding crude oil spectra. The obtained prediction values were compared with the ones produced by the models from the spectra of the samples that were prepared physically in the laboratory. This Chapter describes the results of this study and shows that the LR property values predicted from real and mimicked IR spectra of blends match very well.

5.2. Methods and materials

Two sets of 16 IR spectra of crude oil blends have been submitted to 7 different PLS regression models in order to predict the following LR properties of these samples: yield-long-on-crude (YLC), density (D_{LR}), viscosity (V_{LR}), sulfur content (S), pour point (PP), asphaltenes (Asph) and carbon residue (CR). The first set of spectra consisted of the data recorded from 16 blends that were prepared in-house by physically mixing 2 crude oils from a collection of 47 crude oil samples that have been used to build the PLS models. For this purpose, crude oil samples were selected that represented a wide variety of LR properties. Details on the composition of the blends are given in Table 1. The 1:1 blends were prepared by mixing about 4 grams of each crude oil, while for the 2:1 ratio about 8 and 4 grams quantities were mixed. After firmly shaking, the blends were treated in agreement with the protocols reported in Chapter 2 and the IR spectra were recorded accordingly. The second set also consisted of 16 IR spectra of principally the same blends, but these spectra were created mathematically by linearly co-adding the IR spectra of the 2 corresponding crude oils in exactly the same ratio as the weight ratio of the physically prepared samples (Table 1).

Table 1. Composition of the 16 crude oil blends under investigation.

Blend nr.	Composition and origin	
	Crude 1	Crude 2
1	50% Denmark	50% Libya
2	50% Denmark	50% China
3	50% Denmark	50% Iraq
4	50% Denmark	50% Venezuela
5	50% Italy	50% Kazakhstan
6	50% Italy	50% Ecuador
7	50% Italy	50% Iran
8	50% Italy	50% Congo
9	33% Denmark	66% Libya
10	33% Denmark	66% Italy
11	33% Denmark	66% China
12	33% Denmark	66% Iraq
13	33% Italy	66% Venezuela
14	33% Italy	66% Kazakhstan
15	33% Italy	66% Ecuador
16	33% Italy	66% Congo

5.2.1. IR spectroscopy

IR measurements were carried out at room temperature on a Bruker Tensor-27 Fourier transform spectrometer equipped with a DTGS detector. The sample compartment was flushed with dry air to reduce interference of H_2O and CO_2. Spectra were recorded with a horizontal ATR accessory (FastIR^{TM}, Harrick) with a ZnSe crystal as the internal reflection element. The spectral resolution was 4 cm^{-1} for all spectra and 50 scans were accumulated with medium apodization for each spectrum. ATR-intensity correction was not applied. All spectra were recorded at room temperature (20°C) in closed cells to prevent evaporation of light-ends. Samples were homogenized at least 1 h before measurement by shaking the sample every 10 min. Experimental protocols on sample storage, pre-treatment, preparation and spectral recording have been applied throughout this study to assure reproducible, high quality IR spectra. Details on these protocols can be found in Chapter 2.

5.2.2. Chemometric prediction

The PLS regression models, as obtained from the study described in Chapter 2, have been used to predict the LR properties of 2 series of crude oil blends. Prior to submission to the models, the IR spectra of the blends were pre-processed similar to the spectra used for modeling by first taking the 1st derivative (25 pt. Savitzky-Golay smoothing), followed by Multiplicative Signal Correction (MSC) and Mean Centering (MC) of the data. Prior knowledge of artifacts like baseline drift and atmospheric spectral contributions were taken into account during pre-processing. A baseline correction was applied by subtracting a 3rd degree polynomial fit using the regions 4000-3500, 2500-2000, 1900-1800, 1560-1520, 1000-990 and 650-600 cm^{-1}. Subsequently the region 2500-1800 cm^{-1} was removed from the spectra. No extra pre-processing was applied on the mathematically created spectra. The spectral region from 3200-670 cm^{-1} was used as input, since it was concluded from the previous study based on trial and error that these models led to the best results [53]. Evaluation of the 7 LR property values, predicted by the models, properties, i.e., YLC, D_{LR}, V_{LR}, S, PP, Asph and CR, was carried out in two ways. First, for each property, the predicted values from the physically prepared blends were plotted against the values obtained from the set of mimicked blend spectra in order to determine the rate of correlation. Secondly, the difference in the predicted values of both data sets was expressed in a root mean square error (RMSE) parameter for each of the LR properties. Since no physical/chemical data were available for the 16 blends, these RMSE values were calculated for the set of mathematically created spectra relative to the property values predicted for the measured blends. As already described in Chapter 2, the developed PLS models have proved to be valid for the prediction of blends since 4 blends with known LR properties were included in the applied test set and predicted correctly.

5.3. Results and discussion

The spectra of the physically prepared blends and the mathematically composed blend spectra were found to be virtually the same, particularly after intensity normalization and pre-processing. This is illustrated in Fig. 1, showing the overlay of the real and the mimicked spectrum of the 50% Denmark, 50% Libya blend. On the other hand, as shown in Fig. 2, closer examination reveals that minor differences can be present in the fingerprint region 1100-700 cm^{-1}. The high similarity between the spectra of the real and the artificial blends is also reflected in the LR property values predicted by the models. This is illustrated in Fig. 3, showing the values predicted from the measured blend spectra plotted against the values obtained for the artificial blend spectra.

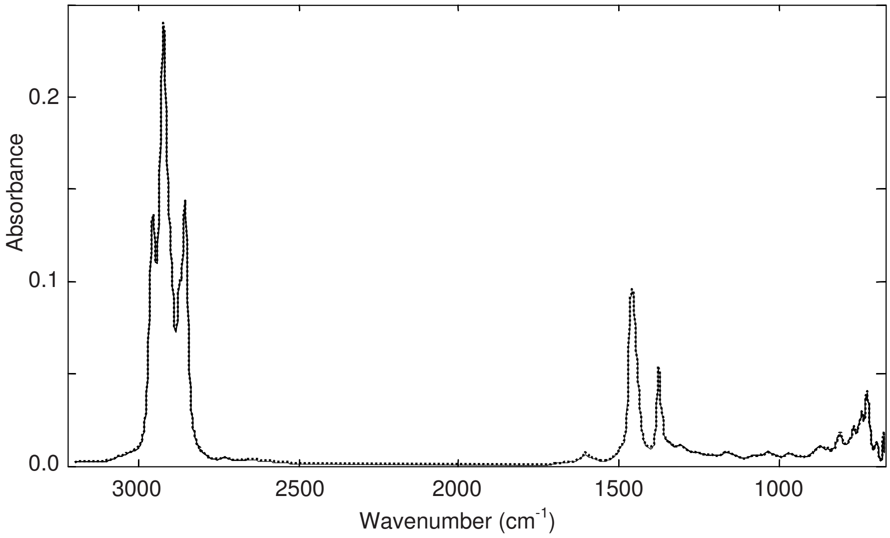


Fig. 1. Overlay of the real IR spectrum of Blend 1 (50% Denmark, 50% Libya) (solid line) and the mathematically created spectrum of this blend (dotted line), obtained by mixing the spectra of the corresponding crudes in the appropriate ratio.

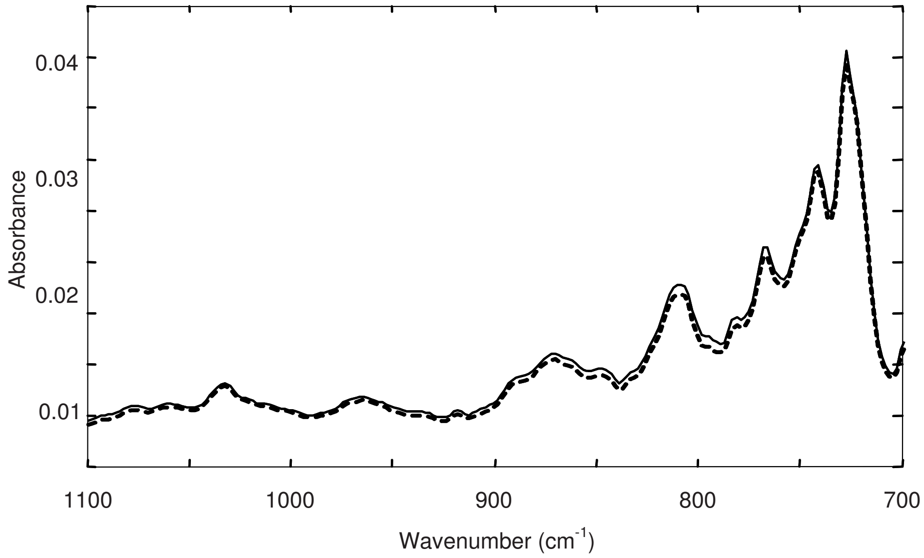


Fig. 2. Exploded view of the fingerprint region of the IR spectra of Blend 1 (50% Denmark, 50% Libya) shown in Fig. 1. Solid line: real blend, dotted line: mathematically created spectrum.

Table 2. RMSE values of the LR property values predicted from the artificially created IR spectra of blends versus the property values predicted from the IR spectra of the physically prepared samples. Additionally, the average property values of the crude oils used to build the prediction models and the corresponding RMSEP$_{VS}$ values are given, as well as the reproducibility of the ASTM and IP methods.

LR-property	Original models		Artificial/Measured blends	
	Mean value	RMSEP$_{VS}$	RMSE	ASTM/IP reproduc.
YLC (wt %)	52.5	3.5	5.2	1.2
D_{LR} (g/mL)	0.930	0.012	0.005	0.0005
V_{LR} (-)	35.3	2.1	0.9	-
S (wt %)	2.2	0.5	0.21	0.09xS (S>0.9%)
PP (°C)	29.6	9.1	2.2	9
Asph (wt %)	4.1	2.1	0.8	0.2xAsph
CR (wt %)	9.7	2.0	0.8	0.046x(3+CR)

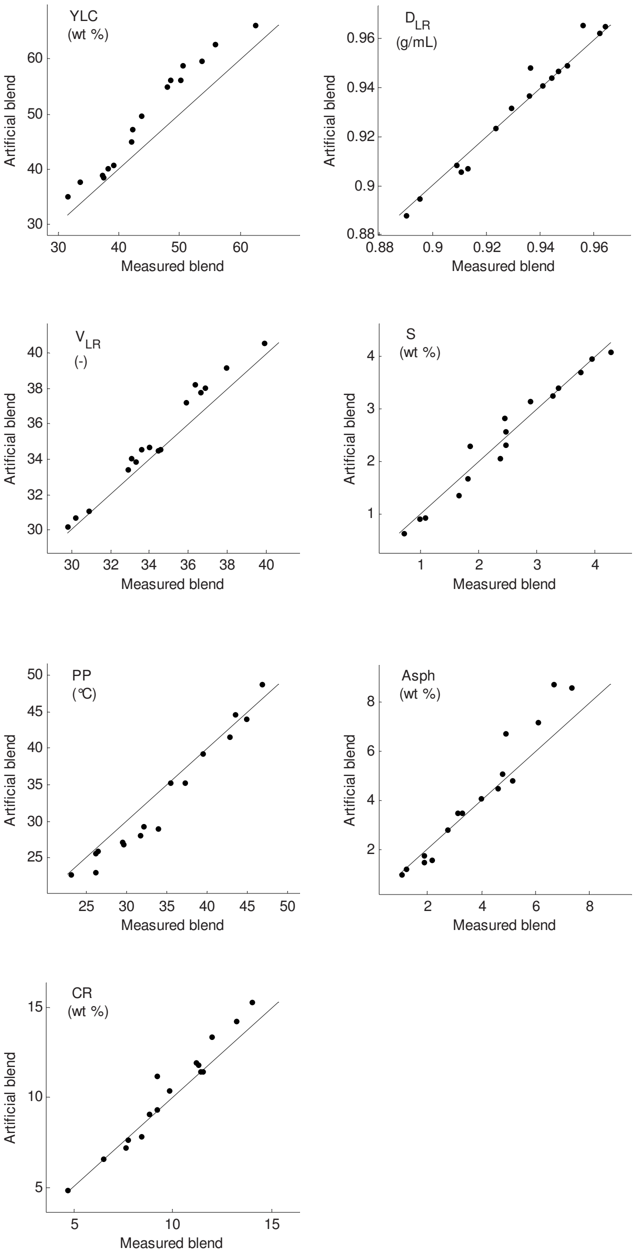


Fig. 3. LR property values predicted from the IR spectra of the real blends versus the spectra of the artificially created blends.

Similar conclusions can be drawn from the RMSE values presented in Table 2. For reasons of comparison, the RMSEP values (denoted as

$RMSEP_{VS}$) of the initial models have been added to this Table, as well as the corresponding average LR property values and the errors of the physicochemical ASTM and IP methods. Although it should be emphasized that the $RMSEP_{VS}$ values represent a different data set, it is evident that, apart from YLC, the scores for the artificial blends stay well within the errors of the prediction models. For that reason, it is concluded that straightforward mathematically mixing the IR spectra of crude oils offers a useful method to predict the LR properties of blends without having to physically prepare such mixtures.

It could be argued that further improvement of this method might be achievable when taking into account non-linear effects. For instance, the weight percentage of sulfur will change proportionally with the weight ratio of the crude oils, whereas the effect on parameters like viscosity, density and pour point will certainly be non-linear. In principle, these non-linearity effects will be reflected in differences in the IR spectra of the blends that are physically prepared and the ones that have been mathematically created. It might explain the small but significant differences in the fingerprint region of these spectra (Fig. 2). In a univariate model, these deviations would result in relatively larger prediction errors for the LR-properties that exhibit non-linear behaviour. For multivariate (PLS) models however, it is known that terms which are linear in absorbance, can account for non-linear effects in the samples and vice versa [52]. For that reason, it is not surprisingly that most of the plots in Fig. 3 show more or less linear correlations.

Closer examination of the plots also reveals that for YLC and V_{LR}, a bias is present. For YLC in particular, this is reflected in a relatively large RMSE value (Table 2). This ‘offset’ might as well originate from nonlinearities, but is not fully clear yet. At this point, it should be noted, that the fact that the IR spectra in this study have been recorded by means of the ATR technique, might play a role in the linearity and accuracy of the predictions. With this sampling method, the penetration depth of the IR radiation into the sample depends on the refractive index of the sample and is the largest at longer wavelength. It follows that non-linear effects will be the strongest at low wavenumbers. This is not only in line with the observations in Fig. 3, but also with the difference spectra of the mimicked and the measured blends in Fig. 4, which were obtained by spectral subtraction. As can be seen, considerable residues are present across the complete spectral region, particularly in the range 1600-500 cm^{-1}. Furthermore, the extreme appearance of the dotted line difference spectrum points to an outlier, but it is actually due to the relatively high water content in one of the involved crude oil components. However, the fact that even the property values for this artificial blend were predicted reasonably well proves that the models, including the pre-processing used, are quite robust and (partly) compensating for nonlinearities.

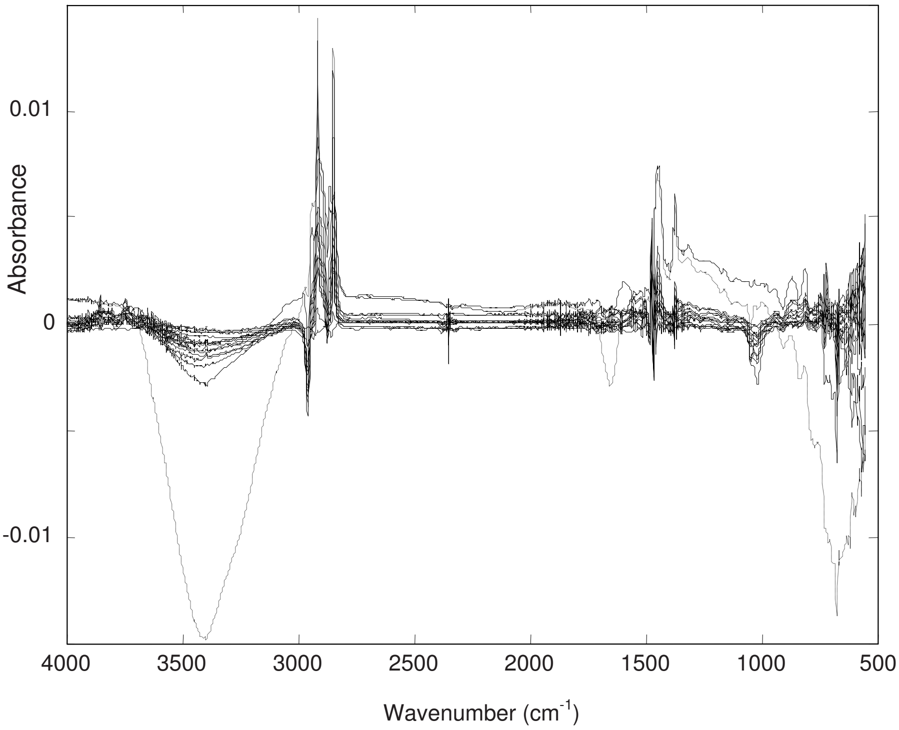


Fig. 4. Difference spectra obtained by subtracting the measured blendspectra from the artificially created ones.

In our opinion, taking into account such non-linear effects offers the possibility to achieve even more accurate results from the mimicked spectra. Especially for the LR properties that are sensitive to these effects, such as YLC, V_{LR}, D_{LR} and PP, this might lead to improvement of the reliability of the predicted values. A way to obtain insight into this would be the prediction of the LR properties from the spectra recorded in transmission. Unfortunately, the high viscosity of several of the blends prevented the use of transmission cells of sufficient short path length. Extra pre-processing steps and a calibration transfer protocol for the artificial spectra might be other ways to achieve this. On the other hand, it should be noted that nonlinearities may as well be composition dependent, which makes this matter highly complex. Further research on this issue is therefore needed. However, even without this improvement, it can be concluded that mimicking the IR spectra of potential blends by mathematically mixing crude oil spectra offers a rapid and ‘clean’ screening alternative for the elaborate laboratory preparation and testing of blends. Most likely, this will also be valid for the prediction of SR properties.

5.4. Conclusions

PLS regression models, based on IR spectra of crude oils, can be applied in a straightforward manner for the prediction of LR properties of IR blend spectra that have been mathematically created by linearly co-adding the IR spectra of the corresponding crude oils in the appropriate weight ratio. The PLS models predict the LR property values for the mathematically created blend spectra largely the same as for the IR spectra of the physically prepared blends. Therefore, it is concluded that ‘artificial blending’ of crude oil IR spectra is a valuable alternative for the elaborate preparation of blends in the laboratory. Besides, the method can be used as a rapid testing and screening tool for large series of potential blends. Compared to the spectra of real blends, slightly larger deviations are observed for some of the LR properties when submitting artificial blend spectra. This is assigned to non-linear effects that occur when blends are physically prepared, but which are not reflected in the mathematically created blend spectra. Further reduction of the prediction errors is foreseen by implementation of an extra pre-processing step that corrects for this type of non-linearity.

References

1. G.B. Dickakian, US patent (1989) 4853337.
2. W.E.M. Schermer, P.M.J. Melein, F.G.A. van den Berg, Petrol. Sci. Technol. 22 (2004) 1045.
3. K.B. MacDougall, Proc. 80th Int. School Hydrocarb. Measur. (2005) 2075/1.
4. J.H. Gary, G.E. Handwerk, “*Petroleum Refining Technology and Economics”*, Marcel Dekker, New York (1994).
5. J. de Jesus Rubio, W. Yu, IEEE Catalogue 04EX865 (2004) 427.
6. J.D. Kelly, J.L. Mann, Hydrocarb. Process. June 2006, 47.
7. J.D. Kelly, J.L. Mann, Hydrocarb. Process. June 2006, 72.
8. S.P. Li, F. Tang, B.X. Shen, B.L. Zhang, J.Y. Yang, X.R. Xu, Petrol. Sci. Technol. 26 (2008) 1499.
9. S. Zhao, B.D. Sparks, L.S. Kotlyar, K.H. Chung, Catal. Today 125 (2007) 122.
10. www.process-nmr.com/fox-app/crd_blndng.htm.
11. www.jiskoot.com/Applications/Crude_Oil/Crude_Oil.html.
12. T.A. Adedosu, O.O. Sonibare, J. Appl. Sci. 5 (2005) 906.
13. E. Evans, B. Batts, N. Cant, Fuel 66 (1987) 327.
14. H. Ganz, W. Kalkreuth, Erdoel & Kohle, Erdgas, Petrochem. 43 (1990) 116.
15. W.I. Friesen, K.H. Michaelian, Y. Long, T. Dabros, Am. Chem. Soc., Div. Fuel Chem. 50 (2005) 228.
16. B. Pavoni, N. Raido, R. Plassa, S. Frignani, Ann. Chim. 94 (2004) 521.
17. N. Pieri, J. Kister, L. Quoniam, J.P. Planche, L. Germanaud, Polycycl. Arom. Comp. 9 (1996) 29.
18. P.W. Yang, H.H. Mantsch, L.S. Kotlyar, J.R. Woods, Energ. Fuels 2 (1988) 26.
19. H. Yuan, X. Chu, H. Li, Y. Xu, Fuel 85 (2006) 1720.
20. M.V. Reboucas, E.C. Santos, F.S.V. Vieira, Vibr. Spectrosc. 44 (2007) 187.
21. S. Bages, B. Descales, D. Lambert, J.-R. Llinas, A. Martens, S. Osta, M. Sanchez, PCT Int. Appl. (1996) WO 9611399.
22. M. Blanco, S. Maspoch, I. Villarroya, X. Peralta, J.M. González, J. Torres, Appl. Spectr. 55 (2001) 834.
23. S. Macho, M.S. Larrechi, Trends Anal. Chem. 21 (2002) 799.
24. R.J. Mikula, V.A. Munoz, N. Wang, B. Bjornson, D. Cox, B. Moisan, K. Wiwchar, J. Can. Petr. Techn. 42 (2003) 50.
25. S. Satya, R.M. Roehner, M.D. Deo, F.V. Hanson, Energ. Fuels 21 (2007) 998.
26. Z. Sikora, W. Salacki, Petrol. Coal 38 (1996) 65.
27. J. Kelly, D. Lambert, A. Martens, PCT Int. Appl. (1998), WO 9836274.
28. M. He, X. Lu, S. Yao, Y. Liu, W. Fan, Dizhi Keji Qingbao 24 (2005) 67.
29. J. Jehlička, O. Urban, J. Pokorný, Spectrochim. Acta A 59 (2003) 2341.

30. K.H. Michaelian, R.H. Hall, J.T. Bulmer, Can. J. Anal. Sci. Spectrosc. 49 (2004) 43.
31. D. Felscher, Chemia Analityczna 22 (1977) 451.
32. J.G. Reynolds, E.L. Jones, J.A. Bennett, W.R. Biggs, Fuel Sci. Technol. Int. 7 (1989) 625.
33. C.H. Tsai, M.D. Deo, F.V. Hanson, A.G. Oblad, Fuel Sci. Technol. Int. 10 (1992) 1437.
34. F. Ali, Z.H. Khan, N. Ghaloum, Energ. Fuels 18 (2004) 1798.
35. J. Bryan, K. Mirotchnik, A. Kantzas, J. Can. Petrol. Technol. 42 (2003) 29.
36. C. Giavarini, C. Vecchi, Fuel 66 (1987) 868.
37. M.U. Hasan, M.F. Ali, A. Bukhari, Fuel 62 (1983) 518.
38. J.G. Speight, Fuel 49 (1970) 76.
39. J. Thiel, M.R. Gray, AOSTRA J. Research 4, 63 (1988) 63.
40. Q. Abbas, N. Dupuy, C. Rebufa, L. Vrielynck, J. Kister, A. Permanyer, Appl. Spectrosc. 60 (2006) 304.
41. O.Y. Begak, A.M. Syroezhko, V.V. Fedorov, M.N. Malechkina, M.B. Nassonova, Russ. J. Appl. Chem. 75 (2002) 1173.
42. P. Ghosh, B. Chawla, P.V. Joshi, S.B. Jaffe, Energ. Fuels 20 (2006) 609.
43. K.J. Johnson, R.E. Morris, S.L. Rose-Pehrsson, Energ. Fuels 20 (2006) 727.
44. S. Kokot, N. Asker, Proc. 8th Int. Conf. Fourier Transform Spectrosc., SPIE- Int. Soc.Opt. Engin. 1575 (1992) 495.
45. S. Kokot, N. Asker, Proc. 6th Int. Conf. Near Infr. Spectrosc. 22 (1995) 1994.
46. A.T. Pauli, F.P. Miknis, A.G. Beemer, J.J. Miller, Preprints - ACS Division of Petroleum Chemistry 50 (2005) 255.
47. V.R. Puttagunta, A. Miadonye, AOSTRA J. Research, 7 (1991) 241.
48. J. Sanchez Caba, A. Crespo, A. Aguirre, Ingenieria Quimica 29 (1997) 77.
49. R. Van Doorn, F.H. Hooyman-Spaargaren, R.J. Neugebauer, L.M. Smeets, PCT Int. Appl. (1997) WO 9714953 A1 19970424.
50. J.M. Brown, C.J. Chrostowsky, PCT Int. Appl. (2006) WO 2006023800 A2 20060302.
51. D.A. LeFebre, L.M. Lane US Patent (1995) US 5430295 A 19950704.
52. D. Stratiev, Z. Belchev, P. Petkov, K. Kirilov, Oil & Gas 34 (2008) 31.
53. P. de Peinder, D.D. Petrauskas, F. Singelenberg, F. Salvatori, T. Visser, F. Soulimani, B. M. Weckhuysen, Appl. Spectrosc. 62 (2008) 414.
54. R. DiFoggio, Appl. Spectrosc. 49 (1995) 67.

Chapter 6

Robustness of the IR Prediction Models

Abstract

Research has been carried out to instrumentation effects on the PLS regression models based on IR spectra to predict the LR properties of crude oils. Three factors that may affect the predictive accuracy have been studied: 1) the spectral reproducibility in time, 2) the use of different instrumental set-ups for data acquisition, and 3) the property prediction from spectral data that have been recorded on a different instrument than the one that was used to acquire the calibration data.

The reproducibility of the spectra in time appears to be very high as expressed by a negligible effect on the predicted LR property values. Furthermore, the effect of using a different spectrometer or a different ATR accessory on the performance of the prediction models was found to be very small, provided the spectra for calibration and validation were recorded on the same experimental set-up and with more or less the same scanning conditions. In contrast, the performance of the models to predict LR properties from a spectrum that was recorded on another spectrometer than the one that was used for modeling, turned out to be negatively affected. In order to maintain the predictive power of the original models for this type of spectra, the development of a calibration transfer protocol is inevitable.

6.1. Introduction

As demonstrated in the previous Chapters of this PhD thesis, IR spectroscopy in particular, combined with PLS regression, is a valuable method to predict the quality of crude oils for bitumen application from a single spectrum. As such, it forms a fast alternative for the elaborate and time-consuming laboratory tests, currently used in oil industries, to determine the LR and SR properties of crude oils and blends [1, 2].

Thus far, the IR spectra that were used for calibration and validation were recorded under well-defined identical experimental conditions to assure the highest possible reproducibility and reliability of the prediction models. This included the use of only one IR spectrometer with fixed scanning conditions and the same attenuated total reflection (ATR) sampling accessory.

It is reasonable to assume that the same high prediction accuracy can be achieved when spectral data are used that have been recorded on other IR instruments and/or accessories, provided that the data for modeling and prediction originate from the same set-up.

It is to be expected, however, that the performance of a prediction model will decrease when spectra are submitted, that have been recorded on a different instrument than the one that was used to build the model. This effect can be attributed to the (small) differences in the numerical output accuracy and non-matching data point spacing when comparing IR instruments form different manufacturers. Although practically all companies agreed to enable conversion of their spectra to a uniform, so-called JCAMP-DX format to facilitate data exchange [3-6], small but significant differences still exist due to different optical configurations of instruments [7, 8]. In mathematical and chemometrical operations, this can be a complicating factor that negatively affects the results.

For that reason, several studies have been undertaken to circumvent this problem, particularly, in applications of NIR spectroscopy [9-19]. One strategy is to create robust models that can be achieved by selecting robust spectral variables (wavenumbers) [9-11], and/or by the removal of differences by data pre-processing [12, 13]. Many examples can be found in literature of protocols that facilitate the transfer of a calibration model from one instrument to another [14-20]. Other strategies include bias correction and the transfer the spectral data instead of the model, i.e. Direct Standardization or Piecewise Direct Standardization [15, 20]. In order to determine whether such a protocol will also be needed to realize a more universal application of the LR prediction models, developed in this PhD thesis, a robustness study for the models has been carried out, based on 3 different IR spectrometers and 2 different ATR accessories. This chapter describes the results of this study.

6.2. Experimental

6.2.1. *Methods and materials*

IR data acquisition has been carried out on 5 different experimental set-ups as listed in Table 1. The Perkin-Elmer 2000 FT-IR spectrometer was equipped with a DTGS detector. With this instrument 25 scans of 4 cm^{-1} optical resolution were accumulated for one spectrum. The Bruker Tensor-27 and Tensor-37 FT-IR instruments were also equipped with a DTGS detector. Spectral scanning with these instruments was carried out with an optical resolution of 4 cm^{-1} too, but 50 scans were accumulated for each spectrum. The optical element of the MIRacle ATR accessory (Pike Technologies) was diamond/ZnSe. The one of the FastIR^{TM} (Harrick) was a large area ZnSe crystal.

Table 1. Overview of the different combinations of IR spectrometers and ATR units.

Set-up	IR-instrument	ATR accessory
A	Perkin-Elmer 2000	MIRacle^{TM}
B	Bruker Tensor-27	FastIR^{TM}
C	Perkin-Elmer 2000	FastIR^{TM}
D	Bruker Tensor-37	FastIR^{TM}
E	Bruker Tensor-37	MIRacle^{TM}

The effect of the reproducibility of the spectra on the accuracy of the LR prediction models was studied first. For this purpose, the IR spectra of the 28 crude oils of the calibration set and 20 of the validation set, as listed in Table 1 of Chapter 2 (samples C1-C28 and V1-V20), have been measured on Set-up A with a time gap of 2 months. During this period, samples were stored in a refrigerator at 4℃.

The effect on the results of modeling for the LR property prediction of the IR spectra recorded on two different instrumental set-ups was carried out with Set-ups A and B on the same set of 28 crude oil samples C1-C28. The results obtained for the two different spectral datasets were compared.

The study to establish the robustness of the LR prediction models, built from the 28 calibration spectra recorded with Set-up B, was carried out on 10 crude oils that were selected in order to obtain a representative variety of geographical origin and LR properties. For this purpose, spectra recorded on all 5 different experimental set-ups A-E were submitted to the models. The root-mean-square-error-of-prediction (RMSEP) values obtained for the different data sets were mutually compared.

The following LR properties were studied; the long residue yield-on-crude (YLC), density (D_{LR}), viscosity (V_{LR}), sulfur content (S), pour point (PP), asphaltenes (Asph) and carbon residue (CR). As input values for these properties, the data as presented in Table 1 of Chapter 2 and obtained with the corresponding ASTM and IP methods were used.

6.2.2. Chemometrics

Partial Least Squares (PLS) modeling of the different data sets was carried out with the PLS Toolbox (Eigenvector Research, Inc.) for MatLab (The MathWorks, Inc.). Pre-processing and PLS modeling was performed according to the method described in Chapter 2. In order to enable submission of spectra for LR property prediction, which have been recorded on the Perkin-Elmer instrument (set-up A and C) to models built from Bruker Tensor-27 spectra (set-up B), matching of the data point position and resolution was necessary. For that reason, the Perkin-Elmer spectra were pre-processed to obtain the same data point resolution.

6.3. Results and discussion

The results of modeling the spectra that have been recorded with a time gap of two months on the same instrumentation, namely Set-up A, have been summarized in Table 1. As appears, the RMSECV values of modeling the LR properties are very similar for the duplicate measurements when based on the total data set of 28 crude oil spectra. The 20 validation samples were used to determine the RMSEP.

Table 1. Results of PLS modeling of the IR spectra of 28 crude oils, as recorded at 2 different moments in time with Set-up A, for the LR properties.

		Measurement 1			Measurement 2		
LR-property	Mean value	RMSECV	LV's	RMSEP	RMSECV	LV's	RMSEP
YLC (wt %)	52.5	3.0	8	2.3	2.6	7	2.1
D_{LR} (g/mL)	0.930	0.011	8	0.016	0.010	8	0.010
V_{LR} (-)	35.3	1.9	4	1.9	1.6	5	1.5
S (wt %)	2.16	0.5	8	0.4	0.5	9	0.3
PP (°C)	29.6	9.2	6	10.5	9.0	5	11.3
Asph (wt %)	4.1	2.0	9	3.2	2.1	10	3.1
CR (wt %)	9.7	1.8	5	1.7	1.7	5	1.5

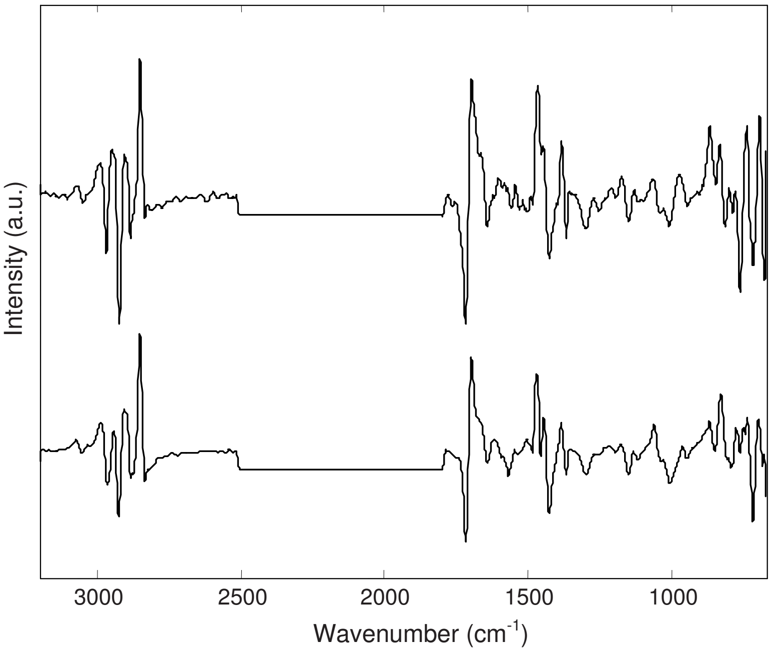


Fig. 1. Regression vectors of the PLS models for V_{LR} based on spectra from Measurement 1 (lower graph) and Measurement 2 (upper graph) on set-up A.

The obtained regression vectors of the 2 models for V_{LR}, presented in Fig. 1, are very similar. This means that not only the RMSEP values are comparable, but also the PLS models itself. For the other properties the regression vectors were also very much the same. Based on these findings, it was concluded that data acquisition over a longer period, including the storage of samples in a refrigerator at 4℃, does not significantly affect the model performance or the IR spectra.

The results of modeling the 28 IR calibration spectra that have been recorded on the 2 different spectrometers with different ATR-units are listed in Table 2, including the RMSEP values of the 20 validations samples. The average peak-to peak noise in the IR spectra (range 2200-2000 cm^{-1}) recorded with Set-up B was a factor of 2 better than for Set-up A (0.0004 against 0.0008 absorbance units). This is ascribed particularly to the larger contact area of the FastIR^{TM} accessory and could be one of the reasons that the RMSEP values obtained with this accessory are slightly better for D_{LR}, V_{LR}, S and Asph. However, as appears from Table 2, the effect of the difference in the S/N ratio is too small to affect the results of modeling significantly. It indicates that the performance of the prediction models does not depend on the instrumentation used, provided the spectral quality is about the same. Furthermore, it can be concluded that the chosen experimental settings and scan parameters were correct for the purpose of this modeling study.

Closer examination of the different regression vectors revealed a high similarity, but also some significant differences in relative intensities. This is illustrated in Fig. 2 by the regression vectors obtained for V_{LR} from data of Set-up's A and B. It indicates that models built with the spectra collected with one set-up cannot be used straightforward to predict a property value using spectra obtained from the other set-up. It should be noted that the minor shifts observed in the peak positions in the regression vectors were also observed when comparing the original Bruker and Perkin-Elmer spectra. It implies that the observed differences in the regression vectors are largely due to the different optical configuration of both set-ups. Especially the combination of the ATR accessory and beam shape of the instrument is considered to be the main source of the differences.

Table 2. Results of PLS modeling for the LR properties of 28 crude oil IR spectra recorded on Set-up A and Set-up B.

		Set-up A			Set-up B		
LR-property	Mean value	RMSECV	LV's	RMSEP	RMSECV	LV's	RMSEP
YLC (wt %)	52.5	2.6	7	2.1	3.7	9	2.8
D_{LR} (g/mL)	0.930	0.010	8	0.010	0.012	8	0.007
V_{LR} (-)	35.3	1.6	5	1.5	2.0	6	1.3
S (wt %)	2.16	0.5	9	0.3	0.6	8	0.2
PP (°C)	29.6	9.0	5	11.3	9.2	5	10.9
Asph (wt %)	4.1	2.1	10	3.1	2.2	1	2.5
CR (wt %)	9.7	1.7	5	1.5	2.1	6	1.6

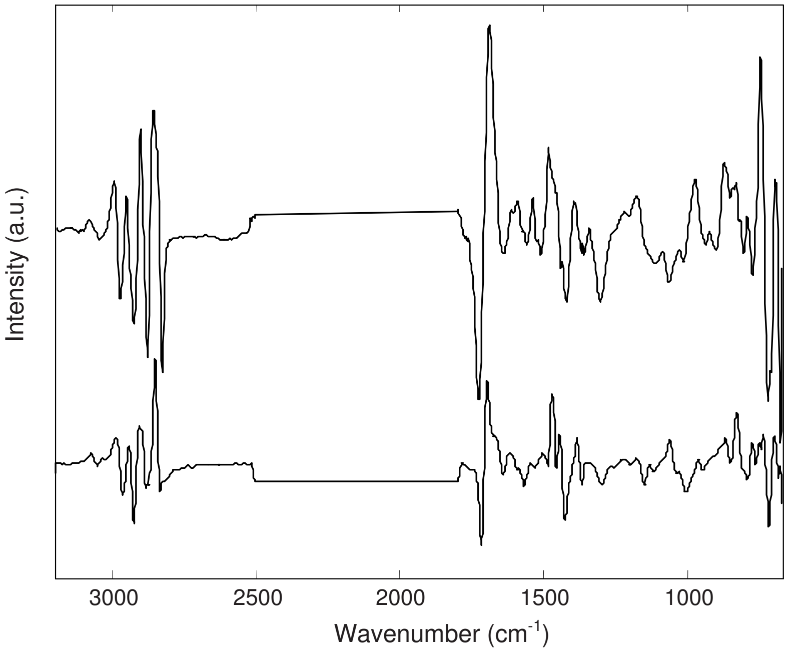


Fig. 2. Regression vectors of the PLS models for V_{LR} based on spectra from Set-up A (lower graph) and Set-up B (upper graph).

The results of predicting the LR properties from the 10 spectra recorded on a different experimental set-up than the one that was used for the spectra to build the models (*i.e.*, Set-up B), are shown in Table 3 and Fig. 3. Table 3 lists the RMSEP values obtained for each of the LR-properties and each of the 5 different instrumental combinations (Set-up A-E). Fig. 3 represents the LR properties predicted for the different sets plotted against the LR reference values. The RMSEP values clearly show that the data measured with the Tensor-37/FastIR^{TM} combination (set-up D) match the best with the prediction results of the original models based on Tensor-27/FastIR^{TM} spectra (set-up B). This is rather obvious regarding the similarity in optical lay-out and data acquisition features of both instrumental combinations. Apart from the property Pour Point (PP), which is not connected to a very reliable model, all RMSEP values are the lowest for this configuration. However, the results obtained for the other combinations are at maximum a factor 2-3 worse. On the one hand, it implies that the robustness of the original prediction models is fairly large, particularly since a rather straightforward pre-processing method, i.e., baseline correction, 1st derivative and MSC, was applied to the spectra. On the other hand, a calibration transfer method is expected to further improve the similarity between the spectra recorded on different instrumental set-ups.

Table 3. RMSEP values of the LR properties predicted from 10 spectra recorded on 5 different experimental set-ups (A-E), using the models from spectra recorded on Set-up B.

	RMSEP values				
LR-property	Set-up A	Set-up B	Set-up C	Set-up D	Set-up E
YLC (wt %)	3.6	1.6	4.1	2.9	4.5
D_{LR} (g/mL)	0.008	0.005	0.012	0.006	0.013
V_{LR} (-)	0.9	1.0	1.6	0.9	1.7
S (wt %)	0.4	0.2	0.6	0.3	0.4
PP (°C)	9.5	5.5	6.8	8.3	13.0
Asph (wt %)	1.6	1.4	1.6	1.6	2.2
CR (wt %)	1.5	1.2	1.1	1.1	1.6

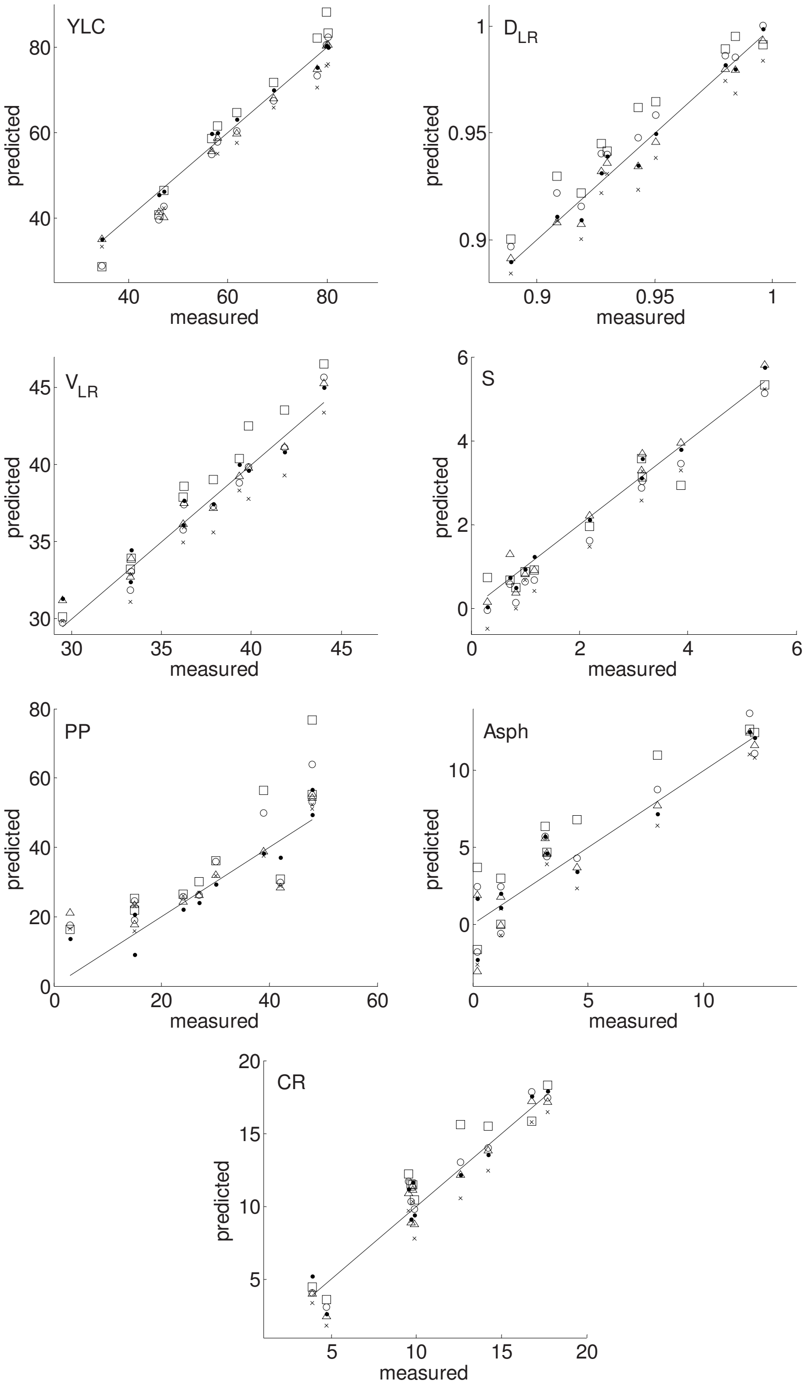


Fig. 3. Correlation plots of predicted LR values from 10 spectra recorded on 5 different experimental set-ups (A-E), using the models built from spectra recorded on Set-up A.

6.4. Conclusions

It is found that the performance of the PLS models to predict the LR properties of crude oils, is not affected by a large time delay in the acquisition of the data that are used for calibration and validation. Moreover, the effect of the applied spectrometer and/or ATR accessory on the performance of the prediction models is very small when the spectra for calibration and validation are recorded on the same instrumentation and with comparable scan settings. The performance of the models to predict the LR properties from spectra that have been recorded on a different spectrometer than the instrument that is used to build the models decreases with a maximum factor of 2. This accuracy is still sufficiently good for most of the LR properties to apply the models for indicative prediction purposes. However, in order to assure the highest reliability and facilitate a more universal application of the models, further research on this issue is necessary. In this respect, the development of a calibration transfer protocol for each data acquisition station to be used next to the original set-up is essential. In addition, the issue of instrument independent wavenumber selection should be explored.

References

1. P. de Peinder, D.D. Petrauskas, F. Singelenberg, F. Salvatori, T. Visser, F. Soulimani, B. M. Weckhuysen, Appl. Spectrosc. 62 (2008) 414.
2. P. de Peinder, D.D. Petrauskas, F. Singelenberg, F. Salvatori, T. Visser, B. M. Weckhuysen. PCT Int. Appl. (2008) WO/2008/135411.
3. R.S. McDonald, P.A. Wilks, Appl. Spectrosc. 42 (1988) 151.
4. J.G. Grasselli, Pure Appl. Chem. 63 (1991) 1781.
5. P. Lampen, J. Lambert, R.J. Lancashire, R.S. McDonald, P.S. McIntyre, D.N. Rutledge, T. Fröhlich, A.N. Davies, Pure Appl. Chem. 71 (1999) 1549.
6. R. Lancashire, A. Davies, Chem. Int. 28 (2006) 10.
7. A.N. Davies, H. Hillig, M. Linscheid, Proc. 4th Workshop Comput. Chem., Software Dev. Chem. 4, (1990) 147.
8. E. Smith, G. Dent (Eds), *Modern Raman Spectroscopy. A Practical approach*, Wiley & Sons, London (2005).
9. W. Cai, Y. Li, X. Shao, Chemom. Intell. Lab. Syst. 90 (2008) 188.
10. F. Rossi, D. Francois, V. Wertz, M. Meurens, M. Verleysen, Chemom. Intell. Lab. Syst. 86 (2007) 208.

11. T. Chen, E.E. Martin, Anal. Chim. Acta 631 (2009) 13.
12. M. Zeaiter, J.-M. Roger, V. Bellon-Maurel, Trends Anal. Chem. 24 (2005) 437.
13. K.E. Kramer, R.E. Morris, S.L. Rose-Pehrsson, Chemom. Intell. Lab. Syst. 92 (2008) 33.
14. C.F. Pereira, M.F. Pimentel, R.K.H. Galvao, F.A. Honorato, L. Stragevitch, M.N. Martins, Anal. Chim. Acta 611 (2008) 41.
15. T. Fearn, J. Near Infrar. Spectrosc. 9 (2001) 229.
16. E. Bouveresse, B. Campbell in *Handbook of Near-Infrared Analysis* (3rd Edition), D.A. Burns, C.W Ciurczak (Eds), Marcel Dekker, New York (2008) 231.
17. B. Igne, C.R. Hurburgh, J. Near Infrar. Spectrosc. 16 (2008) 539.
18. C. Tan, M. Li, Anal. Sciences 23 (2007) 201.
19. P. Dardenne, R. Biston, G. Sinnaeve in *Near Infra-Red Spectroscopy,* K.I. Hildrum (Ed), Ellis Horwood , Chichester (1992) 291.
20. K.E. Kramer, G.W. Small, Appl. Spectrosc. 63 (2009) 246.

Chapter 7

Sulfur Speciation of Crude Oils by PLS Regression Analysis of IR Spectra

Abstract

Research has been carried out to determine the feasibility of PLS modeling of IR spectra of crude oils as a tool for fast sulfur speciation of these materials. The study is a continuation of the methods to predict long and short residue properties of crude oils from IR and NIR spectra as outlined in the previous chapters. Retention data of two dimensional gas chromatography (2D-GC) of 47 crude oil samples have been used as input for modeling the corresponding IR spectra. 10 different PLS prediction models have been built: 1 for the total sulfur content and 9 for the sulfur compound classes 1) sulfides, thiols, disulfides and thiophenes, 2) aryl-sulfides, 3) benzothiophenes, 4) naphthenic-benzothiophenes, 5) dibenzothiophenes, 6) naphthenic-dibenzothiophenes, 7) benzo-naphthothiophenes, 8) naphthenic-benzonaphthothiophenes and 9) dinaphthothiophenes. From the total set of 47 spectra, 28 were used for calibration. The remaining 19 spectra were used as a test set to validate the PLS regression models. The results obtained confirm the conclusion from the previous chapters that PLS modeling of IR spectra to predict the total sulfur concentration of a crude oil is a valuable alternative for the commonly applied physicochemical ASTM method D2622. Besides, it is demonstrated that the concentration of dibenzothiophenes and the related benzothiophene compound classes benzothiophenes, naphthenic-benzothiophenes and naphthenic-dibenzothiophenes can be predicted with reasonable accuracy. It implies that these models offer a valuable tool for quick on-site screening on these compounds, which are potentially harmful for the production plant. The models for the other sulfur compound classes are insufficiently accurate to be used as a method for detailed sulfur speciation of crude oils.

7.1. Introduction

Crude oils are very complex mixtures of organic compounds with a large variety in elemental composition and chemical structures. All crude oils contain sulfur in concentrations between 0.1 weight % in light samples up to 10 % in for example bitumen and tar sands [1]. The majority of the sulfur is present as organic molecules in more than 10.000 different structures, ranging from aliphatic sulfides, disulfides and alkyl-substituted thiophenes to a variety of large polycyclic benzothiophenes [2]. The presence of sulfur species in crude oils has a severe impact on the oil production and refinery process. Next to direct corrosive effects on the plant infrastructure and equipment, macromolecular sulfur compounds form a substantial part of the solid asphaltenes and may cause clogging of the pipelines [3]. Therefore, an important task at production platforms and refineries is to quickly identify the compounds that are harmful for the production plant.

Another, well-known drawback of sulfur in crude oils is the release of sulfur oxides (SO_x) upon combustion of crude oil based fuels. This environmental effect has lead to more and more severe directives on SO_x emission. As a result, novel or improved hydrodesulfurization (HDS) catalysts have been developed and as a consequence the sulfur content of fuels is dramatically reduced. Nowadays, the maximum sulfur concentration in Europe is 10 ppm for gasoline and diesel [4] and 1000 ppm for marine diesel [5]. Desulfurization is therefore a big topic in oil industries. The current method of choice in refineries is HDS by means of e.g., a cobalt-molybdenum based catalyst. This catalytic method is expected to stay the dominant technology for the coming years, even though it is still not possible to eliminate the sulfur completely [6]. On the other hand, HDS is an expensive treatment for deep desulfurization, while the removal of heterocyclic aromatic sulfur compounds is not very effective. This is particularly relevant since the exploration of the tar sand fields in Canada and China has brought large amounts of crude oils with high sulfur concentrations onto the world market. For that reason, research for alternative methods and ways to enhance the efficiency of the HDS process is ongoing [4]. Obviously, also in this process, detailed knowledge of the qualitative and quantitative composition of sulfur compounds in crude oils is essential. Besides, the type and molecular structure of the sulfur compounds are found to affect the crackability and detachability [7], while the desulfurization efficiency for an individual sulfur compound differs with the type of crude [8]. Evidently, sulfur speciation of crude oils, either into detail or indicative and fast, is an important task in oil industries.

Many analytical techniques have been explored for this purpose, ranging from square wave voltammetry [9] and liquid chromatography [10] to conventional gas chromatography (GC) and two dimensional GC (2D-GC) [11-13]. A variety of sulfur selective detectors have been used in combination with GC, such as atomic emission detection (GC-AED) [14], sulfur chemiluminescence detection (GC-SCD) [1, 7, 15-18] and mass spectrometric detection (MSD) [19-24]. Other detection techniques have been based on X-ray spectroscopy including X-ray fluorescence (XRF) [25-28] and X-ray absorption near-edge structure (XANES) spectroscopy [29-32]. Furthermore, the potentials of temperature programmed reduction and oxidation methods have been studied [27, 33] as well as the new, but powerful technique of Fourier transform ion cyclotron resonance (FT-ICR)-MS [3, 34-38]. Occasionally, IR spectroscopy has been used, either including an oxidation pre-treatment [39] or without it [40]. The advantage of IR is that it can be easily performed on location without any preparation of the sample. In the previous Chapters, we have demonstrated the viability of chemometric modeling IR spectra of crude oils to predict long and short residues properties of crude oils straightforward from their spectra. This method, based on Partial Least Squares (PLS) regression models has been patented [41] and is currently tested on-site as a fast alternative for the much more elaborate ASTM and IP methods used thus far. Also, the method turned out to be able to predict the sulfur content with high accuracy. For that reason, a study to the potentials of PLS modeling of IR spectra as a tool for sulfur speciation is a logical next step. This Chapter describes the results of that study, using the speciation data obtained from standard 2D-GC analysis as reference values.

7.2. Experimental

Protocols for sample storage, pre-treatment and preparation, as described in Chapter 2, have been applied throughout this study to assure the acquisition of reproducible, high quality data. The same set of samples has been used as in Chapters 3-5 except sample V20, for which no 2D-GC data could be obtained. The resulting set of 47 crude oil samples represents a wide range of oil wells and hence a large variety of different sulfur compounds and concentrations. Modeling has been carried out for 10 different sulfur classes, as listed in Table 1, i.e., the total sulfur content and 9 commonly used sulfur speciation groups.

Table 1. Sulfur compound classes as applied in this study.

1	STD	Sulfides, Thiols, Disulfides, Thiophenes
2	Ar-S	Aryl-sulfides
3	BT	Benzothiophenes
4	NBT	Naphthenic-benzothiophenes
5	DBT	Dibenzothiophenes
6	NDBT	Naphthenic-dibenzothiophenes
7	BNaT	Benzonaphthothiophenes
8	NBNaT	Naphthenic-benzonaphthothiophenes
9	DNaT	Dinaphthothiophenes
10	S Total	Total sulfur amount (including elemental S)

7.2.1. IR-spectroscopy

IR measurements have been carried out at room temperature on a Bruker Tensor-27 Fourier transform spectrometer equipped with a DTGS detector. The sample compartment was flushed with dry air to reduce interference of H_2O. Spectra were recorded with a horizontal ATR accessory (FastIR^{TM}, Harrick Scientific) with a ZnSe crystal as the internal reflection element. The spectral resolution was 4 cm^{-1} for all spectra and 50 scans were accumulated with medium apodization for each spectrum. ATR-intensity correction was not applied. Although the high viscosity of several of the crude oils would make it reasonable to perform the IR measurements at elevated temperatures, all IR measurements were carried out at room temperature (20℃) for practical reasons and to obtain a high screening velocity.

7.2.2. Gas chromatography

GC x GC analysis was performed on a double column Hewlett-Packard P 6890 gas chromatograph (Agilent Technologies) equipped with a CIS4 PTV injector, a sulfur chemiluminescence detector, and a liquid nitrogen cryogenic modulation assembly (Zoex Corp.). The first column was a nonpolar DB-1, dimethylpolysiloxane, 10 m, 0.25 mm i.d., 0.25 μm Df. (J&W Scientific) and the second one a medium polarity stationary phase BPX-50, 50% phenyl (equiv.)-polysilphenylene-siloxane, 2 m, 0.10 mm i.d., 0.10 μm Df. (SGE). The modulation capillary was comprised of DPTMDS fused silica tubing, 2 m (1 m in loop), 0.10 mm i.d. (BGB Analytik Vertrieb, Germany). The initial oven temperature for the first dimension column was 40℃. After an initial hold of 5 min, the oven was programmed at a rate of 2.5℃/min to 320℃, which was maintained for 20 min. The secondary oven chamber for the second dimension column had an initial temperature of 90℃. After an initial hold of 5 min., it was programmed at a rate of 2.5℃/min up to 370℃, which was maintained for 20 min. The hot-pulse duration was set to 500 ms, and the modulation time was 10 s. Samples were injected either pure or, when viscosity at 60℃ or S-content did not allow so, diluted with toluene and/or cyclohexane. Concentrations of components in parts per million sulfur (ppm S) were calculated by means of a classified internal standard.

7.2.3. Chemometrics

Modeling was performed using the PLS Toolbox (Eigenvector Research, Inc.) for MatLab (The MathWorks, Inc.) on the IR spectra of the 47 crude oils. This set was divided into a group of 28 spectra for calibration (samples C1-C28) and 19 spectra for validation (samples V1-V19) listed in Table 1 in Chapter 2. As input for modeling of the 9 different sulfur compound classes, the concentrations as determined with 2D-GC have been used. Modeling for the total sulfur content was carried out on the data as determined according to ASTM method D2622.

Prior to modeling, a baseline correction was applied to the IR spectra by subtracting a 3rd degree polynomial fit using the regions 4000-3500, 2500-2000, 1900-1800, 1560-1520, 1000-990 and 650-600 cm^{-1}. Subsequently, the region 2500-1800 cm^{-1} was removed from the spectra since no absorbance bands were observed in this region. Next, pre-processing of the IR spectra was optimized for all 10 sulfur classes by systematic varying pre-process parameters like scaling, smoothing, region selection and spectrum derivative options. This resulted in a selection of 18 different pre-processing methods which are listed in Table 2. MC refers to mean centering and was applied in all cases. The spectral range was either 3500-650 or 1800-650 cm^{-1}. For scaling either the option “none”, Multiplicative Signal Correction (MSC) or Standard Normal Variate (SNV) with and without detrending, 2nd or 3rd order polynomial, was applied. The Savitzky-Golay (SG) smoothing and differentiation parameters were varied from 25-49 points, using a 2nd order polynomial and none, 1st or 2nd derivative. As an example, pre-processing method 13 comprises a SG smoothing with 25 points using a 2nd order polynomial and taking the 1st derivative followed by SNV, detrending with a 2nd order polynomial and MC on the 3500-650 cm^{-1} region. For each of the 18 pre-processing methods, PLS modeling was carried out for the 10 sulfur classes, which resulted in 180 models. From these, the 10 models with the lowest root mean square error of prediction (RMSEP) value for each of the sulfur classes were selected for concentration prediction.

Table 2. Schematic representation of the 18 pre-processing methods used.

1	MC, 1800-650
2	MC, 3500-650
3	MSC, MC, 3500-650
4	SNV, MC, 3500-650
5	SNV, Detrend (2), MC, 3500-650
6	SNV, Detrend (3), MC, 3500-650
7	SG (25 2 0), MC, 3500-650
8	SG (25 2 0), MC, 1800-650
9	SG (25 2 1), MC, 3500-650
10	SG (25 2 1), MC, 1800-650
11	SG (25 2 1), MSC, MC, 3500-650
12	SG (25 2 1), MSC, MC, 1800-650
13	SG (25 2 1), SNV, Detrend (2), MC, 3500-650
14	SG (25 2 1), SNV, Detrend (2), MC, 1800-650
15	SG (35 2 2), MC, 3500-650
16	SG (25 2 2), MSC, MC, 3500-650
17	SG (49 2 2), MSC, MC, 3500-650
18	SG (49 2 2), MSC, MC, 1800-650

7.3. Results and discussion

7.3.1. Gas chromatography

The concentrations in ppm of the 9 different sulfur compound classes, as determined with 2D-GC analysis and S Total analyzed with ASTM method 2622, are presented in Table 3. As appears, the calibration samples C1-C28, used for building the models, as well as the samples applied for validation (V1-V19) represent a wide range of concentrations for the different sulfur speciation classes. It implies that PCA of the IR spectra, as described in Chapter 2, is also valid for this study. To illustrate the results from 2D-GC, a retention time-intensity plot for crude oil C21 is shown as an example in Fig. 1. Note that the physicochemically determined amount of "total S" differs from the summed concentrations as measured with 2D-GC. This is due to the fact that the latter method only covers compounds that elute in the boiling point range from ambient to 465°C, whereas the ASTM method includes elemental and inorganic sulfur.

Table 3. Crude oil samples for calibration (C1-C28) and validation (V1-V19) used for modeling of 10 different sulfur classes. Compound class abbreviations refer to names listed in Table 1. Concentrations (ppm) have been determined with 2D-GC.

Sample	Concentration (ppm)									
	STD	Ar-S	BT	NBT	DBT	NDBT	BNaT	NBNaT	DNaT	S Total
C1	14	8	359	57	742	114	187	26	15	7180
C2	4	1	20	3	59	4	13	2	1	1370
C3	62	254	1083	759	1521	683	519	197	62	11700
C4	53	59	461	191	490	195	171	61	38	5630
C5	11	67	598	332	1038	442	436	164	111	8350
C6	26	15	98	24	85	18	24	9	6	2530
C7	586	272	811	308	412	116	63	18	9	9900
C8	1803	607	6791	929	5335	1068	2678	804	870	54200
C9	138	91	111	58	124	37	50	16	11	3000
C10	62	70	366	167	218	64	31	13	6	4330
C11	385	318	2458	600	2097	638	728	309	205	21900
C12	170	830	2129	2149	3312	1425	1529	632	527	30500

C13	2488	557	6340	1193	4357	1301	1943	726	589	50900
C14	130	612	2096	2020	3193	1270	1415	571	522	31400
C15	269	1412	2704	2306	4248	1345	1597	589	506	38800
C16	1848	421	4907	974	2810	757	773	270	120	47500
C17	1039	262	950	200	1371	250	322	74	78	10800
C18	1622	640	3400	867	2033	756	664	289	105	20500
C19	623	289	5689	1075	4872	1166	1816	438	336	45500
C20	1179	293	1052	209	1475	277	331	72	62	11100
C21	1228	451	2821	929	1694	608	540	268	158	21500
C22	71	239	1606	492	1790	533	705	297	246	14700
C23	1022	501	3070	875	2448	844	1053	471	436	32200
C24	333	182	5114	928	4264	1124	1577	471	315	31700
C25	1448	520	5452	1292	2652	599	566	112	38	30900
C26	57	35	160	58	208	51	57	29	0	4580
C27	871	332	4506	717	4290	1086	1987	714	670	48700
C28	757	247	4264	756	3323	681	1101	201	154	37600
V1	23	22	119	54	175	44	42	24	0	4200
V2	31	48	1700	357	1534	351	364	65	32	15400
V3	684	217	5322	1063	4053	1005	1476	326	274	43000
V4	1759	687	4306	1055	1980	557	486	164	90	33900
V5	20	21	212	68	380	92	121	36	0	4840
V6	966	268	5603	915	3651	785	1247	266	137	41200
V7	252	403	2213	760	1717	415	430	83	42	25200
V8	242	130	5449	991	5010	1195	2070	581	456	32300
V9	2073	665	2902	994	1760	642	531	259	149	28200
V10	881	292	3981	760	3094	876	1298	538	466	44700
V11	1210	437	5146	862	3994	1010	1642	618	473	42600
V12	736	235	3537	600	2614	751	1183	518	444	43100
V13	1279	374	3191	705	1470	271	223	53	23	20300
V14	807	473	2715	668	1248	236	223	61	34	25000
V15	263	150	1121	338	805	191	174	79	38	11300
V16	638	235	608	124	370	77	73	18	9	8130
V17	1149	446	4123	862	3096	871	1097	294	175	31400
V18	174	90	418	150	362	93	99	41	35	7010
V19	675	518	3935	1109	3082	1030	1148	480	479	28500

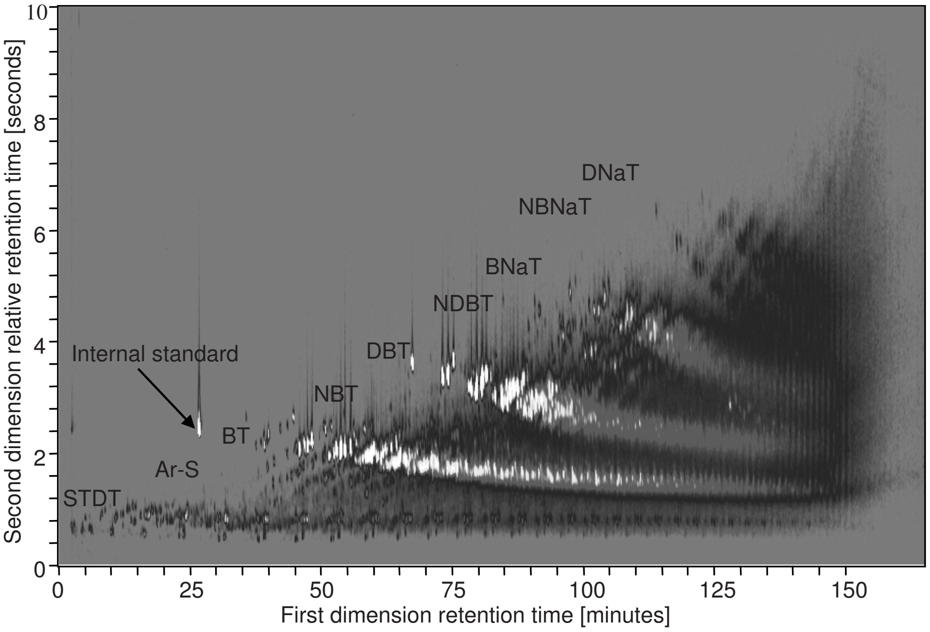


Fig. 1. 2D-GC plot of crude oil C19. S-compound classes and internal standard have been indicated. White colors represent high concentrations, black colors low concentrations.

7.3.2. Infrared spectroscopy

As reported in Chapters 2 and 3, the IR spectra of crude oils are very similar, particularly after intensity normalization and pre-processing. This is illustrated in Fig. 2, showing the overlay of the 28 baseline corrected crude oil spectra of the calibration set C1-C28. All spectra are dominated by strong absorption bands of aliphatic C-H stretching (3000-2800 cm^{-1}) and bending (1470-1350 cm^{-1}) vibrations. Small differences are present in the fingerprint region (1300-650 cm^{-1}). The absorption bands in this region can be merely attributed to aromatic skeletal modes. In general, specific C-S, S-H and/or S-S vibrations are not very IR active because of the small dipole moment change during the vibration of these structural elements [42]. However, for example thiophene rings exhibit several sharp bands related to ring stretching (1550-1350 cm^{-1}) and =C-H out of plane vibrations (800-690 cm^{-1}) [42, 43].

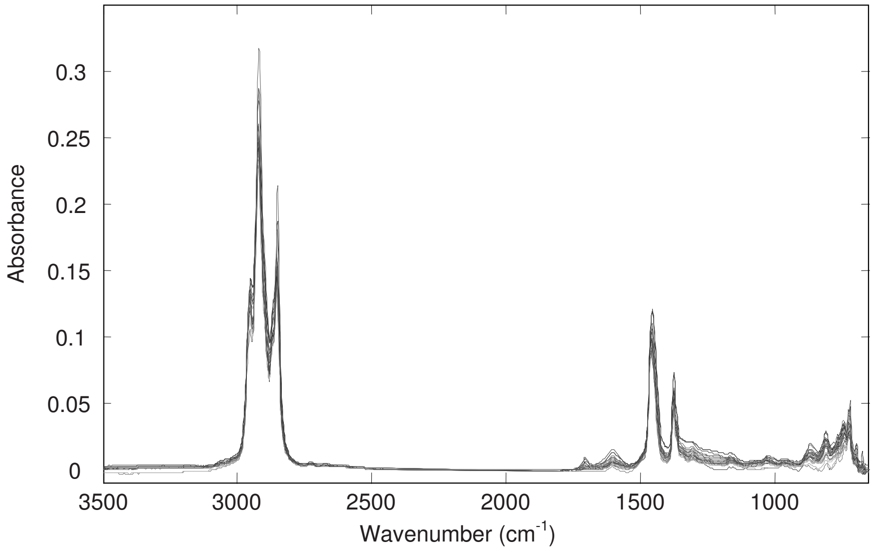


Fig. 2. Overlay of 28 spectra of crude oils as used for calibration of the PLS-models.

7.3.3. Data analysis

As already mentioned, PLS modeling of the 10 sulfur classes, using 18 different pre-processing methods, resulted in 180 models. For these 180 models, the RMSEP values obtained for each speciation class were divided by the standard deviation of the calibration values to express the relative error. These relative errors have been plotted as a function of the pre-processing method for each sulfur speciation class in Fig. 3. The figure illustrates that, independent of the applied pre-processing method, some classes (e.g. NBT) are better predicted than others (e.g. STD). In our opinion, this demonstrates the ability of the models to extract structure related correlations from the IR spectra. Next, the models with the lowest RMSEP values for each of the 10 classes were selected for further evaluation. This is summarized in Table 4, showing for each class, the applied pre-processing method, the mean concentration value for the calibration set, the corresponding standard deviation, the root mean square error of cross validation (RMSECV) value, the number of latent variable (LV's) that was used for the model and the RMSEP values obtained for the validation set. In addition, the corresponding plots of the predicted versus the measured concentrations for the calibration set (●) and the validation set (x) are shown in Fig. 4.

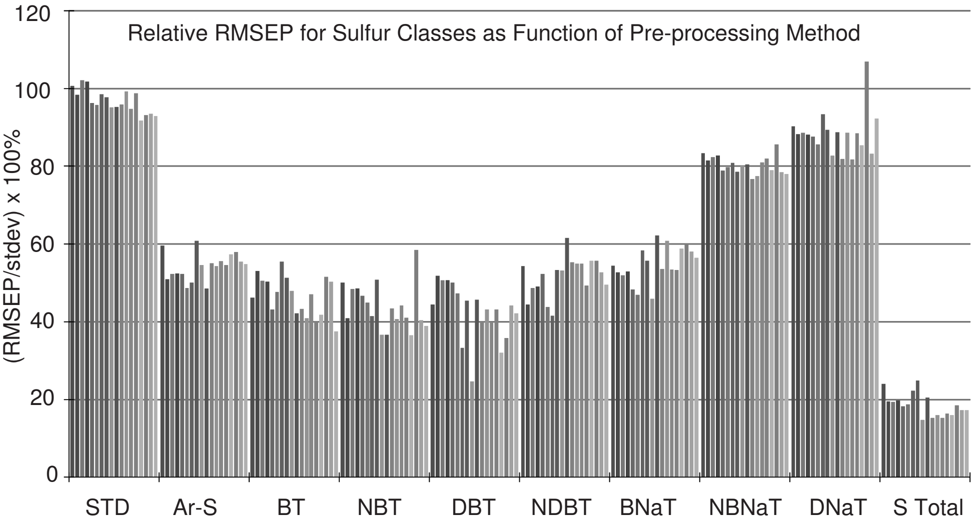
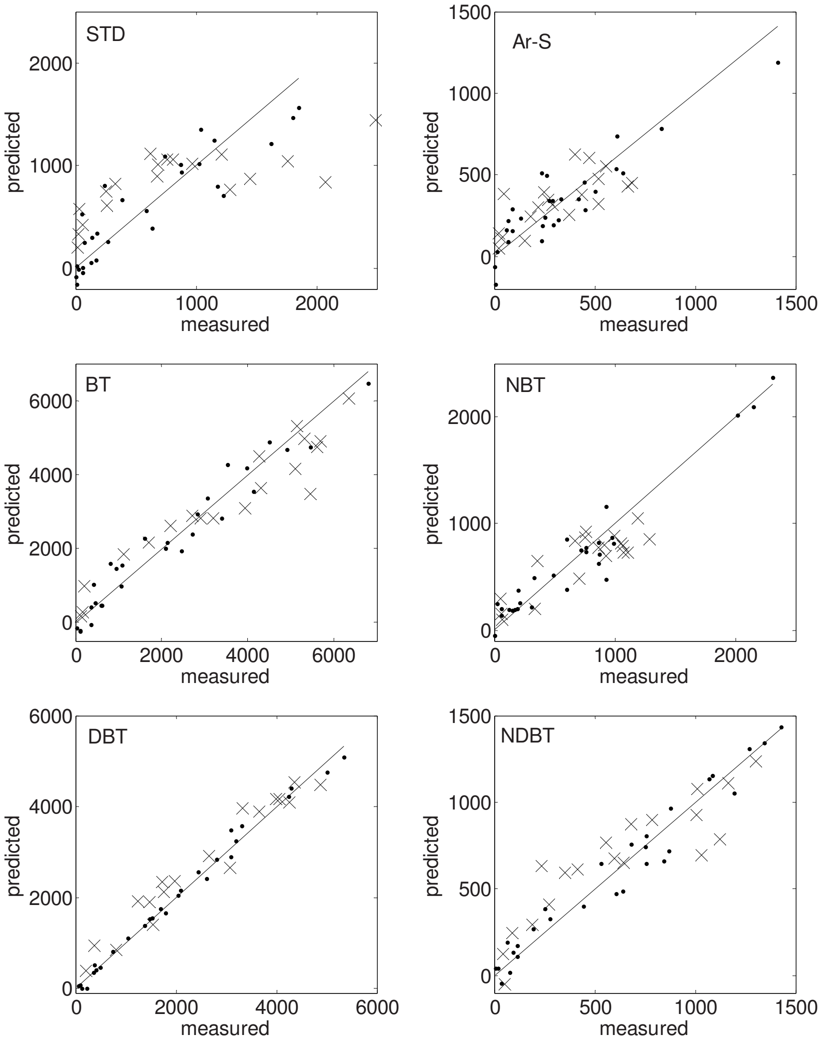


Fig. 3. RMSEP values divided by the standard deviation for the 10 sulfur speciation classes for 18 different pre-processing methods.

Table 4. Results of optimized PLS models to predict the concentrations of 10 different sulfur compound classes in crude oils. Mean concentration, STDEV, RMSECV and LV values refer to the calibration set, RMSEP values to the validation set.

S class	Pre-processing (method number)	Mean (ppm)	STDEV (ppm)	RMSECV (ppm)	LV's	RMSEP (ppm)
STD	15	586	585	448	6	537
Ar-S	10	328	302	229	4	147
BT	18	2161	1864	769	6	700
NBT	15	666	625	306	6	228
DBT	9	1976	1553	632	10	383
NDBT	7	585	451	213	8	187
BNaT	9	773	722	367	5	331
NBNaT	11	281	253	95	8	194
DNaT	13	227	243	131	4	199
S Total	9	21587	16351	5403	7	2520


STD
2000
1000
0
predicted
0
1000
2000
measured
1500
Ar-S
1000
500
0
predicted
0
500
1000
1500
measured
BT
6000
4000
2000
0
predicted
0
2000
4000
6000
measured
NBT
2000
1000
0
predicted
0
1000
2000
measured
6000
DBT
4000
2000
0
predicted
0
2000
4000
6000
measured
1500
NDBT
1000
500
0
predicted
0
500
1000
1500
measured

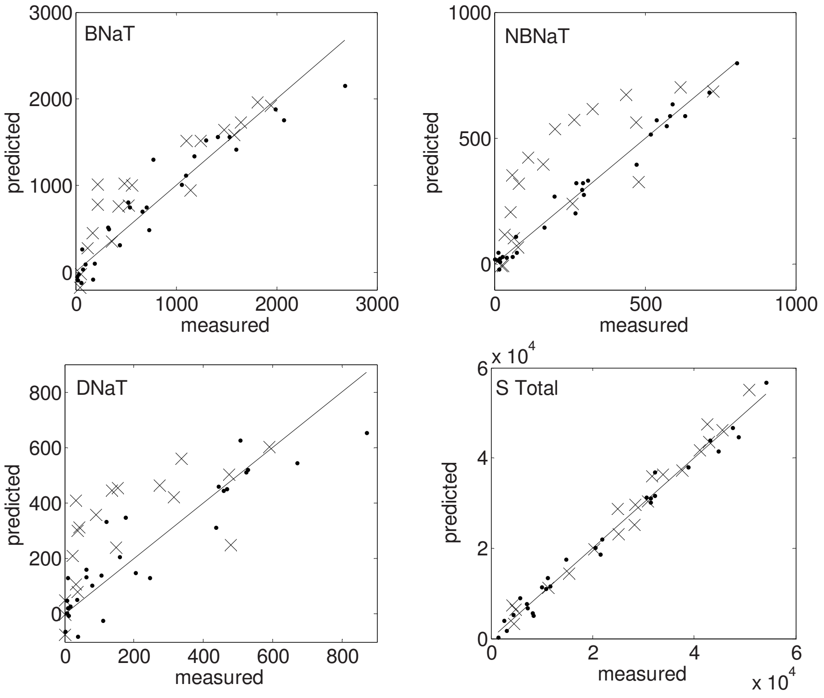


Fig. 4. Prediction plots of PLS modeling the concentration of 10 sulfur speciation classes of crude oils based on their IR spectra. Calibration spectra (•) and validation spectra (x). The corresponding pre-processing methods are listed in Table 4.

First of all, the results confirm the conclusion from the previous chapters that the prediction of the total sulfur concentration of crude oils by means of PLS modeling of the IR spectra is a valuable alternative for ASTM method 2622. Furthermore, the model to predict the dibenzothiophenes (DBT) is promising followed by the related benzothiophene compound classes BT, NBT and NDBT. The correct prediction of DBT concentrations is particularly interesting in view of the fact that these compounds are the major sulfur containing species left in fuels after desulfurization. Moreover, the models for the speciation of DBT together with BT, NBT and NDBT might be useful, as this type of compounds is known to hamper efficient crude oil processing and refining. Finally, the models for the remaining classes STD, Ar-S, BNaT, NBNaT and DNaT are not very useful for concentration prediction. The differences in the predictive power of the models can be explained by the assumption that vibrations related to benzothiophene structures are well represented in the IR spectra, whereas other sulfur containing functional groups lack specific sulfur related absorption bands.

7.4. Conclusions

PLS modeling of the IR spectra of crude oils is a valuable alternative to ASTM method 2622 to predict the total sulfur content of these materials. The application as a tool for sulfur speciation, however, is limited. From the 9 different sulfur compound classes that are usually determined with standard 2D-GC analysis, the model to predict the concentration of dibenzothiophenes is promising. Also, the regression models for the related benzothiophene compound classes BT, NBT and NDBT perform reasonably well. However, the models for the remaining classes STD, Ar-S, BNaT, NBNaT and DNaT are not very useful. As such, PLS regression modeling is not as widely applicable for sulfur speciation as 2D-GC. On the other hand, it can be a fast method for the qualitative and quantitative on-site screening on dibenzothiophenes, a class of compounds which is known to be detrimental in crude oil processing and a predominant sulfur-residual in fuels.

References

1. R. Hua, Y. Li, W. Liu, J. Zheng, H. Wei, J. Wang, X. Lu, H. Kong, G. Xu, J. Chromatogr. A 1019 (2003) 101.
2. J. Beens, R. Thijssen, J. High Resol. Chromatogr. 20 (1997) 131.
3. A.G. Marshal, R.P. Rodgers, Acc. Chem. Res. 37 (2004) 53.
4. M.F. Ali, A. Al-Malki, B. El-Ali, G. Martinie, M.N. Siddiqui, Fuel 85 (2006) 1354.
5. Directive on fuel quality, 98/70/EC as amended by 2003/17/EC.
6. Z. Ring, J. Chen, H. Yang, H. Du, Y. Briker, Proceed. AIChE Spring National Meeting, New Orleans, April 25-29 (2004) 1386.
7. Z. Xialolan, J. Jun, L. Jianhua, Y. Yongtan, Chin. J. Anal. Chem. 34 (2006) 1546.
8. Y.T. Yang, H.Y. Yang, W.Z. Lu, Chin. J. Chromatogr. 20 (2002) 493.
9. D.M. Serafim. N.R. Stradiotto, Fuel 87 (2008) 1007.
10. S. Sinkkonen, J. Chromatogr. 475 (1989) 421.
11. J. Beens, J. Blomberg, P.J. Schoenmakers, J. High Resol. Chrom. 23 (2000) 182.
12. J. Blomberg, P.J. Schoenmakers, U.A.Th. Brinkman, J. Chromatogr. A, 972 (2002) 137.
13. J. Blomberg, T. Riemersma, M. Van Zuijlen, H. Chaabaani, J. Chromatogr. A, 1050 (2004) 77.

14. A.H. Hegazi, J.T. Andersson, M.A. Abu-Elgheit, M. Sh. El-Gayar, Polyc. Arom. Comp. 24 (2004) 123.
15. M.K. Andari, H. Behbehani, A. Stanislaus, Fuel Science & Technol. Int. 14 (1996) 939.
16. H.S.J. Behbehani, Abstracts, 219th ACS National Meeting, San Francisco, CA, March 26-30, (2000) PETR-057.
17. R. Hua, J. Wang, H. Kong, J. Liu, X. Lu, G. Xu, J. Sep. Sci. 27 (2004) 691.
18. I.C. Lee, H.C. Ubanyionwu, Fuel 87 (2008) 312.
19. O. Glinzer, D. Severin, C. Beduerftig, C.D. Czogalla, U. Puttins, Fres. Zeitschr. Anal. Chem. 315 (1983) 208.
20. I. Dzidic, M.D. Balicki, I.A.L. Rhodes, H.V. Hart, Haskell, J. Chromatogr. Sci. 26 (1988) 236.
21. J.D. Payzant, D.S. Montgomery, O.P. Strausz, AOSTRA J. Research 4 (1988) 117.
22. M. Nishioka, R.S. Tomich, Fuel 72 (1993) 1007.
23. J. Sinninghé Damsté, W.I.C. Rijpstra, J.W. de Leeuw, G.W.M. Lijmbach, J. High Res. Chrom. 17 (1994) 489.
24. X. Ma, K. Sakanishi, T. Isoda, I. Mochida, Fuel 76 (1997) 329.
25. G.S. Waldo, O.C. Mullins, J.E. Penner-Hahn, S.P. Cramer, Fuel 71 (1992) 53.
26. C.E. Snape, K. Ismail, S.C. Mitchel, K. Bartle, in *Composition, geochemistry and conversion of oil shales*, C.E. Snape (Ed.), Kluwer Academic Publishers (1995) 125.
27. C.E. Snape, J. Yperman, D. Franca, K. Bartle, Europ. Comm. Report (1998) EUR 17947, 1.
28. L.R. Barker, W.R. Kelly, W.F. Guthrie, Energ. Fuels 22 (2008) 2488.
29. G.S. Waldo, R.M.K. Carlson, J.M. Moldowan, K.E. Peters, J.E. Penner-Hahn, Geochim. Cosmochim. Acta 55 (1991) 801.
30. M. Kasrai, G.M. Bancroft, R.W. Brunner, R.G. Jonasson, J.R. Brown, K.H. Tan, X. Feng, Geochim. Cosmochim. Acta 58 (1994) 2865
31. G. Sarret, J. Connan, M. Kasrai, L. Eybert-Berard, G.M. Bancroft, J. Synchrotr. Rad. 6 (1999) 670.
32. A. Mijovilovich, L.G.M. Pettersson, S. Mangold, M. Janousch, J. Susini, M. Salomé, F.M.F. de Groot, B.M. Weckhuysen, J. Phys. Chem. A 113 (2009) 2750.
33. C.E. Snape, S.C. Mitchel, K. Ismail, R. Garcia, Rev. Anal. Chem. Euroanalysis VIII (1994) 103.
34. S. Guan, A.G. Marshall, S.E. Scheppele, Anal. Chem. 68 (1996) 46.
35. C.A. Hughey, P.R. Rodgers, A.G. Marshall, K. Qian, W.K. Robbins, Org, Geochem. 33 (2002) 743.
36. G.C. Klein, R.P. Rodgers, A.G. Marshall, Fuel 85 (2006) 2071.
37. C.A. Hughey, S.A. Galasso, J.E. Zumberge, Fuel 86 (2007) 758.
38. S.K. Panda, W. Schrader, A. Al-Hajji, J.T. Anderson, Energ. Fuels 21 (2007) 1072.
39. R. Saetre, A. Somogyvari, Preprints – Am. Chem. Soc., Div. Petrol. Chem. 34 (1989) 268.

40. F.I. Samedova, G.S. Martynova, Y.G. Yusifov, B.A. Guseinova, E.G. Ismailov, Azarbaycan Neft Tasarrufati 4 (2008) 39.
41. P. de Peinder, D.D. Petrauskas, F. Singelenberg, F. Salvatori, T. Visser, B. M. Weckhuysen, PCT Patent Appl. (2008) WO 2008/135411.
42. D. Lin-Vien, N.B. Colthup, W.G. Fateley, J.G. Graselli, (Eds), *Handbook of Infrared and Raman Characteristic Frequencies of Organic Molecules*, Academic Press, New York (1991).
43. Infrared and Raman Interpretation Support software (IRIS 3.0), Thiophenes, http://www.vibspec.com.

Chapter 8

Summary and Concluding Remarks

8.1. General considerations

Nowadays, molecular spectroscopic analysis is the method of choice to obtain detailed qualitative and quantitative information of complex structures and mixtures. Valuable chemical and physical data can be extracted from the spectral data, particularly, when chemometric modeling techniques are used. This combined approach might as well be useful to determine the physicochemical properties of crude oils. Knowledge of parameters such as density and viscosity, for instance, are important to determine whether it is economically justified to exploit an oil reservoir, whereas *e.g.*, the sulfur and asphaltene content is crucial to define the catalytic and refining conditions. If successful, a combined spectroscopic-chemometric method offers a fast alternative for the elaborate and occasionally complex crude oil assays that are currently used in petrochemical industries. This challenging objective has been the major motivation of this PhD work to explore the potential of a number of molecular spectroscopic methods in combination with chemometric modeling techniques, to classify crude oils in terms of a series of physicochemical properties.

The primary goal of the research conducted was to develop models to predict so-called long residue (LR) and short residue (SR) properties directly from a single crude oil spectrum. An additional aim was to study the usefulness of this integrated approach as a tool for sulfur speciation and possibly to determine structural information of these species on a molecular scale. This topic is important to improve the efficiency of the current catalytic hydrodesulfurization processes. Six different spectroscopic techniques have been explored: IR, NIR, Raman, UV-Vis, ^{1}H-NMR and ^{13}C-NMR. Principal Component Analysis (PCA) and Partial Least Squares (PLS) regression have been selected as modeling techniques, since these standard methods allow direct implementation of the developed hard- and software in crude oil production plants and related refineries. The approach that has been followed to achieve the stated goals is schematically listed in Table 1 together with the topics and techniques that have been studied.

Table 1. Schematic representation of the research performed in this PhD thesis.

Chapter	Technique						Properties		Topic
	IR	NIR	NMR (^{1}H)	NMR (^{13}C)	Raman	UV-Vis	LR	SR	
2	+	+	+	+	+	+	+	+	Utility testing
3	+		+	+			+		Complementarity
4	+						+	+	Temperature effects
5	+						+		Mathematical blending
6	+						+		Model robustness
7	+						+		Sulfur speciation
8	+						+	+	Model implementation

8.2. Prediction of long and short residue properties

From the six spectroscopic techniques that have been explored in this work, Raman and UV-Vis spectroscopy were found not to be useful to establish the goals set. As outlined in **Chapter 2**, the acquisition of useful Raman spectra was hampered by extreme fluorescence and self-absorption phenomena, while the interpretative value of the UV-Vis spectra was limited as result of the lack of details in the electronic transition bands. PLS modeling of the ^{1}H- and ^{13}C-NMR spectra provided rather disappointing prediction models in view of the intrinsically high information content of NMR data. This appeared to be due to a poor reproducibility of the spectral data points, a drawback that is frequently encountered in NMR spectrometry combined with multivariate data analysis. The best results in predicting the LR properties were obtained for PLS modeling of the IR and NIR spectra. For that reason, modeling of these data types has been studied in great detail, including the development of models to predict SR properties. It was concluded that both methods provide a valuable alternative for the time-consuming ASTM and IP methods, which are currently used in petrochemical industries to determine LR and SR properties. The IR models perform slightly better than the NIR ones probably since the corresponding spectra reflect the fundamental vibrational transitions, whereas NIR only exhibits overtones and combination bands. From the 7 different LR properties for which IR models were built, *i.e.*, yield-long-on-crude (YLC), density (D_{LR}), viscosity (V_{LR}), pour point (PP), asphaltenes (Asph), carbon residue (CR) and sulfur content (S), the ones to predict YLC, D_{LR} and V_{LR} were found to be the most reliable while the inaccuracy in the prediction of PP is similar to that of the corresponding ASTM method. The scores for S, Asph and CR are less good, but still useful for indicative purposes. The results to predict SR properties from IR spectra are also promising, particularly for the density (D_{SR}) and the viscosity (V_{SR}), when modeling was carried out as a function of the Atmospheric equivalent Flash Temperature (AFT). The use of yield-short-on-crude (YSC) instead of AFT as input did not provide better results. The resulting methodology has been patented.

Chapter 3 summarizes the results of PLS modeling of the spectra of two complementary methods, *i.e.*, IR and NMR spectroscopy, to predict crude oil LR properties. After applying a data point shift correction to the NMR spectra in order to compensate for the data point inaccuracies, as observed in Chapter 2, IR spectra were merged with ^{1}H-NMR and/or ^{13}C-NMR data and used as input for modeling. Surprisingly, this did not lead to a significant enhancement of the results compared to the models, based on separate IR, ^{1}H-NMR or ^{13}C-NMR data. It implies that combining the spectra of complementary spectroscopic methods does not automatically lead to better results. For all combined data sets, prediction accuracies for the properties YLC, V_{LR}, D_{LR} and PP were obtained that are of the same order of magnitude as the reproducibility of the previously mentioned ASTM and IP methods. However, since the best results were obtained with the models based on only IR spectra, this technique is to be preferred over NMR for the prediction of LR properties of crude oils at site, especially since IR spectroscopy can be realized much easier at lower cost and does not require any special sample preparation.

As a next step to improve the LR and SR prediction models, temperature effects were studied in **Chapter 4**. The exposure of the crude oils for 24 h to 65°C to reduce the contributions of volatile constituents in the spectra, did not lead to an increase in the results of predicting the LR and SR properties. Modeling of the spectra of the same sample, recorded over the range 20 to 65°C with increments of 5°C, showed a linear correlation with the temperature. This effect was assigned to a decrease of the density with increasing temperature. Application of mean centering for temperature prior to modeling of the IR spectra recorded at different temperatures also revealed a linear relationship with the temperature. More detailed analysis of the prediction results showed a decrease in the number of used LV's for the prediction models of the LR-properties. A minor improvement in the RMSECV values for the SR prediction models was observed if the variable temperature data were used.

Chapter 5 describes the results of the study to determine the applicability of the IR models to predict the LR properties of mathematically created spectra of blends. These mimicked blend spectra were produced by co-adding the IR spectra of two crude oils in various weight ratio. The results of prediction were compared with those obtained from the spectra of the same, physically prepared blends. The predicted LR properties of the artificial and the physically mixed blends turned out to be largely the same. For that reason, it is concluded that 'artificial blending' of crude oil IR spectra is a fast and clean desk-top alternative for physically mixing blends in the laboratory. Moreover, the method can be used as a rapid testing and screening tool for large series of potential blends. Compared to the spectra of real blends, slightly larger deviations are observed for some of the LR properties when submitting artificial blend spectra to the prediction models. This is assigned to non-linear effects that occur when blends are physically prepared, but which are not reflected in the mathematically created blend spectra.

In order to determine the utility of the developed IR prediction models on different crude oil production sites, the robustness of the LR prediction models was studied in **Chapter 6** by modeling spectra that were recorded on different instrumental set-ups. First of all, it was concluded that spectra, recorded at different instruments and with different sample cells, are highly reproducible. This confirms that the experimental protocol is correct. Also, large time delays in the acquisition of the spectra did not influence the accuracy of the predicted property values. The effect of the applied spectrometer and ATR accessory on the performance of the prediction models appeared to be very small if the calibration spectra and the ones used for validation were recorded on the same instrument. In case validation spectra were recorded on a different instrument than the one that was used for the calibration set, the accuracy decreases with a factor of 2.

Finally, the result of PLS-modeling of IR spectra as a tool for sulfur speciation of crude oils was investigated in **Chapter 7**. Although, this approach was found to be successful for the prediction of the total sulfur content, as demonstrated in Chapters 2-6, the potential for sulfur speciation turned out to be limited compared to the 2D-GC analysis, currently used for this purpose. On the other hand, it was concluded that the models to predict the concentration of dibenzothiophenes (DBT) and the benzothiophene classes BT, NBT and NDBT perform reasonably well. It offers the possibility to use these models as a tool for fast indicative on-site screening on benzothiophenes. Knowledge of the concentration of this compound class is very helpful to determine the optimal parameters for crude oil processing.

8.3. Future developments

Thus far, the prediction models based on IR spectra have shown to provide the highest accuracies. From a theoretical point of view, it is not very likely that models based on NIR data will provide better results than the IR ones. However, NIR spectroscopy has a number of practical advantages over IR, particularly with regards to sampling and signal to noise ratio. As such, the design of an NIR instrument, dedicated to crude oil measurement, might lead to more reproducible spectra and thus better models. Another option for improvement of the IR and NIR models is the use of more accurate property values as input for modeling. Particularly, the physicochemically determined values for the LR property PP and the SR properties P and R&B are rather insecure.

The currently available models for the prediction of SR properties at various AFT values are only capable of prediction linear correlations. It is known that the P values, especially for softer crudes, show non linear behavior for the complete AFT range. In order to overcome this limitation the utility of non linear multivariate regression methods should be explored. An alternative approach is the combination of the IR spectrum of the crude oil with AFT and an additional input value, e.g. YSC, to describe the non linear relationship.

Optimization could also be achieved by selecting specific parts of the spectrum for modeling on the basis of prior IR knowledge. For example, modeling of only that part of the spectrum that is characteristic for aromatic compounds, might give rise to improved results for the Asph prediction model. The use of advanced algorithms for wavelength selection in combination with PLS could also be employed to determine the important spectral regions. In addition, the application of alternative spectral pre-processing will provide insight in ways to optimize modeling.

Spectral data acquisition at different temperatures resulted in small improvement for some of the models. However, a more detailed study of this topic is necessary to determine the significance of the observed effect. The same holds for the non-linear trends that showed up in the spectra of the physically prepared blends. The artificial blends are constructed by straight forward linear addition of the IR spectra. An additional pre-processing step in the mathematical blend creation that corrects for this type of non-linearity might be a way to reduce the prediction errors.

Finally, the most important issue is to allow the application of the prediction models at different crude oil production plants. Developing robust prediction models which are insensitive to instrument configuration and environment, for instance by applying wavelength selection, could be part of a solution. Alternatively a calibration transfer protocol could be developed. This transfer protocol should adjust the available model to work with the spectra of the new set-up or it should convert the spectra of the new set-up to a format that can be used with the available model. Further research on this field, particularly focused on the development of a calibration transfer protocol for each data acquisition station to be used, is considered inevitable.

Samenvatting en Conclusies

Algemeen

Moleculair spectroscopische analyse is een belangrijke manier om gedetailleerde kwalitatieve en kwantitatieve informatie van complexe structuren en mengsels te verkrijgen. Waardevolle chemische en fysische data kunnen uit deze spectroscopische data worden geëxtraheerd, vooral in combinatie met chemometrische modelleertechnieken. Deze gecombineerde benadering zou ook zeer nuttig kunnen zijn om de fysisch-chemische eigenschappen van ruwe aardolie te bepalen. Kennis van de parameters zoals bijvoorbeeld dichtheid en viscositeit zijn belangrijk om te bepalen of het economisch gerechtvaardigd is om een aardoliereservoir te exploiteren. De hoeveelheid zwavel en asfaltenen zijn bijvoorbeeld cruciaal om de katalytische en raffinagecondities te bepalen. Een geslaagde combinatie van een spectroscopische en chemometrische methode biedt een snel alternatief voor de arbeidsintensieve en soms complexe analyse van ruwe aardolie die momenteel gebruikt wordt in de petrochemische industrie. Dit uitdagende doel vormde het startpunt van dit promotieonderzoek naar de mogelijkheden van een aantal moleculair spectroscopische methoden gecombineerd met chemometrische modelleer technieken, om ruwe aardolie te classificeren op basis van een reeks fysisch-chemische eigenschappen.

Het belangrijkste doel van het onderzoek was om modellen te ontwikkelen die de zogenaamde 'long residue' (LR) en 'short residue' (SR) eigenschappen voorspellen aan de hand van een spectrum van een ruwe aardolie. Het LR ontstaat na atmosferische destillatie van een ruwe aardolie (>370℃) en het SR na vacuümdestillatie van het LR. Deze LR en SR eigenschappen zijn ook essentieel bij de productie van bitumen, een belangrijk onderdeel van asfalt. Een bijkomend doel was te onderzoeken of deze geïntegreerde aanpak ook geschikt zou kunnen zijn als zwavel speciatie methode en om structuurinformatie van zwavel componenten op een moleculaire schaal te verkrijgen. Dit onderdeel is belangrijk om de efficiëntie van de huidige katalytische ontzwavelingsprocessen te verbeteren. Zes verschillende spectroscopische technieken werden onderzocht: Infrarood (IR), Nabij Infrarood (NIR), Raman, UV-Vis, ^{1}H-NMR en ^{13}C-NMR. Principale componenten analyse (PCA) en 'Partial Least Squares' (PLS) regressie werden geselecteerd als de multivariate modelleer technieken, omdat deze standaardmethoden directe implementatie van de ontwikkelde software mogelijk maken bij zowel de winning van ruwe aardolie als bij de raffinage. De gevolgde strategie is schematisch weergegeven in Tabel 1 samen met de eigenschappen en technieken die onderzocht zijn.

Tabel 1. Schematische weergave van het wetenschappelijk onderzoek beschreven in dit proefschrift.

Hoofdstuk	Techniek						Eigenschappen		Onderwerp
	IR	NIR	NMR (^{1}H)	NMR (^{13}C)	Raman	UV-Vis	LR	SR	
2	+	+	+	+	+	+	+	+	Haalbaarheid studie
3	+		+	+			+		Complementariteit
4	+						+	+	Temperatuur effecten
5	+						+		Mathematisch mengen
6	+						+		Model robuustheid
7	+						+		Zwavel speciatie
8	+						+	+	Model implementatie

Voorspelling van de LR en SR eigenschappen

Van de zes spectroscopische technieken die onderzocht zijn in dit promotieonderzoek bleken Raman en UV-Vis niet geschikt om de gestelde doelen te bereiken. Zoals beschreven in **Hoofdstuk 2**, konden geen bruikbare Raman spectra worden opgenomen vanwege extreme fluorescentie en zelfabsorptie. Tevens werd de interpreteerbaarheid van de UV-Vis spectra beperkt door een gebrek aan detail in de UV-Vis absorptiebanden. De bruikbaarheid van de ontwikkelde PLS modellen om de LR eigenschappen van ruwe aardolie te voorspellen aan de hand van ^{1}H- en ^{13}C-NMR spectra bleek kleiner dan werd verwacht op basis van de intrinsieke hoge informatiedichtheid van NMR data. Dit kon echter met name worden toegeschreven aan de beperkte reproduceerbaarheid van de datapunten in de NMR spectra. Dit verschijnsel blijkt vaker problemen op te leveren bij de toepassing van multivariate data analysetechnieken op NMR gegevens. De beste resultaten voor de voorspellingen van de LR eigenschappen werden behaald met de modellen gebaseerd op IR en NIR spectra. Om die reden zijn, voor deze spectrale data, de modellen tot in detail bestudeerd, inclusief de ontwikkeling van modellen om de SR eigenschappen te voorspellen. Het bleek dat beide methoden een waardevol alternatief bieden voor de tijdrovende ASTM en IP standaardmethoden die momenteel in de petrochemische industrie gebruikt worden om de LR en SR eigenschappen te bepalen. De IR modellen presteren in het algemeen iets beter dan de NIR modellen. Dit kan worden toegeschreven aan het feit dat IR spectra met name de fundamentele vibratieovergangen weergeven, terwijl NIR spectra alleen boventonen en combinatiebanden bevatten. Van de 7 verschillende LR eigenschappen waarvoor IR modellen zijn ontwikkeld, nl. de opbrengst-LR-uit-ruwe aardolie (YLC), de dichtheid (D_{LR}), de viscositeit (V_{LR}), het schenkpunt (PP), de asfaltenen concentratie (Asph), het koolstof residu (CR) en het zwavel gehalte (S), bleken de meest betrouwbare modellen die voor YLC, D_{LR} en V_{LR}. Tevens was de fout in de voorspelling voor PP vergelijkbaar met die van de corresponderende ASTM methode. De prestaties voor S, Asph en CR waren minder goed, maar de modellen zijn nog wel bruikbaar voor indicatieve doeleinden. De resultaten om de SR eigenschappen te voorspellen op basis van de IR spectra waren ook veelbelovend, met name voor de dichtheid (D_{SR}) en de viscositeit (V_{SR}), als de modellen ontwikkeld werden als functie van de atmosferisch equivalente vlampunt temperatuur (AFT). Het gebruik van opbrengst-SR-uit-ruwe aardolie (YSC) in plaats van AFT leverde minder goede voorspellingen op. De ontwikkelde methode is gepatenteerd.

In **Hoofdstuk 3** zijn de resultaten beschreven van het PLS modelleren van de spectra van twee complementaire methoden om de LR eigenschappen van ruwe aardolie te voorspellen, nl. IR en NMR spectroscopie. Na het toepassen van een datapunt verschuiving op de NMR spectra om te corrigeren voor de onnauwkeurigheid in de x-as, zoals vermeld in Hoofdstuk 2, werden de IR spectra gecombineerd met de ^{1}H-NMR en/of ^{13}C-NMR data en gebruikt als input voor het modelleren. Tegen de verwachting in leidde dit niet tot een significante verbetering van de voorspellingen in vergelijking met de modellen die gebaseerd waren op de afzonderlijke IR, ^{1}H-NMR of ^{13}C-NMR data. Hieruit is geconcludeerd dat het combineren van spectra van complementaire technieken niet per definitie leidt tot betere resultaten. Voor alle gecombineerde data sets bleek de fout in de voorspelde waarden van YLC, V_{LR}, D_{LR} en PP van dezelfde orde van grootte als de reproduceerbaarheid van de ASTM en IP methoden. Aangezien de beste resultaten werden verkregen met de modellen gebaseerd op alleen de IR spectra geniet deze techniek de voorkeur boven NMR om de LR eigenschappen van ruwe aardolie te voorspellen. Bovendien is IR spectroscopie goedkoper en eenvoudiger en derhalve meer geschikt voor gebruik in een raffinaderij of op een olieplatform.

Als een volgende stap om de LR en SR voorspelmodellen, gebaseerd op IR spectra, te verbeteren werden twee temperatuureffecten bestudeerd in **Hoofdstuk 4**. Het 24 uur lang blootstellen van de ruwe aardolie aan 65℃, om zodoende de bijdrage van de vluchtige componenten in de spectra te verminderen, leverde geen verbetering op in het voorspellen van de LR en SR eigenschappen. Het modelleren van de spectra, verkregen door de olie te verwarmen van 20 tot 65℃ met stappen van 5℃, vertoonde een lineair verband met de temperatuur. Dit effect werd toegewezen aan een afname van de dichtheid van het monster met een toename van de temperatuur. Ook het aftrekken van het gemiddelde spectrum van de temperatuurreeks van ieder individueel spectrum, voorafgaande aan het modelleren, resulteerde in een lineair verband met de temperatuur. Een meer gedetailleerde analyse van de voorspellingen liet een afname in het aantal LV's zien voor de LR-voorspelmodellen. Ook werd een kleine verbetering in de RMSECV-waarden van de SR voorspelmodellen waargenomen indien de variabele temperatuur spectra werd gebruikt.

In **Hoofdstuk 5** zijn de resultaten beschreven van het onderzoek naar de bruikbaarheid van de IR modellen om LR-eigenschappen te voorspellen aan de hand van nagebootste IR spectra van aardoliemengsels, ook wel genoemd 'blends'. Deze blend-spectra werden gecreëerd door de spectra van twee ruwe aardoliën bij elkaar op te tellen in de gewenste gewichtsamenstellingen. De voorspelresultaten van deze kunstmatig verkregen spectra werden vervolgens vergeleken met de resultaten die werden verkregen met de spectra van de daadwerkelijk geprepareerde aardoliemengsels. De voorspelde LR eigenschappen van de oliemengsels bleken nagenoeg hetzelfde voor de nagebootste en de echte spectra. Voor sommige LR eigenschappen werden iets grotere voorspelfouten geconstateerd bij gebruik van de nagebootste spectra. Dit is toegekend aan niet-lineaire effecten, zoals intermoleculaire interacties, welke optreden wanneer aardoliën echt gemengd worden en die niet verwerkt zijn in de mathematisch gecreëerde spectra. Echter, op basis van de verkregen resultaten kon worden geconcludeerd dat het softwarematig nabootsen van ruwe aardoliespectra een snel en schoon alternatief is voor het daadwerkelijk prepareren van blends in het laboratorium. Bovendien kunnen met de ontwikkelde methode grote series van potentiële blends getest en geselecteerd worden.

Om inzicht te krijgen in hoeverre de ontwikkelde LR voorspelmodellen universeel toepasbaar zijn, dat wil zeggen op verschillende olieproductie- en raffinage-locaties, is onderzoek gedaan naar de robuustheid van de modellen. De resultaten hiervan zijn beschreven in **Hoofdstuk 6**. De studie is uitgevoerd door het modelleren van de IR spectra van dezelfde verzameling aardoliemonsters, welke op verschillende instrumentele opstellingen werden opgenomen. Allereerst werd geconcludeerd dat de spectra, afkomstig van verschillende IR spectrometers en met de verschillende optische accessoires, zeer reproduceerbaar zijn. Dit bevestigde de juistheid van het opgestelde experimentele protocol. Daarnaast bleek dat een groot tijdsverschil tussen opnames van de spectra geen invloed heeft op de nauwkeurigheid van de voorspelde eigenschappen. Ook werd vastgesteld dat het effect van de gebruikte spectrometer en accessoire op de nauwkeurigheid van de voorspelmodellen erg klein is als zowel de calibratie- als validatiespectra op dezelfde instrumentele opstelling worden opgenomen. Indien de validatiespectra op een ander instrument worden opgenomen dan de calibratiespectra, dan vermindert de nauwkeurigheid met een factor 2.

Als laatste werd in **Hoofdstuk 7** de toepasbaarheid van PLS modellen op basis van IR spectra onderzocht als alternatieve methode voor de speciatie van zwavel in ruwe aardolie. Zoals reeds werd vastgesteld in de Hoofdstukken 2-6, is deze aanpak succesvol voor de voorspelling van het totale zwavelgehalte. De potentie voor zwavelspeciatie bleek echter beperkt in vergelijking met de 2D-GC analyse die momenteel voor dit doel wordt gebruikt. Wel kon worden geconcludeerd dat de modellen om de concentratie van dibenzothiophenen (DBT) en de benzothiopheen klassen BT, NBT en NDBT te voorspellen, redelijk presteren. Dit biedt de mogelijkheid om deze modellen te gebruiken voor een snelle indicatieve analyse van aardolie op benzothiophenen. Deze kennis is belangrijk om de optimale parameters voor de opwerking van ruwe aardolie te bepalen.

Toekomstige ontwikkelingen

De modellen gebaseerd op de IR spectra hebben tot nu toe de beste resultaten opgeleverd. Vanuit theoretisch oogpunt is het niet waarschijnlijk dat het gebruik van NIR data tot betere resultaten zal leiden. Echter, NIR spectroscopie heeft een aantal praktische voordelen ten opzichte van IR, met name voor wat betreft bemonstering en de signaal-ruis verhouding. Een NIR instrument geoptimaliseerd voor ruwe aardolie metingen zou tot beter reproduceerbare spectra en dus tot betere modellen kunnen leiden. Een andere optie voor verbetering van de IR en NIR modellen is het gebruik van nauwkeuriger referentiewaarden van de LR- en SR-eigenschappen als input voor het modelleren. De fysisch-chemisch bepaalde waarden voor de LR eigenschap PP en de SR eigenschappen P en R&B zijn bijvoorbeeld niet erg betrouwbaar.

De momenteel beschikbare modellen voor het voorspellen van de SR eigenschappen bij verschillende AFT waarden zijn alleen in staat om een lineair verband te voorspellen. Het is bekend dat de P-waarden, vooral van ‘zachte’ olie, een niet lineair gedrag vertonen voor het volledige AFT bereik. Om deze beperking op te heffen zullen niet-lineaire multivariate regressie methoden onderzocht moeten worden. Een alternatieve aanpak om deze niet-lineariteit te kunnen beschrijven is het combineren van een IR spectrum van een ruwe aardolie met de AFT-waarde en een additionele inputwaarde, zoals bijvoorbeeld YSC. Ook kan worden geprobeerd te optimaliseren door specifieke delen van het spectrum te selecteren voor het modelleren op basis van IR kennis. Het modelleren van het deel van het spectrum dat karakteristiek is voor aromatische componenten zou bijvoorbeeld kunnen leiden tot een verbetering voor het Asph voorspelmodel. De belangrijke spectrale gebieden kunnen ook worden bepaald door het gebruik van geavanceerde algoritmes voor golflengteselectie in combinatie met PLS. Ook het toepassen van alternatieve spectrale voorbewerkingen kan inzicht geven in de model optimalisatie strategie.

Opname van de spectra bij verschillende temperaturen leidde tot kleine verbeteringen voor enkele modellen. Een meer gedetailleerde studie is nodig om de significantie van de waargenomen effecten te bepalen. Hetzelfde geldt voor de niet-lineaire trends die optraden in de spectra van de fysische geprepareerde aardoliemengsels. Tot nu toe werden de gesimuleerde spectra geconstrueerd door lineaire additie van de ruwe aardoliespectra. Toevoeging van een extra voorbewerking, welke corrigeert voor deze niet-lineariteit, zou de fout in de voorspelling kunnen verminderen.

Het belangrijkste onderwerp ten slotte, is de vergroting van de robuustheid van de modellen. Dit maakt de bruikbaarheid onafhankelijk van de spectrometer configuratie en dus ook van de olie-productie en -raffinage locatie. Het toepassen van golflengteselectie zou een deel van de oplossing kunnen zijn. Een andere mogelijkheid is om een calibratietransfer protocol te ontwikkelen. Een dergelijk protocol zou het beschikbare model zodanig aan moeten passen dat het spectra kan verwerken die afkomstig zijn van spectrometers met een andere configuratie. Als alternatief kan een protocol ontwikkeld worden dat de spectra van de andere configuratie zodanig aanpast dat het beschikbare model direkt gebruikt kan worden. Onderzoek gericht op de ontwikkeling van een calibratietransfer protocol voor ieder spectrometer type dat gebruikt gaat worden, wordt noodzakelijk geacht.

List of Abbreviations

AFT	Atmospheric equivalent Flash Temperature
Asph	Asphaltenes content
ASTM	American Society for Testing and Materials International
ATR	Attenuated total reflection
CCD	Charge-coupled device
CR	Carbon residue
CV	Cross validation
D_{LR}	Density (long residue)
D_{SR}	Density (short residue)
DTGS	Deuterated triglycine sulfate
FT	Fourier-transform
GC	Gas chromatography
HDS	Hydrodesulfurization
IP	Institute for petroleum
IR	Infrared
LOO	Leave one out
LR	Long residue
LV	Latent variable
MC	Mean centering
MLR	Multiple linear regression
MSC	Multiplicative signal correction
NIR	Near infrared
NMR	Nuclear magnetic resonance
P	Penetration
PARAFAC	Parallel Factor(s)
PC	Principal component
PCA	Principal component analysis
PLS	Partial least squares
PP	Pour point
R&B	Ring and ball (or softening point)
RMSEC	Root mean square error of calibration

RMSECV	Root mean square error of cross validation
RMSEP	Root mean square error of prediction
S	Sulfur content
SG	Savitzky-Golay
SNV	Standard normal variate
SR	Short residue
STDEV	Standard deviation
UV-Vis	Ultraviolet-visible
V_{LR}	Viscosity (long residue)
V_{SR}	Viscosity (short residue)
YLC	Yield long on crude
YSC	Yield short on crude

List of Related Publications and Patents

Publications

P. de Peinder, T. Visser, D.D. Petrauskas, F. Salvatori, F. Singelenberg, F. Soulimani, B.M. Weckhuysen, *Prediction of Long and Short Residue Properties of Crude Oils from Their Infrared and Near-Infrared Spectra*, Applied Spectroscopy 62 (2008) 414.

P. de Peinder, T. Visser, D. D. Petrauskas, F. Salvatori, F. Soulimani, B. M. Weckhuysen, *Prediction of Long Residue Properties of Potential Blends from Mathematically Mixed Infrared Spectra of Pure Crudes by Partial Least Squares Regression Models*, Energy & Fuels 23 (2009) 2164.

P. de Peinder, T. Visser, D. D. Petrauskas, F. Salvatori, F. Soulimani, B. M. Weckhuysen, *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 (2009) 205.

P. de Peinder, T. Visser, D. D. Petrauskas, F. Salvatori, F. Soulimani, B. M. Weckhuysen, *Sulfur Speciation of Crude Oils by Partial Least Squares Regression of Their Infrared Spectra,* Energy & Fuels, accepted for publication.

Patent Applications

P. de Peinder, F. Singelenberg, T. Visser and B. Weckhuysen, *Method for Predicting a Physical Property of a Residue Obtainable from a Crude Oil*, (2008) WO/2008/135411.

Dankwoord

Deze bladzijden zijn misschien wel de belangrijkste van het proefschrift. Als het goed is zal niemand verrast worden op deze pagina's omdat ik de meeste mensen al persoonlijk bedankt heb gedurende de afgelopen jaren. Door het vast te leggen op papier (en internet) wordt het voor een groter publiek duidelijk wie een onmisbare bijdrage hebben geleverd.

Om bij mijn begin te beginnen wil ik mijn liefdevolle ouders Wout en Henny bedanken. Vanaf dat ik me kan herinneren, en eerder, zijn ze er voor mij geweest en hebben ze mij de ruimte gegeven om me in mijn eigen tempo te ontplooien. Stimulerend maar nooit dwingend. Wat mij betreft was die gelukkige jeugd de basis voor alles wat ik bereikt heb en waar ik tot nu toe van heb mogen genieten.
Bert Weckhuysen, mijn promotor, wil ik bedanken voor het risico dat hij aangedurfd heeft om mij het beschreven promotie werk toe te vertrouwen. Door zijn snelle respons op alle ingeleverde teksten en het stellen van deadlines (voor mij in deze zeker noodzakelijk) is dit proefschrift er gekomen.
Tom Visser, mijn co-promotor en meer vriend dan collega. Door zijn vertrouwen in mijn spectroscopische en chemometrische kennis ben ik betrokken geraakt bij het beschreven Bitumen project. De samenwerking was wat mij betreft geweldig en ik hoop dat ondanks zijn nakende pensioen we op welke manier dan ook, blijven samenwerken. Desnoods in zijn tuin. De afgelopen jaren was hij mijn kamergenoot, reisgenoot (in binnen- en buitenland), spectroscopiegenoot en Betuwegenoot. Dus veel genoten. Ik had het zonder Tom niet gehaald, zelfs niet gewild. "Het staat nu geschreven, dus is het waar!"
Joop van der Maas, die al tijdens het 2^{e} jaar van mijn studie scheikunde als opvallend goede docent invloed kreeg op mijn keuzes. Hij was het die mijn interesse in de vibratiespectroscopie gewekt heeft en van wie ik nog steeds veel leer. Hij heeft al vanaf het begin geprobeerd mij warm te maken voor een promotie. Dit is hij blijven doen, ook later tijdens de reisjes van de vibratievrienden naar oorden die voldeden aan de criteria zonnig (niet altijd), goed eten en drinken. Zowel in Lissabon, Malta, Porto, Lyon, Londen, Marseille en tijdens de etentjes bij de Griek in Utrecht werd grote druk uitgeoefend op de enige Drs. in het gezelschap.
Fouad Soulimani, die onmisbaar was bij het uitvoeren van veel metingen aan de vaak lastige olie. Goede metingen blijven de basis voor goede resultaten, hoeveel er daarna ook nog aan gerekend wordt. Zijn behulpzame, vriendelijke en sociale opstelling is de afgelopen jaren onmisbaar geweest.
Derek Petrauskas, Fred Singelenberg and Fabien Salvatori from Shell Global Solutions B.V. are thanked for their input and guidance during the Bitumen project.
De studenten Christophe Rebreyend en Bob van de Giessen bedank ik voor het prima uitvoeren van het experimentele werk voor hoofdstuk 6. Monique Lamers en vooral Dymph Serrée van het secretariaat voor alles wat ze voor me gedaan hebben waaronder het op het juiste moment uitvoeren van administratieve handelingen. Vincent Koot voor de ondersteuning bij netwerk problemen en andere computer gerelateerde zaken. The 'colleagues' from the Inorganic Chemistry and Catalysis

group but especially the ones I have worked with: Leticia Espinosa-Alonso, Muriel Lepage, Ana Iglesias Juez, Marianne Kox, Gerbrand Mesu and Jaap Bergwerff.

Aangezien ik al ver voor deze promotie in de vibratiespectroscopie werkzaam was zijn er nog enkele mensen die ik wil bedanken voor hun bijdrage in een eerdere fase.

H.J. Luinge, die mij als collega in de vibratiespectroscopie groep de basis van de chemometrie heeft bijgebracht. Ik hoop dat het hem goed blijft gaan in Florida waar hij de beste spectroscopist en chemometrist is binnen zijn huidige vakgebied.

Bert Lutz, die ik helaas niet meer persoonlijk kan bedanken. Hij heeft meer invloed op mijn leven en werk gehad dan hij ooit geweten heeft. Naast een geweldige collega en spectroscopist was hij een fantastisch mens.

Jan Jansen, mijn baas ("nee, collega"), omdat hij het mij mogelijk heeft gemaakt voor de Universiteit Utrecht te blijven werken tijdens mijn dienstverband met Philips en goedkeurde dat ik een bedrijf oprichtte in hetzelfde vakgebied. Dit heeft uiteindelijk de deur geopend voor dit promotie onderzoek. In deze gaat ook mijn dank uit naar de steun van Arjan Mank, de opvolger van Jan, en sinds dit jaar ook weer ‘collega' bij de Universiteit Utrecht.

RKTVC 3 voor de sportieve ontspanning. Onoverwinnelijk....., in ieder geval als het om feest maken gaat.

Mijn zussen voor 'het zus zijn' en het uiteindelijk niet volledig verstoren van de randvoorwaarden die heel belangrijk zijn in ieders leven, zeker tijdens zo'n drukke periode. Laten we de sibling etentjes in ere houden.

Mijn drie geweldige zonen, Luuk, Mick en Stan (**L**arge, **M**edium en **S**mall) voor hun pesterijen en grappen die gewoon doorgingen tijdens het schrijven ("Zóóó saaaaiii") van dit boekje. Het positieve voor LMS is de vrije dag van school én het feest. Binnenkort weer MOH met jullie, online, 3 tegen 1, kom maar op!

Als laatste, maar zeker niet de minste, bedank ik mijn geliefde Henriëtte Sloot (Jet). Doordat zij zo lief, sterk en stabiel is kan ik de meest wilde plannen uitvoeren, zelfs een promotie, zonder dat ons gezinsleven of onze relatie er onder gebukt gaat. Ze is mijn desk Jet, privé Jet, (meestal) easy Jet en super Jet. Dit boekje staat bol van de modellen, maar het mooiste, beste en meest betrouwbare model is wat mij betreft pas in deze laatste alinea beschreven...

Peter

Curriculum Vitae

Peter de Peinder werd geboren op 29 augustus 1967 te Tiel. In 1987 behaalde hij het VWO diploma aan de Rijksscholengemeenschap te Tiel. De militaire dienstplicht werd vervuld van 1987 tot 1988 waarna hij Scheikunde ging studeren aan de Universiteit Utrecht. In 1994 studeerde hij af bij de sectie Biochemie van Membranen onder leiding van Prof. dr. B. de Kruijff. In datzelfde jaar ging hij werken bij de vakgroep Analytische Molecuul Spectrometrie in de werkgroep Vibratiespectroscopie onder leiding van Prof. dr. J. H. van der Maas aan de Universiteit Utrecht. In 1999 trad hij in dienst bij Philips Electronics B.V. waar hij senior onderzoeker infrarood en Raman spectroscopie werd in de afdeling Materiaal Analyse bij Philips Research te Eindhoven. Van 1999 tot 2004 bleef hij vanuit Philips in deeltijd betrokken bij de cursussen en het onderzoek op het gebied van vibratiespectroscopie aan de Universiteit Utrecht. In 2004 richtte hij in samenwerking met van der Maas het bedrijf VibSpec op en in juli 2006 verliet hij Philips Research. In 2005 startte hij zijn promotieonderzoek onder leiding van Prof. dr. ir. B. M. Weckhuysen en dr. T. Visser bij de sectie Anorganische Chemie en Katalyse aan de Universiteit Utrecht. In 2008 heeft hij VibSpec Software B.V. opgericht.