



**Sandra Lopes  
da Silva**

**RMN na Caracterização de Correntes Processuais  
de Resíduos Pesados**

**NMR in the Characterization of Heavy Residual  
Procedural Streams**





**Sandra Lopes  
da Silva**

## **RMN na Caracterização de Correntes Processuais de Resíduos Pesados**

### **NMR in the Characterization of Heavy Residual Procedural Streams**

Tese apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Engenharia da Refinação, Petroquímica e Química, realizada sob a orientação científica do Doutor Artur Manuel Soares da Silva, Professor Catedrático do Departamento de Química da Universidade de Aveiro e do Doutor Fernando Gomes Martins, Professor Auxiliar do Departamento de Engenharia Química da Faculdade de Engenharia da Universidade do Porto.

Apoio financeiro da Galp Energia S:A. Apoio financeiro da FCT e do FSE no âmbito do III Quadro Comunitário de Apoio.



UNIÃO EUROPEIA  
Fundo Social Europeu



## **o júri**

Presidente

**Professor António Carlos Matias Correia**  
Professor Catedrático da Universidade de Aveiro

Vogais

**Professor Sebastião José Cabral Foyo de Azevedo**  
Professor Catedrático da Faculdade de Engenharia da Universidade do Porto

**Professor Artur Manuel Soares da Silva**  
Professor Catedrático da Universidade de Aveiro (Orientador)

**Prof. Doutor Marco Paulo Seabra dos Reis**  
Professor Auxiliar da Faculdade de Ciências e Tecnologia da Universidade de Coimbra

**Prof<sup>a</sup>. Doutora Margarida Maria Silva Monteiro Bastos**  
Professora Auxiliar da Faculdade de Engenharia da Universidade do Porto

**Prof. Doutor Fernando Gomes Martins**  
Professor Auxiliar da Faculdade de Engenharia da Universidade do Porto (Coorientador)

**Mestre Jorge Correia Ribeiro**  
Chefe do Laboratório da Refinaria de Matosinhos – Galp Energia S.A.



## Agradecimentos

Gostaria de expressar os meus sinceros agradecimentos ao Professor Doutor Artur Silva, orientador deste trabalho, pela sua orientação, pela sua disponibilidade, paciência, pelo acompanhamento e por todo o apoio e amizade prestados ao longo deste Doutoramento.

Ao Doutor Fernando Martins, coorientador deste trabalho, pela preciosa orientação científica, pelo apoio incondicional, pela sua enorme disponibilidade, pela dedicação, sabedoria e amizade demonstrada ao longo deste trabalho.

Ao Dr. Jorge Ribeiro, coordenador empresarial por parte da Galp Energia, pela orientação, pelo fundamental acompanhamento, por toda a ajuda e esclarecimentos prestados, pela importantíssima contribuição para o desenvolver deste trabalho, pela sua amizade e por toda a sua disponibilidade.

Ao Doutor Carlos Manuel Silva e ao Doutor Francisco Avelino da Silva, pela oportunidade concedida, pelo acompanhamento científico, pelo apoio demonstrado, pela colaboração e amizade.

Ao Doutor Hilário Tavares, pela ajuda, disponibilidade e contributo prestado na obtenção dos espectros de RMN.

A todos os meus colegas do laboratório da refinaria de Matosinhos, Jorge Monteiro, Luís Cunha, Ricardo Albuquerque, Maria João Pinto, Cristina Busto, Sara Coelho, Álvaro Magalhães, Zé Luís, Paulo Santos, Pedro Vale, Sr. Graça, Sr. Albino, Andrea Garcia, Felisbela Tinoco, e em especial à Márcia Silva, pelo bom ambiente existente, por estarem sempre disponíveis para me ajudar e pela amizade. A todos os chefes de turno, ao Sr. Alberto de Castro e à Rita Valença por toda a partilha de conhecimentos, colaboração, apoio e amizade. A todas as pessoas da refinaria de Matosinhos que, direta ou indiretamente, contribuíram para a realização deste trabalho. Um agradecimento especial à Márcia Gonçalves, à Ana Rita Marques, à Rita Pinto e à Patrícia Alves.

A todos os meus colegas do grupo EGICHEM da Universidade de Aveiro, Simão, Marcelo, Zé Pedro, Rui, Patrícia, Ana, Joana, e em especial à Raquel, Bruno Figueiredo, Bruno Antunes e Eduardo, agradeço-lhes pela amizade, ajuda, companheirismo e boa disposição. Aos meus colegas do Departamento de Química e dos Laboratórios Tecnológicos da Universidade de Aveiro pela ajuda, simpatia e boa disposição.

Aos meus colegas/amigos do EngIQ, em especial à Ana Rute Ferreira e Catarina Varanda, por todos os momentos que passámos, por todas as experiências que partilhámos, por toda a motivação, apoio incondicional, pela amizade e companheirismo.

Aos meus amigos, pela compreensão, pela paciência e pela ternura manifestada durante este período de tempo. À Carina, Susana Marques, Joana Garcia, Joana Coimbra, Vitor Ribeiro, Ângela, Ana Afonso, Ana Filipa, Belmifer Maio, Zé, Paulo Geraldo e Catarina Novo pela amizade incondicional.

À Fundação para a Ciência e a Tecnologia e à Galp Energia agradeço a bolsa de Doutoramento (SFRH/BDE/33879/2009).

Finalmente, agradeço especialmente aos meus pais, às minhas irmãs, à minha avó e a toda a minha família por estarem sempre do meu lado, por me apoiarem incondicionalmente, pelo incentivo, paciência e compreensão durante este período de tempo.

A todos, os meus sinceros agradecimentos.





## Palavras-chave

Derivados de petróleo, propriedades físico-químicas, espectroscopia de ressonância magnética nuclear, métodos estatísticos multivariáveis, análise por componentes principais, regressão múltipla por componentes principais, regressão por mínimos quadrados parciais, redes neuronais artificiais, análise estatística.

## Resumo

O principal objetivo deste trabalho foi monitorizar um conjunto de propriedades físico-químicas de correntes processuais pesadas através da espectroscopia de ressonância magnética nuclear, com o intuito de propor um procedimento de análise e processamento de dados em linha para o controlo processual. Vários métodos estatísticos que permitiram relacionar os resultados obtidos por espectroscopia de ressonância magnética nuclear com os resultados obtidos pelos métodos convencionais, aquando da caracterização das diferentes correntes, foram implementados a fim de desenvolver modelos de previsão dessas mesmas propriedades. O conhecimento em tempo real das propriedades físico-químicas dos derivados de petróleo é essencial para otimizar as operações de refinação, garantindo assim operações técnica, económica e ambientalmente adequadas.

A primeira parte deste trabalho envolveu a realização de um elevado número de experiências, efetuadas na refinaria de Matosinhos, seguindo métodos convencionais normalizados, importantes para avaliar e caracterizar as correntes de gasóleo de vácuo leve, gasóleo de vácuo pesado e fuel óleo. As propriedades analisadas foram a massa volúmica, viscosidade cinemática, teor em enxofre, ponto de inflamação, resíduo carbonoso, valor P e a destilação atmosférica e de vácuo. Para além da determinação de todas estas propriedades físico-químicas, usando os métodos convencionais, as mesmas amostras foram analisadas por espectroscopia de ressonância magnética nuclear.

A segunda parte deste trabalho esteve relacionada com a aplicação de métodos estatísticos multivariáveis que relacionam as propriedades físico-químicas com a informação quantitativa adquirida por espectroscopia de ressonância magnética nuclear. Vários métodos foram aplicados, incluindo a análise por componentes principais, a regressão múltipla por componentes principais, as regressões múltiplas parciais e as redes neuronais artificiais. A análise de componentes principais foi utilizada para reduzir o número de variáveis preditivas e transformá-las em novas variáveis, os componentes principais. Estes componentes principais foram utilizados como variáveis de entrada da regressão múltipla por componentes principais e das redes neuronais artificiais. Na regressão por mínimos quadrados parciais os dados originais foram usados como variáveis de entrada. Tomando em consideração o desempenho dos modelos desenvolvidos, com recurso a parâmetros estatísticos, foi possível concluir que a regressão múltipla por componentes principais contribuiu para piores desempenhos. Melhores desempenhos foram obtidos com a aplicação da regressão por mínimos quadrados parciais e das redes neuronais artificiais. No entanto, foi com os modelos de redes neuronais artificiais que melhores desempenhos foram obtidos em quase todas as propriedades analisadas.

Tendo em conta os resultados obtidos, foi possível concluir que a espectroscopia de ressonância magnética nuclear combinada com métodos estatísticos multivariáveis pode ser usada para prever as propriedades físico-químicas de derivados de petróleo. Demonstrou-se que esta técnica pode ser considerada como uma potencial alternativa aos métodos convencionais tendo-se obtido resultados bastante promissores.



**Keywords**

Crude oil fractions, physical-chemical properties, nuclear magnetic resonance spectroscopy, multivariate statistical methods, principal component analysis, principal component regression, partial least squares, artificial neural networks, statistical performance indexes.

**Abstract**

The main objective of this work was to monitor a set of physical-chemical properties of heavy oil procedural streams through nuclear magnetic resonance spectroscopy, in order to propose an analysis procedure and online data processing for process control. Different statistical methods which allow to relate the results obtained by nuclear magnetic resonance spectroscopy with the results obtained by the conventional standard methods during the characterization of the different streams, have been implemented in order to develop models for predicting these same properties. The real-time knowledge of these physical-chemical properties of petroleum fractions is very important for enhancing refinery operations, ensuring technically, economically and environmentally proper refinery operations.

The first part of this work involved the determination of many physical-chemical properties, at Matosinhos refinery, by following some standard methods important to evaluate and characterize light vacuum gas oil, heavy vacuum gas oil and fuel oil fractions. Kinematic viscosity, density, sulfur content, flash point, carbon residue, P-value and atmospheric and vacuum distillations were the properties analysed. Besides the analysis by using the standard methods, the same samples were analysed by nuclear magnetic resonance spectroscopy.

The second part of this work was related to the application of multivariate statistical methods, which correlate the physical-chemical properties with the quantitative information acquired by nuclear magnetic resonance spectroscopy. Several methods were applied, including principal component analysis, principal component regression, partial least squares and artificial neural networks. Principal component analysis was used to reduce the number of predictive variables and to transform them into new variables, the principal components. These principal components were used as inputs of the principal component regression and artificial neural networks models. For the partial least squares model, the original data was used as input. Taking into account the performance of the develop models, by analysing selected statistical performance indexes, it was possible to conclude that principal component regression lead to worse performances. When applying the partial least squares and artificial neural networks models better results were achieved. However, it was with the artificial neural networks model that better predictions were obtained for almost of the properties analysed.

With reference to the results obtained, it was possible to conclude that nuclear magnetic resonance spectroscopy combined with multivariate statistical methods can be used to predict physical-chemical properties of petroleum fractions. It has been shown that this technique can be considered a potential alternative to the conventional standard methods having obtained very promising results.



# Table of Contents

<b>Table of Contents</b> .....	<b>i</b>
<b>Figure Index</b> .....	<b>v</b>
<b>Table Index</b> .....	<b>xi</b>
<b>List of Abbreviations</b> .....	<b>xv</b>
<b>Notation</b> .....	<b>xxi</b>
<b>1. INTRODUCTION</b> .....	<b>1</b>
1.1. Scientific Relevance.....	3
1.2. Thesis Structure.....	5
<b>2. PETROLEUM INDUSTRY</b> .....	<b>7</b>
2.1. Crude Oil.....	9
2.2. Crude Oil Refining.....	14
2.2.1. Crude Oil Refining in Portugal.....	21
2.2.2. Matosinhos Refinery.....	22
2.2.3. Fuel Plant.....	25
2.3. Upgrading the “Bottom of Barrel”.....	29
2.3.1. Vacuum Distillation.....	29
2.3.2. Visbreaker Unit.....	30
2.4. Crude Oil Derivatives.....	34
2.4.1. Vacuum Gas Oil (LVGO and HVGO).....	35
2.4.2. Fuel Oil.....	36
2.4.3. Physical-Chemical Characterization of Petroleum Fractions.....	37
2.5. Final Remarks.....	42
<b>3. CHROMATOGRAPHIC AND SPECTROSCOPIC ANALYSIS OF PETROLEUM FRACTIONS WITH EMPHASIS IN NMR</b> .....	<b>43</b>
3.1. Chromatographic and Spectroscopy Characterization of Petroleum Fractions.....	45
3.1.1. Gas Chromatography (GC).....	45

3.1.2.	High Performance Liquid Chromatography (HPLC) .....	47
3.1.3.	Thin Layer Chromatography (TLC) .....	48
3.1.4.	Infrared (IR) Spectroscopy .....	49
3.1.5.	Raman Spectroscopy.....	51
3.1.6.	Mass Spectrometry (MS) .....	52
3.1.7.	Nuclear Magnetic Resonance (NMR) Spectroscopy .....	54
3.2.	NMR Spectroscopy - Overview .....	56
3.2.1.	<sup>1</sup> H NMR Spectroscopy.....	65
3.2.2.	<sup>13</sup> C NMR Spectroscopy.....	67
3.2.3.	NMR Spectrometer .....	68
3.3.	Applications.....	70
3.3.1.	<sup>1</sup> H NMR Spectroscopy in the Analysis of Petroleum Fractions .....	70
3.3.2.	<sup>13</sup> C NMR Spectroscopy in the Analysis of Petroleum Fractions.....	74
3.3.3.	<sup>1</sup> H and <sup>13</sup> C NMR Spectroscopy in the Analysis of Petroleum Samples .....	82
3.3.4.	NMR Spectroscopy in Asphaltenes Characterization.....	86
3.3.5.	Spectral editing NMR Techniques.....	89
3.3.6.	NMR Relaxation Effects.....	98
3.4.	Final Remarks.....	99
<b>4.</b>	<b>EXPERIMENTAL PROCEDURES .....</b>	<b>101</b>
4.1.	Samples .....	103
4.2.	Physical-Chemical Characterization .....	104
4.2.1.	Kinematic Viscosity (ASTM D445) .....	104
4.2.2.	Density (ASTM D4052 and ASTM D5002).....	109
4.2.3.	Carbon Residue (ASTM D4530) .....	113
4.2.4.	Sulfur Content (IP 336).....	116
4.2.5.	Flash Point (ASTM D93).....	119
4.2.6.	P-Value (SMS 1600).....	121
4.2.7.	Distillation at Atmospheric Pressure (ISO 3405) .....	123

4.2.8.	Distillation at Reduced Pressure (ASTM D1160) .....	124
4.3.	NMR Experiments.....	127
4.3.1.	<sup>1</sup> H NMR Experiments .....	127
4.3.2.	<sup>13</sup> C NMR Experiments .....	129
4.3.3.	Pre-Measuring Steps .....	131
4.3.4.	Data Pre-Processing .....	133
4.3.4.1.	Phase and Baseline Correction .....	134
4.3.4.2.	Chemical Shift Alignment.....	135
4.3.4.3.	Data Reduction .....	137
4.3.4.4.	Binning .....	137
4.3.4.5.	Normalization .....	138
4.3.4.6.	Intensity Scaling .....	138
4.3.4.7.	Creation of Data Set for Multivariate Analysis.....	139
4.4.	Final Remarks.....	140
<b>5.</b>	<b>PREDICTION USING MULTIVARIATE METHODS .....</b>	<b>141</b>
5.1.	State of the Art .....	143
5.2.	Methods Applied in this Thesis.....	146
5.2.1.	Principal Components Analysis - PCA.....	147
5.2.2.	Principal Component Regression - PCR.....	150
5.2.3.	Partial Least Squares – PLS.....	151
5.2.4.	Artificial Neural Networks – ANN.....	157
5.3.	Models Pre-Treatment.....	164
5.4.	Performance Indexes .....	168
5.5.	Final Remarks.....	173
<b>6.</b>	<b>RESULTS AND DISCUSSION.....</b>	<b>175</b>
6.1.	Data .....	177
6.1.1.	<sup>13</sup> C NMR Spectroscopy.....	177
6.1.2.	<sup>1</sup> H NMR Spectroscopy.....	181

6.1.3.	$^1\text{H}$ and $^{13}\text{C}$ NMR Spectroscopy.....	185
6.2.	$^1\text{H}$ NMR Spectroscopy in the Analysis of Petroleum Fractions .....	186
6.2.1.	Fuel Oil .....	187
6.2.1.1.	Fuel Oil - Final Remarks .....	221
6.2.2.	LVGO .....	224
6.2.1.2.	LVGO – Final Remarks.....	243
6.2.3.	HVGO.....	246
6.2.1.3.	HVGO – Final Remarks .....	277
6.3.	LVGO, HVGO and Fuel Oil – Chemical Composition .....	281
<b>7.</b>	<b>CONCLUSIONS AND FUTURE WORK.....</b>	<b>287</b>
7.1.	General Conclusions.....	289
7.2.	Suggestions for Future Work .....	293
<b>8.</b>	<b>REFERENCES .....</b>	<b>295</b>
<b>9.</b>	<b>APPENDIX .....</b>	<b>315</b>



## Figure Index

### Chapter 2

<b>Figure 2.1:</b> Example of a refining flowsheet of a simple refinery.....	16
<b>Figure 2.2:</b> Example of a refining flowsheet of a conventional refinery.....	17
<b>Figure 2.3:</b> Example of a refining flowsheet of petroleum industries.....	18
<b>Figure 2.4:</b> Example of a possible today's refining flowsheet. ....	20
<b>Figure 2.5:</b> Matosinhos refinery - Galp Energia S.A.....	21
<b>Figure 2.6:</b> Flowsheet of Matosinhos refinery. ....	24
<b>Figure 2.7:</b> Flowsheet of the fuels plant of Matosinhos refinery. ....	28
<b>Figure 2.8:</b> Example of a vacuum unit. ....	30
<b>Figure 2.9:</b> Example of a vacuum – visbreaker unit.....	33
<b>Figure 2.10:</b> Flowsheet of the vacuum and visbreaker unit of Matosinhos refinery.....	34

### Chapter 3

<b>Figure 3.1:</b> Example of a GC chromatogram of a light Nigerian crude (fraction with a boiling temperature smaller than 343 °C).....	47
<b>Figure 3.2:</b> Example of TLC-FIDr chromatograms of 500 N VGO cut and its fractions. Chromatogram a) VGO, b) extract, c) raffinate, d) wax and e) base oil. ....	49
<b>Figure 3.3:</b> Example of an IR spectrum of heavy fuel oil. ....	50
<b>Figure 3.4:</b> Example of a Raman spectrum of a gasoline sample. ....	52
<b>Figure 3.5:</b> Example of an ESI FT-ICR mass spectra of condensed thiophenes in different Arabian crude oils.....	54
<b>Figure 3.6:</b> Representation of the possible energy levels for nuclei with spin of $I = 12$ . .	58
<b>Figure 3.7:</b> Precession of a single proton around the external magnetic field. ....	59
<b>Figure 3.8:</b> The nuclear magnetic resonance process. ....	60
<b>Figure 3.9:</b> Formation of the bulk magnetization vector from the excess spin population.	61
<b>Figure 3.10:</b> Representation of the net magnetization vector $M_0$ in the transverse plane after the application of the 90° radiofrequency pulse. ....	62
<b>Figure 3.11:</b> Application of the Fourier transformation on the FID. (A) Time domain data, FID, (B) Frequency domain, spectrum. ....	63

<b>Figure 3.12:</b> $^1\text{H}$ NMR spectrum of a fuel oil. ....	66
<b>Figure 3.13:</b> Inverse gate decoupling $^{13}\text{C}$ NMR spectrum of a crude oil (Albcore Crude Oil).....	67
<b>Figure 3.14:</b> An example of a modern NMR spectrometer. ....	69
<b>Figure 3.15:</b> Typical $^{13}\text{C}$ and $^1\text{H}$ NMR spectra for SARA fractions, obtained in a Varian Unity Inova 400 MHz spectrometer with a 5 mm 4 nucleus probe. ....	82
<b>Figure 3.16:</b> Example of some asphaltene structures generated using Monte Carlo method. ....	87
<b>Figure 3.17:</b> Example of a $^1\text{H}$ DOSY NMR spectrum of a base oil. ....	89
<b>Figure 3.18:</b> Asphaltenes $^1\text{H}$ NMR spectrum at 1 wt% in toluene at 20 °C. ....	89
<b>Figure 3.19:</b> Example of a COSY spectrum of a base oil. Obtained using a 300 MHz NMR spectrometer with a sweep width of 3016 Hz and 4 K data points. ....	91
<b>Figure 3.20:</b> Example of an HETCOR spectrum of a base oil, obtained in a NMR spectrometer operating a 300 MHz for $^1\text{H}$ and 75 MHz for $^{13}\text{C}$ . A = isopropyl methyl groups (0.93 and 22.7 ppm) of S-1 and S-2 structures. A' = -CH group resonance (S-2) at $\delta$ 28.2. B and B' = -CH <sub>2</sub> carbon of S-4 at $\delta$ 20.3 correlated with a methyl group (19 – 21 ppm). E (0.9 and 28.6, 27.4 ppm) and E' (0.99 and 30.2 ppm) = branched methyl of S-10 and S-11. G = $\beta$ – methylene groups (1.3 and 22.4 ppm) of S-2. With S-1, S-2, S-4 and S-10 from Table 3.2. ....	92
<b>Figure 3.21:</b> Example of an a) HSQC and an b) HMBC spectra of a high boiling point fraction VGO, identifying CH <sub>3</sub> (a), CH <sub>2</sub> (b) and CH (c) groups in different side chains - ethyl (a <sub>1</sub> , b <sub>1</sub> ), propyl (a <sub>2</sub> , b <sub>2</sub> b <sub>3</sub> ), butyl (a <sub>3</sub> , b <sub>4</sub> b <sub>5</sub> b <sub>6</sub> ), isopropyl (a <sub>4</sub> , c <sub>1</sub> ), tetraline (b <sub>7</sub> b <sub>8</sub> ), indane (b <sub>9</sub> b <sub>10</sub> ) and 1-methylindane (a <sub>5</sub> , b <sub>11</sub> b <sub>12</sub> c <sub>2</sub> ). Obtained in a Bruker Avance 400 MHz NMR spectrometer equipped with an inverse detecting probe, using a $\pi/2$ pulses of 13.3 and 9.6 $\mu\text{s}$ for $^1\text{H}$ and $^{13}\text{C}$ , respectively, a recycle delay of 2 s and three sine gradients of 1.5 ms with 8:3:2 ratio for HSQC and 5:3:4 ratio for HMBC. ....	93

## Chapter 4

<b>Figure 4.1:</b> Example of a Cannon-Fenske viscometer for transparent liquids. ....	105
<b>Figure 4.2:</b> Viscosity Tamson bath with a viscometer for transparent liquids.....	106
<b>Figure 4.3:</b> Example of a Cannon-Fenske viscometer for opaque liquids. ....	107

<b>Figure 4.4:</b> Histogram for 168 samples on the variation of HVGO kinematic viscosity at 100 °C. ....	108
<b>Figure 4.5:</b> Histogram for 217 samples on the variation of fuel oil kinematic viscosity at 100 °C. ....	108
<b>Figure 4.6:</b> Histogram for 108 samples on the variation of fuel oil kinematic viscosity at 50 °C. ....	109
<b>Figure 4.7:</b> Digital density analyzer, Anton Paar DMA 4500M. ....	110
<b>Figure 4.8:</b> Histogram for 105 samples on the variation of LVGO density. ....	111
<b>Figure 4.9:</b> Digital density analyzer used, Mettler Toledo Density Meter De40. ....	112
<b>Figure 4.10:</b> Histogram for 189 samples on the variation of HVGO density.....	112
<b>Figure 4.11:</b> Histogram for 185 samples on the variation of fuel oil density.....	113
<b>Figure 4.12:</b> Coking oven, Alcor MCRT-160. ....	115
<b>Figure 4.13:</b> Histogram for 142 samples on the variation of HVGO carbon residue.....	115
<b>Figure 4.14:</b> Histogram for 103 samples on the variation of fuel oil carbon residue.....	116
<b>Figure 4.15:</b> Energy dispersive X-ray fluorescence analyzer, Oxford Instruments, Lab-X3500SCL. ....	117
<b>Figure 4.16:</b> Fluorescence of sulfur atom.....	118
<b>Figure 4.17:</b> Histogram for 213 samples on the variation of fuel oil sulfur content. ....	118
<b>Figure 4.18:</b> Pensky-Martens closed flash tester, HFP 360 Flash Point Analyzers. ....	120
<b>Figure 4.19:</b> Histogram for 157 samples on the variation of fuel oil flash point. ....	120
<b>Figure 4.20:</b> Zematra ASA automated stability analyzer for residual refinery streams and fuel oils. ....	122
<b>Figure 4.21:</b> Histogram for 213 samples on the variation of fuel oil P-value.....	122
<b>Figure 4.22:</b> Herzog atmospheric distillation HAD 627 apparatus. ....	123
<b>Figure 4.23:</b> Herzog vacuum distillation HDV 632 apparatus. ....	124
<b>Figure 4.24:</b> Bruker Avance 300 NMR spectrometer. ....	127
<b>Figure 4.25:</b> Example of a LVGO <sup>1</sup> H NMR spectrum. ....	128
<b>Figure 4.26:</b> Example of a HVGO <sup>1</sup> H NMR spectrum.....	128
<b>Figure 4.27:</b> Example of a fuel oil <sup>1</sup> H NMR spectrum.....	129
<b>Figure 4.28:</b> Example of a fuel oil inverse gate decoupling <sup>13</sup> C NMR spectrum.....	130
<b>Figure 4.29:</b> Example of a flow chart of the different steps that can be utilized during the analysis of NMR spectral data.....	133

<b>Figure 4.30:</b> Peak signal in pure absorption or dispersion mode. ....	134
<b>Figure 4.31:</b> <sup>1</sup> H NMR spectrum of a fuel oil with reference of the TMS signal. ....	135
<b>Figure 4.32:</b> Representation of the scaling methods used: after mean-centering and autoscaling all variables will have mean zero and unit variance. ....	139
<b>Figure 4.33:</b> Creation of data set for multivariate analysis after applying data pre-processing to each spectrum. ....	139

## Chapter 5

<b>Figure 5.1:</b> A flow chart of the different multivariate statistical methods used in this work. ....	147
<b>Figure 5.2:</b> Decomposition of the data matrix into scores, loadings and the residual (E). ....	148
<b>Figure 5.3:</b> A flow chart of the complete procedure used when applying the PLS model. ....	154
<b>Figure 5.4:</b> Example of a feedforward artificial neural network. ....	159
<b>Figure 5.5:</b> Steps used in the development of an ANN prediction models. ....	162
<b>Figure 5.6:</b> Schematic representation of variable selection for the PCR model and PLS model with cross validation. ....	165
<b>Figure 5.7:</b> Schematic representation of variable selection for the and ANN models. ....	165

## Chapter 6

<b>Figure 6.1:</b> Quantitative <sup>13</sup> C NMR spectra of some fuel oil samples. ....	178
<b>Figure 6.2:</b> <sup>1</sup> H NMR spectra of some fuel oil samples. ....	182
<b>Figure 6.3:</b> RMSECV plot for predicting the number of latent variables. ....	193
<b>Figure 6.4:</b> Reproducibility and difference between predicted and observed sulfur content when using PLS model. ....	195
<b>Figure 6.5:</b> Prediction of sulfur content for the a) training and validation data sets and b) test data set, when using the three multivariate models. ....	196
<b>Figure 6.6:</b> Correlation between the experimental and predicted sulfur results of the fuel oil samples obtained by: a) PCR, b) PLS and c) ANN models. ....	198
<b>Figure 6.7:</b> Correlation between the experimental and predicted kinematic viscosity results of the fuel oil samples obtained at 100 °C by: a) PCR, b) PLS and c) ANN models. ....	202

<b>Figure 6.8:</b> Correlation between the experimental and predicted kinematic viscosity results of the fuel oil samples obtained at 50 °C by: a) PCR, b) PLS and c) ANN models. ....	205
<b>Figure 6.9:</b> Reproducibility and difference between predicted and observed density when using ANN models. ....	208
<b>Figure 6.10:</b> Correlation between the experimental and predicted density results of the fuel oil samples obtained by: a) PCR, b) PLS and c) ANN models. ....	209
<b>Figure 6.11:</b> Reproducibility and difference between predicted and observed carbon residue when using ANN models. ....	211
<b>Figure 6.12:</b> Correlation between the experimental and predicted carbon residue results of the fuel oil samples obtained by: a) PCR, b) PLS and c) ANN models. ....	213
<b>Figure 6.13:</b> Correlation between the experimental and predicted flash point results of the fuel oil samples obtained by: a) PCR, b) PLS and c) ANN models. ....	216
<b>Figure 6.14:</b> Correlation between the experimental and predicted P-value results of the fuel oil samples obtained by: a) PCR, b) PLS and c) ANN models. ....	220
<b>Figure 6.15:</b> Correlation between the experimental and predicted density results of the LVGO samples obtained by: a) PCR, b) PLS and c) ANN models. ....	227
<b>Figure 6.16:</b> Correlation between the experimental and predicted 5% distillation of the LVGO samples obtained by: a) PCR, b) PLS and c) ANN models. ....	230
<b>Figure 6.17:</b> Correlation between the experimental and predicted 10% distillation of the LVGO samples obtained by: a) PCR, b) PLS and c) ANN models. ....	233
<b>Figure 6.18:</b> Correlation between the experimental and predicted 50% distillation of the LVGO samples obtained by: a) PCR, b) PLS and c) ANN models. ....	236
<b>Figure 6.19:</b> Correlation between the experimental and predicted 90% distillation of the LVGO samples obtained by: a) PCR, b) PLS and c) ANN models. ....	239
<b>Figure 6.20:</b> Correlation between the experimental and predicted 95% distillation of the LVGO samples obtained by: a) PCR, b) PLS and c) ANN models. ....	242
<b>Figure 6.21:</b> Correlation between the experimental and predicted density of the HVGO samples obtained by: a) PCR, b) PLS and c) ANN models. ....	249
<b>Figure 6.22:</b> Correlation between the experimental and predicted kinematic viscosity at 100 °C of the HVGO samples obtained by: a) PCR, b) PLS and c) ANN models. ....	252
<b>Figure 6.23:</b> Correlation between the experimental and predicted carbon residue of the HVGO samples obtained by: a) PCR, b) PLS and c) ANN models. ....	255

<b>Figure 6.24:</b> Correlation between the experimental and predicted 5% distillation cut of the HVGO samples obtained by: a) PCR, b) PLS and c) ANN models.....	259
<b>Figure 6.25:</b> Correlation between the experimental and predicted 10% distillation cut of the HVGO samples obtained by: a) PCR, b) PLS and c) ANN models.....	263
<b>Figure 6.26:</b> Correlation between the experimental and predicted 50% distillation cut of the HVGO samples obtained by: a) PCR, b) PLS and c) ANN models.....	267
<b>Figure 6.27:</b> Correlation between the experimental and predicted 90% distillation cut of the HVGO samples obtained by: a) PCR, b) PLS and c) ANN models.....	270
<b>Figure 6.28:</b> Correlation between the experimental and predicted 95% distillation cut of the HVGO samples obtained by: a) PCR, b) PLS and c) ANN models.....	273
<b>Figure 6.29:</b> Comparison between some HVGO samples presenting the same distillation temperature for the 95% distillation cut (the spectrum of each HVGO sample was differentiate with different colours).....	275
<b>Figure 6.30:</b> Expansion of the aliphatic region of the <sup>1</sup> H NMR spectra of Figure 6.29 obtained in the 95% distillation cut (the spectrum of each HVGO sample was differentiate with different colours). .....	276
<b>Figure 6.31:</b> Expansion of the aromatic region of the <sup>1</sup> H NMR spectra of Figure 6.29 obtained in the 95% distillation cut (the spectrum of each HVGO sample was differentiate with different colours). .....	276

## Table Index

### Chapter 2

<b>Table 2.1:</b> Composition of crude oil.....	12
<b>Table 2.2:</b> Basic operations in petroleum processing.....	15
<b>Table 2.3:</b> ASTM methods and other tests used in the characterization of petroleum fractions.....	38

### Chapter 3

<b>Table 3.1:</b> Chemical shift regions of various types of protons (solvent: deuteriochloroform).....	73
<b>Table 3.2:</b> Some branched structures identified in base oils and their characteristic $^{13}\text{C}$ NMR chemical shifts. P=-CH <sub>3</sub> , S=-CH <sub>2</sub> and *=carbons for which chemical shifts are reported.....	80
<b>Table 3.3:</b> Assignment of $^{13}\text{C}$ NMR chemical shifts to different types of carbon atoms (solvent: deuteriochloroform).....	81
<b>Table 3.4:</b> Spectral editing NMR techniques.....	96

### Chapter 4

<b>Table 4.1:</b> Identification of samples for properties that were analysed.....	104
<b>Table 4.2:</b> Number of samples, test time and minimum and maximum results for properties that were analysed.....	126

### Chapter 5

<b>Table 5.1:</b> The statistical parameters analysed.....	171
<b>Table 5.2:</b> General performance ratings.....	173

### Chapter 6

<b>Table 6.1:</b> Number of samples analysed by $^1\text{H}$ and $^{13}\text{C}$ NMR spectroscopy and minimum and maximum values of the properties analysed by the standard methods.....	177
--	-----

<b>Table 6.2:</b> Results obtained for different properties of fuel oil analysed when using quantitative <sup>13</sup> C NMR in multivariate models. ....	179
<b>Table 6.3:</b> Results obtained for different properties of fuel oil analysed when using <sup>1</sup> H NMR in multivariate models. ....	183
<b>Table 6.4:</b> Number of samples analysed by <sup>1</sup> H NMR spectroscopy and minimum and maximum values of the properties analysed by the standard methods. ....	187
<b>Table 6.5:</b> Performance indexes achieved using PCR, PLS and ANN models when applying different bins (bin = 0.02, 0.03 and 0.04). ....	189
<b>Table 6.6:</b> Performance indexes achieved when using PCR model during training, validation and test data sets, to predict the sulfur content of fuel oil samples. ....	190
<b>Table 6.7:</b> Performance indexes achieved when changing the number of nodes at the hidden layer in the ANN model.....	191
<b>Table 6.8:</b> Performance indexes achieved when changing the number of LVs in the PLS model. ....	193
<b>Table 6.9:</b> Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the viscosity at 100 °C of fuel oil samples. ....	200
<b>Table 6.10:</b> Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the viscosity at 50 °C of fuel oil samples. ....	203
<b>Table 6.11:</b> Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the density of fuel oil samples. ....	207
<b>Table 6.12:</b> Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the carbon residue of fuel oil samples...	211
<b>Table 6.13:</b> Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the flash point of fuel oil samples. ....	215
<b>Table 6.14:</b> Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the P-value of fuel oil samples.....	218
<b>Table 6.15:</b> Results obtained for different properties of fuel oil analysed when using <sup>1</sup> H NMR in multivariate models. ....	222
<b>Table 6.16:</b> Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the density of LVGO samples.....	225



<b>Table 6.17:</b> Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the 5% distillation of LVGO samples...	229
<b>Table 6.18:</b> Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the 10% distillation of LVGO samples.	231
<b>Table 6.19:</b> Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the 50% distillation of LVGO samples.	234
<b>Table 6.20:</b> Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the 90% distillation of LVGO samples.	237
<b>Table 6.21:</b> Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the 95% distillation of LVGO samples.	240
<b>Table 6.22:</b> Results obtained for different properties of LVGO analysed when using <sup>1</sup> H NMR in multivariate models. ....	244
<b>Table 6.23:</b> Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the density of HVGO samples. ....	248
<b>Table 6.24:</b> Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the kinematic viscosity at 100 °C of HVGO samples. ....	250
<b>Table 6.25:</b> Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the carbon residue of HVGO samples. .	253
<b>Table 6.26:</b> Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the 5% distillation cut of HVGO samples. ....	257
<b>Table 6.27:</b> Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the 10% distillation cut of HVGO samples. ....	260
<b>Table 6.28:</b> Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the 50% distillation cut of HVGO samples. ....	264
<b>Table 6.29:</b> Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the 90% distillation cut of HVGO samples. ....	269

<b>Table 6.30:</b> Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the 95% distillation cut of HVGO samples. ....	272
<b>Table 6.31:</b> Results obtained for different properties of HVGO analysed when using $^1\text{H}$ NMR in multivariate models. ....	279
<b>Table 6.32:</b> Comparison between LVGO, HVGO and fuel oil by using $^1\text{H}$ NMR spectroscopy. ....	282
<b>Table 6.33:</b> Comparison between LVGO, HVGO and fuel oil by using $^{13}\text{C}$ NMR spectroscopy. ....	283

## List of Abbreviations

2D COSY	Homonuclear correlation spectroscopy
2D HETCOR	Heteronuclear correlation spectroscopy
2D TOCSY	Homonuclear total correlation spectroscopy
ADC	Analogue-to-digital converter
ANN	Artificial neural network
API	American petroleum institute
APPI	Atmospheric pressure photoionization technique
ASH	Ash content
Asph	Asphaltenes
ASTM	American society for testing and materials
bbI/day	Barrel per day
CCAI	Calculated carbon aromaticity index
CI	Chemical ionization technique
COW	Correlation optimised warping
CR	Carbon residue
Cr(acac) <sub>3</sub>	Trisacetylacetonatochromium(III)
DEA	Diethanolamine
DEPT	Distortionless enhancement by polarization transfer
D	Density
DOSY	Diffusion-ordered spectroscopy

DTW	Dynamic time warping
EI	Electron impact technique
ESI	Electrospray ionization technique
ETBE	Ethyl-t-butyl ether
F	Fluorescence detector
FAB	Fast atom bombardment technique
FCC	Fluid catalytic cracking
FDI	Field desorption ionization technique
FI	Field ionization technique
FID	Free induction decay
FIDr	Flame ionization detector
FPD	Flame photometric detector
FT	Fourier transformation
FT-ICR-MS	Fourier transform ion cyclotron resonance mass spectrometry
GASPE	Gated spin echo
GC	Gas chromatography
GCV	Gross calorific values
HDS	Hydrodesulfurization
HF	Hydrofinished solvent refined oils
HMBC	Heteronuclear multiple quantum correlation
HPLC	High performance liquid chromatography

HSQC	Heteronuclear single quantum correlation
HT	Hydrotreated base oils
HVGO	Heavy vacuum gas oil
HVI	High viscosity index
<i>I</i>	Spin quantum number
IC	Correlation index
INEPT	Insensitive enhancement by polarization transfer
IP	Institute of petroleum
IR	Infrared spectroscopy
K <sub>UOP</sub> factor	UOP characterization factor
LCO	Light cycle oil
LPG	Liquefied petroleum gas
LVGO	Light vacuum gas oil
LVI	Low viscosity index
LVs	Latent variables
MAE	Mean absolute error
MBE	Mean bias error
MCR	Micro-carbon residue
MLR	Multiple linear regression
MON	Motor octane number
MS	Mass spectrometry

MSC	Multiplicative signal correction
MSE	Mean squared error
MTBE	Methyl-t-butyl ether
MVGO	Medium vacuum gas oil
NCV	Net calorific values
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser enhancement
NSE	Nash-Sutcliffe efficiency
PBIAS	Percent bias
PCA	Principal component analysis
PCR	Principal component regression
PCs	Principal components
PCSE	Part-coupled spin echo
PFA	Principal factor analysis
PFGE	Pulse field gradient spin-echo
PLS	Partial least squares
PP	Pour point
ppm	Parts per million
pump	Pump octane number
QUAT	Quaternary-only carbon spectra
R <sup>2</sup>	Coefficient of determination

RELAY	Heteronuclear relayed coherence transfer
RF	Radio frequency
RI	Refractive index detector
RMSE	Root mean squared error
RMSECV	Root mean squared error of cross validation
RMSEP	Root mean squared error of prediction
RON	Octane number
RSR	Root mean squared error - observations standard deviation ratio
RVP	Reid vapour pressure
S	Sulfur content
SHF	Several hydrofinished oils
TAME	Methyl-t-amyl ether
TCD	Thermal conductivity detector
TLC	Thin layer chromatography
TMS	tetramethylsilane
TSA	Total sediment accelerated
UV	Ultraviolet absorption detector
V	Viscosity
VB	Visbroken
VAST	Variable stability scaling
VHVI	Very high viscosity index

VI	Viscosity index
YLC	Yield long-on-crude



## Notation

$A$	Mass of carbon residue
$A_S$	Percentage substitution of aromatic rings
$AT$	Acquisition time
$A_W$	Molecular weight of the aromatics
$B$	Integral intensity in the region 2.4-3.5 ppm
$\mathbf{B}$	Regression parameters matrix
$\mathbf{b}$	Regression parameters vector
$b - Me$	Percentage of branched methyl groups
$\mathbf{B}_{PLS}$	Regression parameters of partial least squares regression
$BS$	Average number of branching sites per molecule
$C$	Fraction of carbon atoms present as C groups
$\mathbf{c}$	$\mathbf{y}$ weight vector
$\mathbf{C}$	$\mathbf{y}$ weight matrix
$C^*$	Average alkyl chain length
$C_a$	Aromatic carbon
$C_{ah}$	Aromatic protonated carbon
$C_{ar}$	Aromatic ring carbons
$C_b$	Bridgehead aromatic carbons
$CH$	Fraction of carbon atoms present as CH groups

$CH_3$	Fraction of carbon atoms present as $CH_3$ groups
$C_n$	Long-chain n-alkyl carbons
$Cov$	Covariance matrix
$C_p$	Percentage of paraffinic carbons
$C_q$	Quaternary aromatic carbons
$C_r$	Integral intensity in the region 2.0-2.4 ppm
$C_s$	Substituted aromatic carbon
$C_T$	Total relative number of carbons
$C_v$	Calibration constant of the viscometer
$f()$	Activation function
$f(CH_n)$	Fraction of $^{13}C$ NMR intensity due to the various $CH_n$ groups
$f_a$	Aromaticity ( $^{13}C$ ) = fraction of aromatic carbon
$fA_m$	Molar content of aromatic groups
$fH_n^o$	Mole fraction of olefinic protons
$H_a$	Aromatic protons
$H_{ar}$	$^1H$ NMR aromaticity
$H_\alpha$	Number of $\alpha$ hydrogens
$H/C$	Hydrogen to carbon ratio
$(H/C)_o$	Hydrogen to carbon atom ration of olefinic bond structure
$(H/C)_t$	Total hydrogen to carbon atom ratio
$(H/C)_\alpha$	Aromatic $\alpha$ -hydrogen to $\alpha$ -carbons atom ratio

$H_{au}/C_a$	Ratio non-bridgehead aromatic carbon to total aromatic carbon
$I$	Identity matrix
$I(CH_n)$	Total $^{13}\text{C}$ NMR intensity due to the groups in parentheses
$I_d$	Integral intensity in the region 7.5-100 ppm
$IP$	Percentage of isoparaffin
$I_T$	Integral intensity in the region 5-160 ppm
$I_x$	Integral intensity in the region x ppm
$L_S$	Average side chain length
$M_A$	Mono-ring aromatics
$N$	Size of data table
$n$	Number of data points
$N(C^{al})$	Number of aliphatic C groups
$N(C^{ar})$	Number of nonprotonated aromatic carbons
$N(CH_3^{al})$	Number of aliphatic $\text{CH}_3$ groups
$N(CH^{al})$	Number of aliphatic CH groups
$N(CH_n)$	Average number of $\text{CH}_n$ groups
$n_a$	Average of alkyl chain length of aromatic substituents
$N_B$	Average number of branches per molecule
$NC$	Average number of carbons per molecule
$N_C$	Average n-alkane chain length
$n_n$	Average number of carbons per alkyl side chain

$NP$	Percentage of normal paraffin
$N_R$	Average number of rings per molecule
$N_s$	Number side chains per molecule
$N_{SB}$	Number of side chain branched per molecule
$\mathbf{P}$	$\mathbf{X}$ loadings matrix
$\mathbf{p}$	$\mathbf{X}$ loadings vector
$P_A$	Global di-plus-ring aromatics
$S/N$	Signal to noise ratio
$SW$	Spectral width
$t$	Time
$\mathbf{T}$	$\mathbf{X}$ score matrix
$\mathbf{t}$	$\mathbf{X}$ score vector
$T_A$	Number of total aromatic
$t - Me$	Percentage of terminal methyl groups
$T_W$	Total group molecular weight
$\mathbf{U}$	$\mathbf{y}$ score matrix
$\mathbf{u}$	$\mathbf{y}$ score vector
$V$	Angular momentum
$Var$	Variance
VI	Viscosity indexes
$W$	Mass of the sample

$W$	$\mathbf{X}$ weight matrix
$w$	$\mathbf{X}$ weight vector
$w_i$	Weight value – ANN model
$X$	Independent variables
$\mathbf{X}$	Matrix of explanatory variables
$\mathbf{X}_0$	$\mathbf{X}$ standardize matrix
$(\mathbf{X})^T$	Transpose of the matrix $\mathbf{X}$
$x_i$	Input value
$X_{min}$	Critical cetane dilution of the sample
$Y$	Dependent variables
$\mathbf{y}$	Vector of dependent variable
$\mathbf{y}_0$	$\mathbf{y}$ standardize vector
$\hat{Y}_i$	Model output
$\bar{Y}_i$	Average of the output variable
$Y_i$	Output value
$Z$	Standard normal variable
$Z_a$	Atomic H/C ratio for non-aromatic groups
$\gamma$	Magnetogyric ratio
$\delta$	Chemical shift
$\theta$	Bias

$\lambda$	Eigenvalue
$\mu$	Mean of each variable
$\nu$	Kinematic viscosity
$\sigma$	Degree of aromatic ring substitution
$\gamma$	

# 1

## INTRODUCTION

The importance of this study is demonstrated in this chapter. The motivation to develop this work and the scientific contributes for industries are also described. In addition, a description of the structure of the thesis is presented.

## **1. INTRODUCTION**

1.1. Scientific Relevance .....	3
1.2. Thesis Structure .....	5



## 1.1. Scientific Relevance

In the last decades, the petroleum industry had a great interest in refining light crude oils as they were easier to process and had lower operational costs. However, the quantities of light crude oil world reserves are decreasing and thus to acquire this type of crude oil becomes more expensive when compared to the purchase costs of heavier crude oils. With this and due to the growing worldwide energy demands and the increase demands of high quality distillates with tighter specifications, petroleum industries need to process increasingly heavier crudes and converted them from low value heavier petroleum fractions into lighter high value products. In the conversion of these heavier fractions, frequent innovations and new processes are implemented at petroleum industry to convert the heavier petroleum fractions and petroleum residues into products with more high commercial values. The problem associated with the refining of heavier petroleum fractions is the higher quantities of heavy residues that are obtained. The heavy residues are characterized for their complex composition rich in asphaltenes, sulfur, nitrogen and metal ions. The characterization of these fractions to obtain the products with desirable characteristics and defined specifications and to optimize the process conditions is a very important challenge in the refineries.

In the refining industry there are some appropriate methods, including the ASTM, which can be applied in the characterization of the physical-chemical properties of such products. However, there are some drawbacks associated with these laboratory analyses as time consuming, the cost of each analysis and the many involving steps such as sampling, sample handling, sample preparation, measurements, data handling and reporting. The problem associated with these laboratory analyses are related to the delay between sampling and validation of the analytical results. This delay prevents process conditions being corrected and adjusted at the right time by applying a corrective response, contributing to obtain products out of specifications and units working at wrong conditions. An alternative would be to implement online quantitative technologies that allow full and real time characterization of the control process conditions and the final products. The production of the required range of fractions at maximum yield and at a minimum cost is one of the refineries goals.

The physical-chemical properties of petroleum fractions are the result of the influence of the physical-chemical properties of each individual substance which is part of the sample and of the interactions that each individual substance experiences with the surroundings. The identification of these compounds contributes to the identification of the physical-chemical properties of the product analysed. Once the physical properties are correlated with the chemical compositions it is possible to apply spectrometric methods to quantitatively determine the physical properties of a composition of chemical substances, such as a petroleum product. Nuclear magnetic resonance (NMR) is the chosen technique, due to its capability to provide great information on the chemical nature of individual types of proton and carbon atoms in different and complex mixtures of petroleum fractions. The great advantage of this technique is that it measures the physical properties of transparent, opaque and dense solutions by only one measurement and it can be applied to many streams of an industrial refining complex (The Qualion Company, 2006).

The progress made in the interpretation of the NMR spectra with the development of different multivariate statistical methods can give relevant information about the identification and structural characterization of hydrocarbons and their physical-chemical properties. Principal component regression (PCR), partial least squares (PLS) and artificial neural networks (ANN) are examples of multivariate statistical methods used in this work to highlight the correlations among the NMR spectra and the target properties. Kinematic viscosity, density, carbon residue, P-value, sulfur content, flash point, vacuum and atmospheric distillation are examples of some of the properties analysed in this work.

The use of industrial applications of NMR spectroscopy and multivariate models can enhance the refining industries operation as a result of a better knowledge of the streams composition in the refining process, as well as in the prediction of the better operating conditions for obtaining refined products with desired specifications and in adequate quantities to achieve the market demands.

With this work it is expected to demonstrate that the combination of  $^1\text{H}$  NMR spectroscopy with multivariate statistical methods can be used to correlate the chemical composition of each stream with their physical-chemical properties. The real time knowledge of the physical-chemical properties of each stream is very important for process control contributing to obtain best performances in the operation of the units.

## 1.2. Thesis Structure

The structure of this thesis is summarily described in this section. This document is divided into seven chapters, with Chapter 1 as the introduction.

In Chapter 2, the characterization and classification of petroleum is briefly referred. An overview of petroleum processing history until the development of the modern petroleum industry is described. Afterward, the characterization of the petroleum industry, where this study was developed, and a more detailed discussion around the products and their physical-chemical properties which are object of study is presented. The aim of this chapter is to give a brief presentation about the main topics covered throughout the current work.

In Chapter 3, a comparison between some chromatographic and spectroscopic techniques capable to characterize petroleum fractions is made. This chapter is important to understand why NMR is the chosen technique for the determination of the physical-chemical properties of petroleum fractions. Subsequently, the NMR spectroscopy technique is described. Additionally, in this chapter a review about the applications of NMR in the analysis of petroleum and heavy petroleum fractions is presented.

In Chapter 4, the experimental section is presented. The description of the ASTM and IP standard methods used to determine the physical-chemical properties of the different analysed samples are described in detail. The set-up of the procedures adopted for the manipulation of the NMR spectra data is explained.

Chapter 5 presents a detailed description of the multivariate statistical methods used to exploit the information content of the NMR data. Principal component analysis is used to select the number of principal components that are needed as inputs for principal component regression and artificial neural networks. Principal component regression, artificial neural networks and partial least squares are used to highlight the correlations among the NMR spectra and the properties proposed in Chapter 4.

In Chapter 6, the results of this work are presented and discussed for the different samples and using different multivariate models. The performance of the models for predicting the

physical-chemical properties of the different samples with NMR are evaluated by using statistical models.

In Chapter 7, the main conclusions of this thesis are pointed out as well as some suggestions for future work.

# 2

## PETROLEUM INDUSTRY

Petroleum refining industry uses crude oil as raw material in the production of a large assortment of commodities and products important to human life and in many other processes. Products such as gasoline, gas oil, jet fuel, naphtha, gas, fuel, bitumen, petrochemicals are examples of some derivatives that are obtain after crude oil processing. The development of more adequate refining processes over the years contributed to increase the production effectiveness and the quality of crude oil derivatives.

## **2. PETROLEUM INDUSTRY**

2.1.	Crude Oil .....	9
2.2.	Crude Oil Refining .....	14
2.2.1.	Crude Oil Refining in Portugal .....	21
2.2.2.	Matosinhos Refinery .....	22
2.2.3.	Fuel Plant .....	25
2.3.	Upgrading the “Bottom of Barrel” .....	29
2.3.1.	Vacuum Distillation .....	29
2.3.2.	Visbreaker Unit .....	30
2.4.	Crude Oil Derivatives .....	34
2.4.1.	Vacuum Gas Oil (LVGO and HVGO).....	35
2.4.2.	Fuel Oil .....	36
2.4.3.	Physical-Chemical Characterization of Petroleum Fractions .....	37
2.5.	Final Remarks.....	42

## 2.1. Crude Oil

Crude oil, also called petroleum, continues to be a very important power source for our civilization. It is flammable at a room temperature and can be very fluid and clear or very dark and viscous, depending on the quantity of light or heavy distillates, respectively. Crude oil is generally lighter than water and it presents a characteristic and strong odour. It is known that crude oil doesn't have a simple chemical formula, but it has a very complex chemical composition that allows wider variety of uses. It is possible to say that crude oil is a mixture of many different hydrocarbons and small amounts of impurities (Table 2.1). These compounds as well as their proportions vary significantly with crude oils. Even crude oils from the same geographical area can be very different depending on its source, although, at elemental level, the typical composition of these compounds are carbon (83-87%), hydrogen (11.5-14%) and heteroatoms (1-5.5%) (Simanzhenkov *et al.*, 2003).

Crude oil can be classified in different ways depending on: i) the type of hydrocarbons; ii) the American Petroleum Institute (API) gravity; iii) the sulfur content; and iv) the acid content. Taking into account the different hydrocarbons, paraffinic, naphthenic and aromatic hydrocarbons, crude oil can be classified as paraffinic, paraffinic-naphthenic, naphthenic, aromatic and aromatic-naphthenic crude. With these designations an idea is given on which type of hydrocarbons exist in higher quantities, since each crude oil contains all different types of these hydrocarbons in different percentages.

**Paraffinic crude** – Crude with higher quantity of paraffinic hydrocarbons, normally more than 75% of paraffinic compounds. This type of crude is characterized as a light and fluid crude, for having a smaller density (smaller than 0.85 g/mL), a smaller viscosity, a smaller quantity of sulfur and a smaller quantity, smaller than 10%, of resins and asphaltenes. Normally a paraffinic crude is responsible for the production of: i) gasoline with lower octane index; ii) kerosene with high quality; iii) gas oil with good combustion conditions; iv) lube oils with high viscosity index and chemical stability; and v) residuum rich in paraffins.

**Paraffinic-naphthenic crude** – This type of crude is characterized for having a rich composition in paraffinic compounds (50 to 75%) and a smaller amount of naphthenic compounds (less than 20%). The quantity of resins and asphaltenes is small as well as the sulfur content. This crude has higher viscosity and density compared with a paraffinic crude.

**Naphthenic crude** – A naphthenic crude is characterized for having a high quantity of naphthenic compounds and a small percentage of sulfur. Normally the use of naphthenic crude contributes to the production of gasoline with higher octane index.

**Aromatic crude** – This crude normally contains about 50% of aromatic compounds. It is characterized for having more asphaltenes and resins than the other crudes (normally 10 to 30%) and higher quantity of impurities such as sulfur (more than 1%).

**Aromatic-naphthenic crude** – An aromatic-naphthenic crude is characterized for having on its composition less than 35% of naphthenic compounds. The quantity of resins and asphaltenes is higher when compared with an aromatic crude and it can be higher than 25%. When compared with an aromatic crude the quantity of sulfur can be lower, normally having values between 0.4 to 1%.

Relatively to the API gravity scale, a higher API values means a lower density and, consequently, means that crude oil has a higher quantity of valuable low boiling point fractions. Taking into account the API gravity scale, crude oil can be classified as heavy or light crude. Heavy crude oil is richer in higher boiling fractions, contains greater amounts of aromatic and naphthenic structures and higher quantity of heteroatoms. Light crude oil contains lower boiling constituents and waxy molecules and smaller amounts of heteroatoms such as nitrogen, oxygen, sulfur and metals (Altgelt *et al.*, 1994). This light



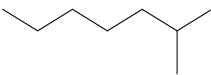
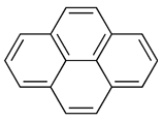

crude oil is the easiest and cheapest to produce and refine but the main problem is that the world's oil resources available are from heavy and viscous crude oils.

Due to the sulfur content, crude oil can be classified as sour (high amount of sulfur) or sweet (low amount of sulfur). The presence of higher quantities of sulfur may have environment impact and effects on refining operations (Speight, 2002).

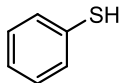
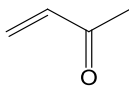

Relatively to the acid content, an acid crude oil is classified for having the total acid number (TAN) higher than 1.0 mg KOH/g. The use of crude oils with a high TAN can give rise to corrosion problems in the refineries and several problems with product quality and environmental protection (Hsu *et al.*, 2006).

Normally the heavier and denser crude oil has a minimal value but it can be used in the production of many fuels, like gasoline, gas oil, naphtha, kerosene and fuel oil, and it can provide raw materials that are used in different industries to make plastics, paints and fibers. When crude oil is refined it gives high valued products although further investments and efforts are needed to take advance of the "bottom of barrel" and with that increase the production of these products with high commercial interest.

**Table 2.1:** Composition of crude oil.

<b>Hydrocarbons</b>	Paraffins ( <i>e.g. isooctane</i> )		Can present different structures, straight chain (normal paraffins) and branched chains (isoparaffins). Of these two different structures (straight and branched chains), the most frequent in the heavier fractions of crude oil are the branched chains (Simanzhenkov <i>et al.</i> , 2003). The branched chains have a high octane index leading to a high quality of gasoline.
	Aromatics ( <i>e.g. pyrene</i> )		The presence of aromatic compounds in heavy crude oils play an important role in the production of gasoline and in the production of aromatic compounds (ex.: toluene and benzene). It is possible to find heavy aromatic or light aromatic compounds depending on the petroleum fraction in analysis. The light ones have huge benefits for some units, as the platforming, as these compounds provide good octane indexes. The heaviest ones are responsible for some problems that may occur in some petroleum products, such as jet and diesel.
	Naphthenes ( <i>e.g. cyclohexane</i> )		Cycloalkanes, known as naphthenes in the petroleum chemistry, are cyclic saturated hydrocarbons defined by the number of rings as mononaphthenes (monocyclic alkanes, $C_nH_{2n}$ ), dinaphthenes (dicyclic alkanes, $C_nH_{2n-2}$ ) and trinaphthenes (tricyclic alkanes, $C_nH_{2n-4}$ ). They are found in high quantities in most crude oils and are relatively stable.

**Table 2.1:** Composition of crude oil (continuation).

		<b>Drawbacks</b>	
<b>Heterocompounds</b>	Asphaltenes (*)	Constitute the heaviest fraction of crude oil and are a complex mixture of molecules that contains significant quantities of heteroatoms (S, N, V, Ni, ...), condensed aromatic rings, aliphatic chains and naphthenic rings.	They are responsible for the obstruction in flow lines, pipelines and petroleum wells and to the poisoning of the catalyst used in the cracking and hydrocracking process (Wauquier, 1995).
	Sulfur (e.g. thiophenol)	 Sulfur may be present in the form of sulfites, thiophenols, cycloalkanethiols, thiophenes, benzothiophenes and alkylbenzothiophenes. Crude can be classified as sour or sweet crude depending on the quantity of sulfur (values higher than 0.5% means sour crude and less than 0.5% means sweet crude).	They contribute to the development of corrosion, which damages the equipment, contributes to the air pollution due to the emission of SO <sub>x</sub> in the atmosphere, to the catalyst contamination and are toxic. Sour crude cause many problems during the refining and its treatment is expensive and difficult because it needs specific equipments.
	Oxygen (e.g. methyl vinyl ketone)	 These compounds may be present, for example, as carboxylic acids (ex: naphthenic acids), phenols, esters, amides and ketones.	They are responsible for the crude acidity. A high acidity means problems in the refineries as it aggravates the effect of corrosion in equipments and pipelines, makes changes in the colour and odour and also in the costs associated with the equipments protection.
	Nitrogen (e.g. pyridine)	 These compounds may be present in the organic form, for example, in the form of pyridines, quinolines, pyrroles, polycyclic compounds with oxygen and sulfur.	They contribute to the catalyst poisoning and are responsible for the change in the quality of the final products.
	Metal ions	It is possible to distinguish nickel, vanadium, iron, zinc, mercury, boron, sodium, potassium, calcium and magnesium.	They are responsible for the catalyst poisoning.

(\*) Asphaltenes do not have a chemical molecular structure. This will be discussed in the Chapter 3.

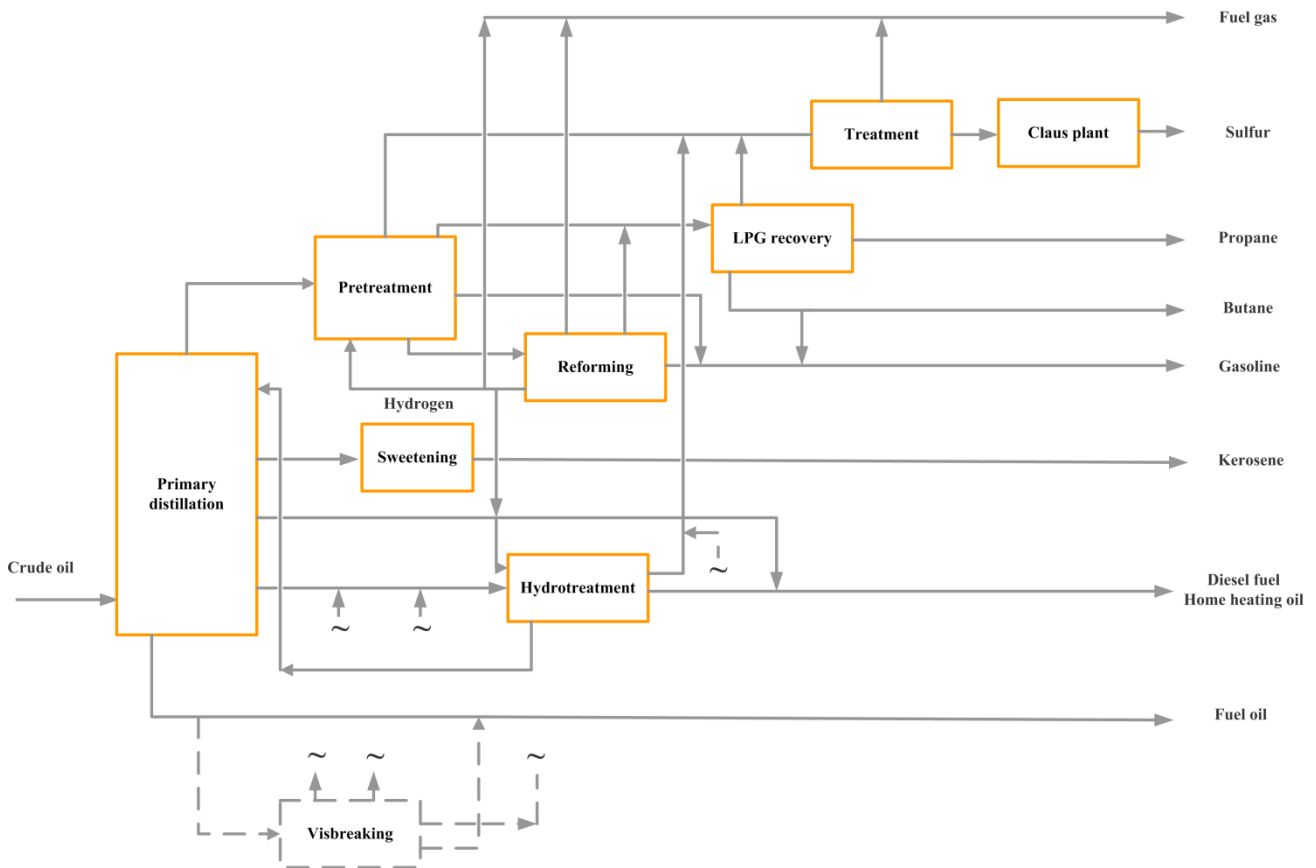
## 2.2. Crude Oil Refining

Since the discovery of petroleum, many refinery industries have been developed all over the world being the first refinery opened in 1861 (Hsu *et al.*, 2006). Initially, the most valuable petroleum fraction was kerosene, produced by simple atmospheric distillation (Hsu *et al.*, 2006; Jones *et al.*, 2006), while the distillates heavier than kerosene were normally used as fuel oils and for making paving and roofing asphalts and the undistilled residues were discarded (Altgelt *et al.*, 1994; Hsu *et al.*, 2006). Nowadays the need in the variation of products as gasoline, gas oil, jet fuel and petrochemicals, contributed to the use of petroleum and its heavy ends. The growing of the automobile industry in the beginning of the twentieth century was one of the main factors that contribute to the development of the petroleum refining. Imagining the markets needs, in the near future, and with the object to increase the efficiency and the quality in the production of crude oil derivatives significant investments were made in the refining processes. With that, the refineries that process light and medium crude oils suffer significant changes to refine heavy crude oil. This heavy crude oil contains higher quantities of heavy fractions that, due to its molecular structure (long molecular chains) and its higher boiling point, requires the use of more energy in the refining processes. This explains the necessity to improve the refinery industry with the aim in improving the quality of petroleum fractions and the efficiency in the conversion of the heavy fractions. It is possible to say that the evolution of the refinery industry was notorious over the last years and it occurred due to the worldwide demand for refined products and environmental laws that become increasingly stringent contributing to the development of new processes (Wauquier, 1995). The most important refining steps can be included in different processes as the separation, conversion, finishing and the environmental protection processes (Simanzhenkov *et al.*, 2003; Speight, 2006). These processes, shown in Table 2.2, have emerged and improved over the years.

**Table 2.2:** Basic operations in petroleum processing.

Process	Definition	Examples
<b>Separation</b>	Separation of the constituents of crude oils into fractions having similar properties taking into account differences of boiling point, melting point, density or solubility.	Distillation Absorption Extraction Crystallization Adsorption
<b>Conversion</b>	Conversion of the separated hydrocarbons into new molecules that contributed to obtain the desirable products with desirable properties.	Catalytic reforming Isomerization Alkylation Ether synthesis Oligomerization Visbreaking Coking Catalytic cracking Steam reforming Hydroconversion
<b>Finishing</b>	Exclusion of elements and compounds that are not desirable to improve the quality of the desirable products	Hydrotreatment / Hydrogenation Sweetening
<b>Environmental Protection</b>	Responsible for the processing of effluents and gases.	Acid gas processing Stack gas processing Waste water treatment

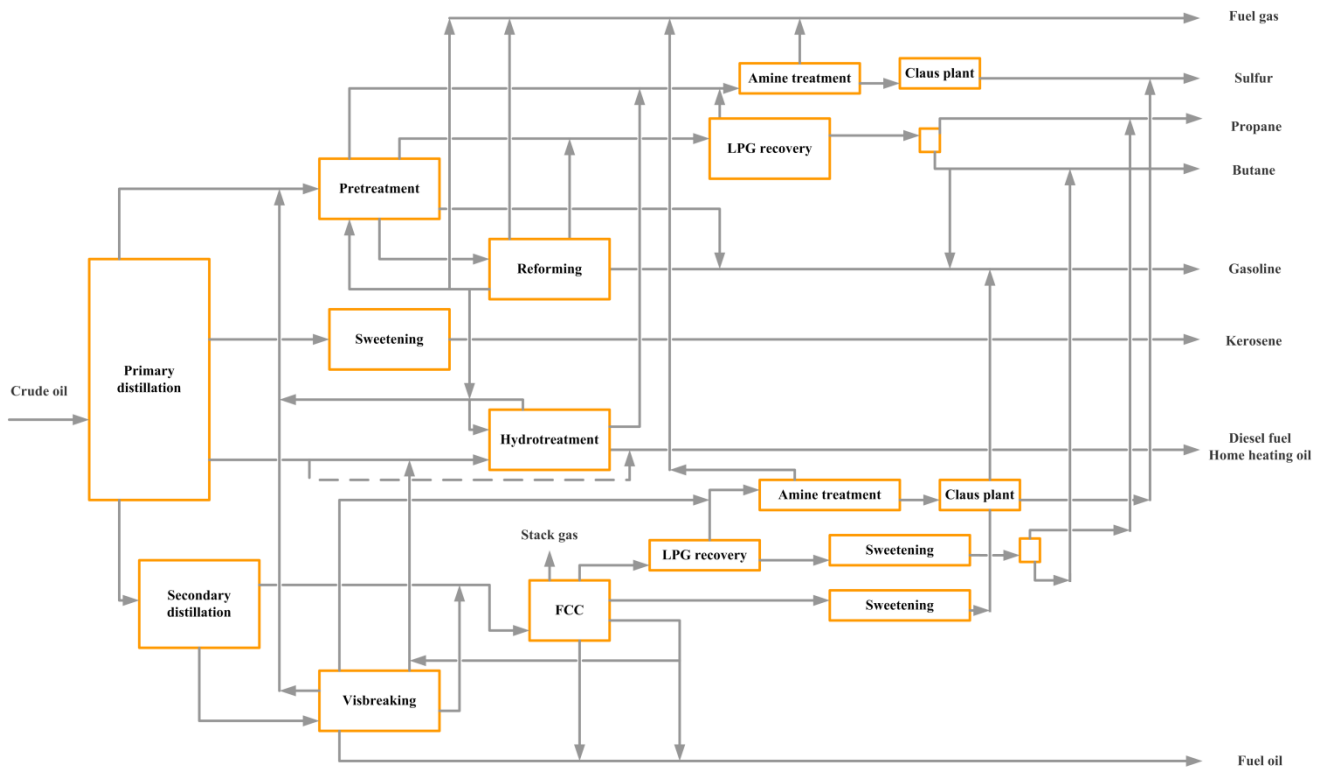
The first petroleum industries, denominated as simple refinery, were characterized for the production of gasoline, gas oil, domestic heating oil and industrial fuel oil. The refining steps included a primary distillation, a catalytic reforming along with a pretreatment step for improving the gasoline pool octane number, a partial hydrodesulfurization of the gas oil fraction and liquefied petroleum gas (LPG) and kerosene sweetening. These refineries were characterized for the absence of conversion process in product streams (excluding heavy naphtha stream) and for producing 40 to 50% of residual fuel oil. Figure 2.1 represents a typical refining flowsheet of a simple refinery, where it is possible to see a catalytic reforming, very important to upgrade the gasoline octane number and to the production of hydrogen. Hydrogen was then used in other processes to enrich other fractions. In the same figure it is also possible to see a visbreaker unit although it was not widely used in that period and so considered as an alternative unit to the production of fuel oil (Wauquier, 1995; Decroocq, 1997).



**Figure 2.1:** Example of a refining flowsheet of a simple refinery. Adapted from Wauquier (1995).

Since the simple refinery was no longer suitable to the market, a rapid growing of petroleum industries was visualized. The implementation of new units as the vacuum distillation, the catalytic cracking and/or hydrocracking of the vacuum distillate and the visbreaking were one of the main factors that helped refineries in the production of more gasoline and distillates. With this, refineries could answer the markets demands for other petroleum fractions. Consequently, there was a decrease in the demand of residual fuel oils which production passed from 40-50% to 20-25% of crude. Although, with this type of conventional refineries it was notorious an increase in the content of pollutants like sulfur, nitrogen and metals (Wauquier, 1995; Decroocq, 1997; Hsu *et al.*, 2006). This increase of pollutants, which did not respect the environmental regulations, could be related to the absence of process responsible for: i) lowering the sulfur content in gas oil and domestic heating oil; ii) the reduction of  $\text{SO}_x$  emissions; and iii) the elimination of lead in gasoline. The presence of lead in gasoline was related to the poisons that contaminated the active

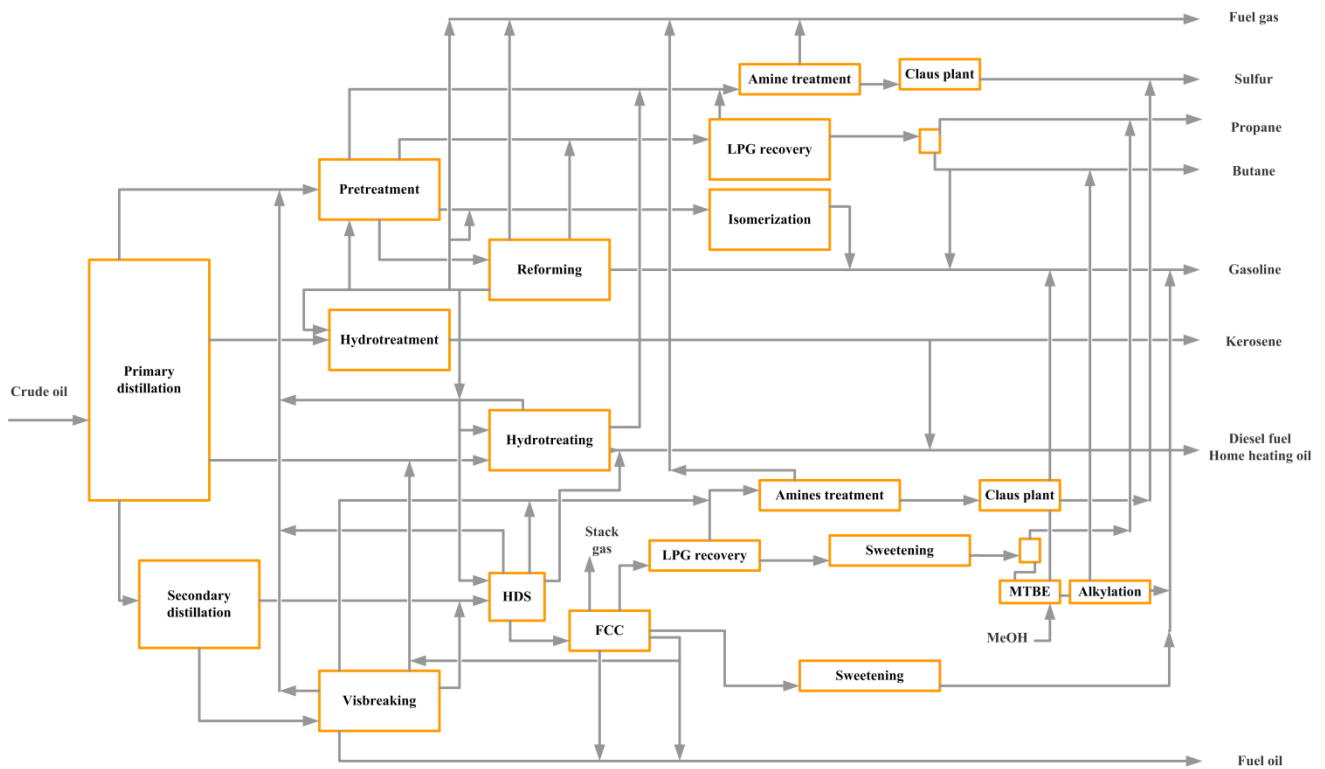
metal of the catalytic converters used to remove carbon monoxide and hydrocarbons from automobile exhaust (Wauquier, 1995; Hsu *et al.*, 2006). In Figure 2.2 it is possible to see an example of this conventional refinery.



**Figure 2.2:** Example of a refining flowsheet of a conventional refinery. Adapted from Wauquier (1995).

To achieve a solution to respect the environmental regulations and to respect the requirements for reformulated gasoline and low sulfur content, the refineries started to develop and implement new processes and units, as shown in Figure 2.3. Isomerization, etherification (MTBE – methyl-t-butyl ether, ETBE – ethyl-t-butyl ether, TAME – methyl-t-amyl ether) and alkylation are examples of three units that helped in the reduction and elimination of lead in the gasoline. The hydrodesulfurization of the fluid catalytic cracking (FCC) feedstock and the hydrodesulfurization of gas oils and domestic heating oils are examples of hydrotreating that contributed to the reduction of  $\text{SO}_x$  emissions and to obtain products with high quality. In these refineries it become regular to use low sulfur light

crude since the conversion of heavy residues was not always made (Wauquier, 1995; Hsu *et al.*, 2006).



**Figure 2.3:** Example of a refining flowsheet of petroleum industries. Adapted from Wauquier (1995).

In today's refineries the great challenge is to convert large quantities of heavy crude oils and to reduce the aromatics content. Since the quantities of the light crude oils became less available and more expensive, the refineries had to introduce complementary conversion processes to transform the heavy crude oils into specific sets of refined products to meet the market demands. These heavy crude oils are generally less expensive and give greater yields of lower value higher boiling products, therefore, complex and expensive processes have been introduced to convert them into lower boiling products. Units as hydrocracking and catalytic cracking have their high importance in the production of diesel motor fuel and gasoline, respectively. Other new processes as the alkylation, polymerization, isomerization and the dehydrogenation of n-butane may have greater importance. The growing of the refineries depends on the type of the used crude oils, light crude oil or



heavy crude oil, on its costs, the market demands and on the final product specifications. To remain viable, to meet the markets demands and every crude supplies, the refineries need to be constantly adapt and upgrade. As a result, the simple refinery, highly dependent on local markets, responsible to prepare feedstocks for petrochemical manufacture, for the production of industrial fuels, large quantities of unfinished oil, desulfurized distilled fuels and high octane gasoline, upgrade to the most versatile conversion refinery. This conversion refinery besides all the bases units that already exist also incorporates new conversion processes, new solvent extraction processes for manufacturing lubricants and petrochemical units to recover propylene, benzene, toluene, and xylenes. Consequently, an increased in the yields and quality of the refined petroleum products is notorious (Wauquier, 1995; Decroocq, 1997; Hsu *et al.*, 2006; Speight, 2006). Figure 2.4 represents a possible flowsheet of a today's refinery.

The intention of the flowsheets presented is to give an idea about the growth and development of new processes in refineries. However the refineries are not all the same, there are many different things that differentiate them, as the location, history and markets drivers. Each refinery has its own flowsheet that depends on the refinery goals. There are some refineries more interested in the production of gasoline whereas other refineries are more oriented toward the production of middle distillates such as jet fuel and gas oil. Besides all the differences, most refineries perform the same basic operations presented in Table 2.2.

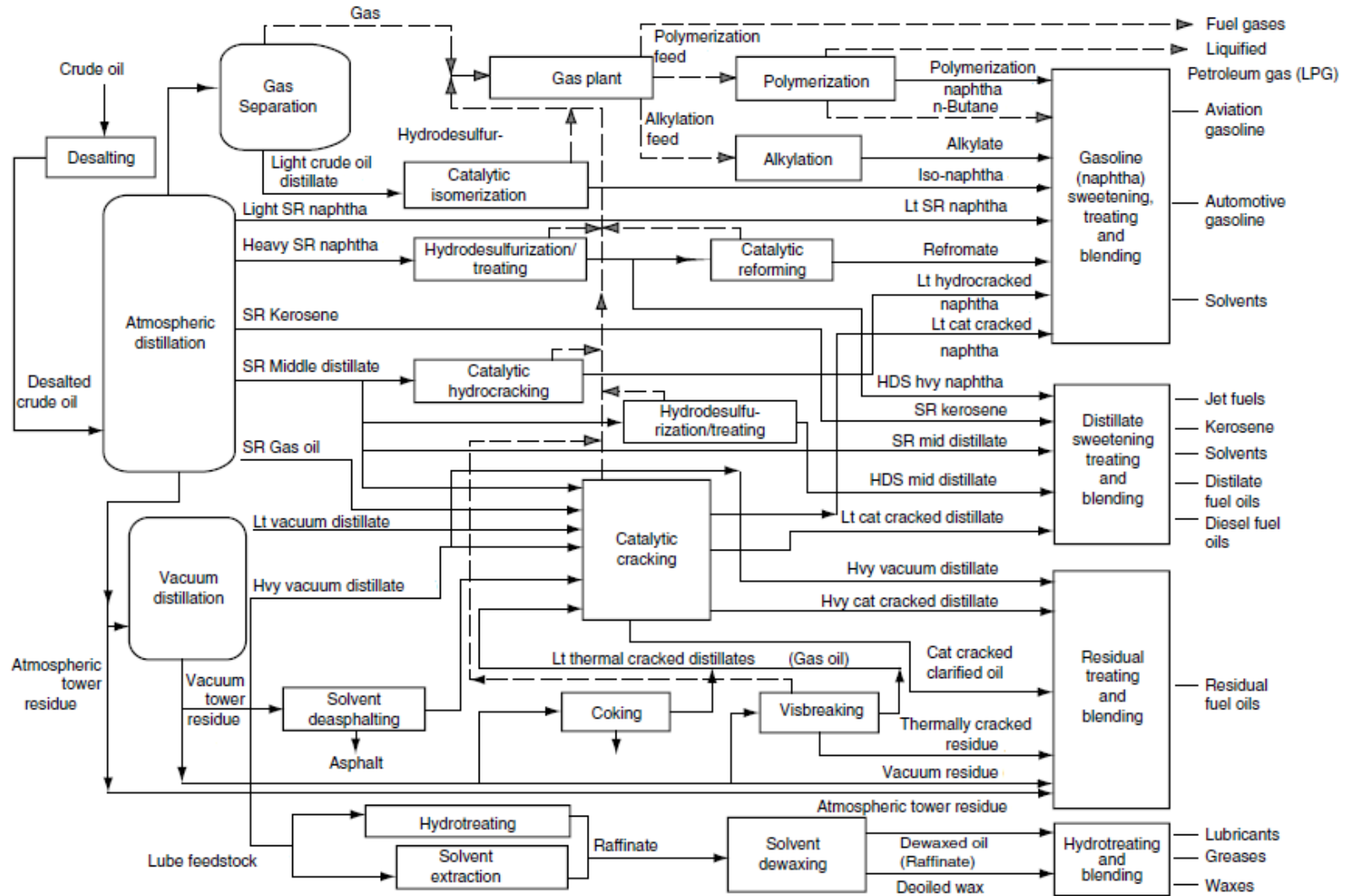


Figure 2.4: Example of a possible today's refining flowsheet (Speight, 2006).

### 2.2.1. Crude Oil Refining in Portugal

Galp Energia S.A. is an example of a company operating with a modern petroleum refining. Galp Energia S.A. is responsible for the exploitation and production of crude oil and natural gas, for the refining and distribution of petroleum products, for the distribution and sale of natural gas and generation and commercialization of electricity (Galp Energia, 2010a). It has a capacity of refining 330 thousand barrels per day of crude oil. Galp Energia is divided into two refineries, Matosinhos refinery (Figure 2.5) which has a 110000 bbl/day (barrel per day) distillation capacity and Sines refinery which has a distillation capacity of 220000 bbl/day. These two refineries working together in a coordinate and integrate way are responsible for the normal supply of fuels, base oils, aromatics, asphalt and other products needed in different sectors of activity including aviation, marine transportation and lubricants industry (Galp Energia, 2010a; Galp Energia, 2010b; Galp Energia, 2011b).



**Figure 2.5:** Matosinhos refinery - Galp Energia S.A. (Galp Energia, 2010b).

### 2.2.2. Matosinhos Refinery

Matosinhos refinery, located in the north of Portugal, started operating in 1969. It is an example of a modern refinery presenting different units and processes, incorporating most of the processes presented in Figure 2.4. Matosinhos refinery doesn't have a hydrocracking and catalytic cracking but produces and sends the feedstocks for these units located at Sines refinery. One of the goals of Matosinhos refinery is to increase the production of gas oil due to the implementation of two new units as the vacuum distillation and the visbreaker unit. With these units the refinery has now the ability of processing different heavy and light crude oils and is prepared to receive acid crude oils. In addition, to improve the environmental performance other units responsible for the treatment of acid water and for sulfur recovery were implemented (Figure 2.6). Moreover, to reduce the environmental impact and to obtain a more energetically efficient refinery a new cogeneration unit is implemented (Galp Energia, 2010b; Galp Energia, 2010c; Galp Energia, 2011b).

Today, Matosinhos refinery has a storage capacity of 1.78 million m<sup>3</sup>, approximately, of which 649000 m<sup>3</sup> are for different crude oils and 1.132 million m<sup>3</sup> are for intermediate and final products. It is responsible for the production of fuels (3700000 ton/year), base oils (150000 ton/year), aromatics and solvents (440000 ton/year), lube oils (1500 ton/year), paraffins (10000 ton/year), bitumen (150000 ton/year) and sulfur (10000 ton/year). Besides, it is responsible for the production of important raw materials (e.g. chemical naphtha, base oils, aromatic solvents, fuel oils, etc.) for other important industries as the chemical, petrochemical, plastics, textiles, fertilizers and paints. The production of all these different products are only possible due to the existence of different and complex operations that are performed in different process units, integrated in five different factories (Galp Energia, 2010b). These factories are:

**1. Lubes plant** - in this area the main purpose is the production of lubricating oils and greases from paraffinic base oils;

**2. Aromatics plant** – consuming reformat, which is its principal raw material, this factory is responsible for the production of important raw materials for other industries as petrochemical;

**3. Base oil plant** – production of base oils and other products like paraffins and bitumens using the Arabian light crude oil as raw material;

**4. Fuels plant** – using crude oil as raw material it is responsible for the production of a great variety of fuel products, some of them used as raw materials for the aromatic and solvent plant. Some of the intermediate and final products produced at this plant were the base to develop this project.

**5. Utilities plant** – this part is very important in the correct operation for all other factories since it is responsible for the production and distribution of energy and utilities. Without the utilities plant the refinery doesn't operate. As visualized in Figure 2.6, in the flowsheet of Matosinhos refinery the facilities of the utilities plant (thermoelectric plant, treatment of boiler water, water cooling system,...) are not evident.

**galp ARL** **PORTO REFINERY BLOCK DIAGRAM**

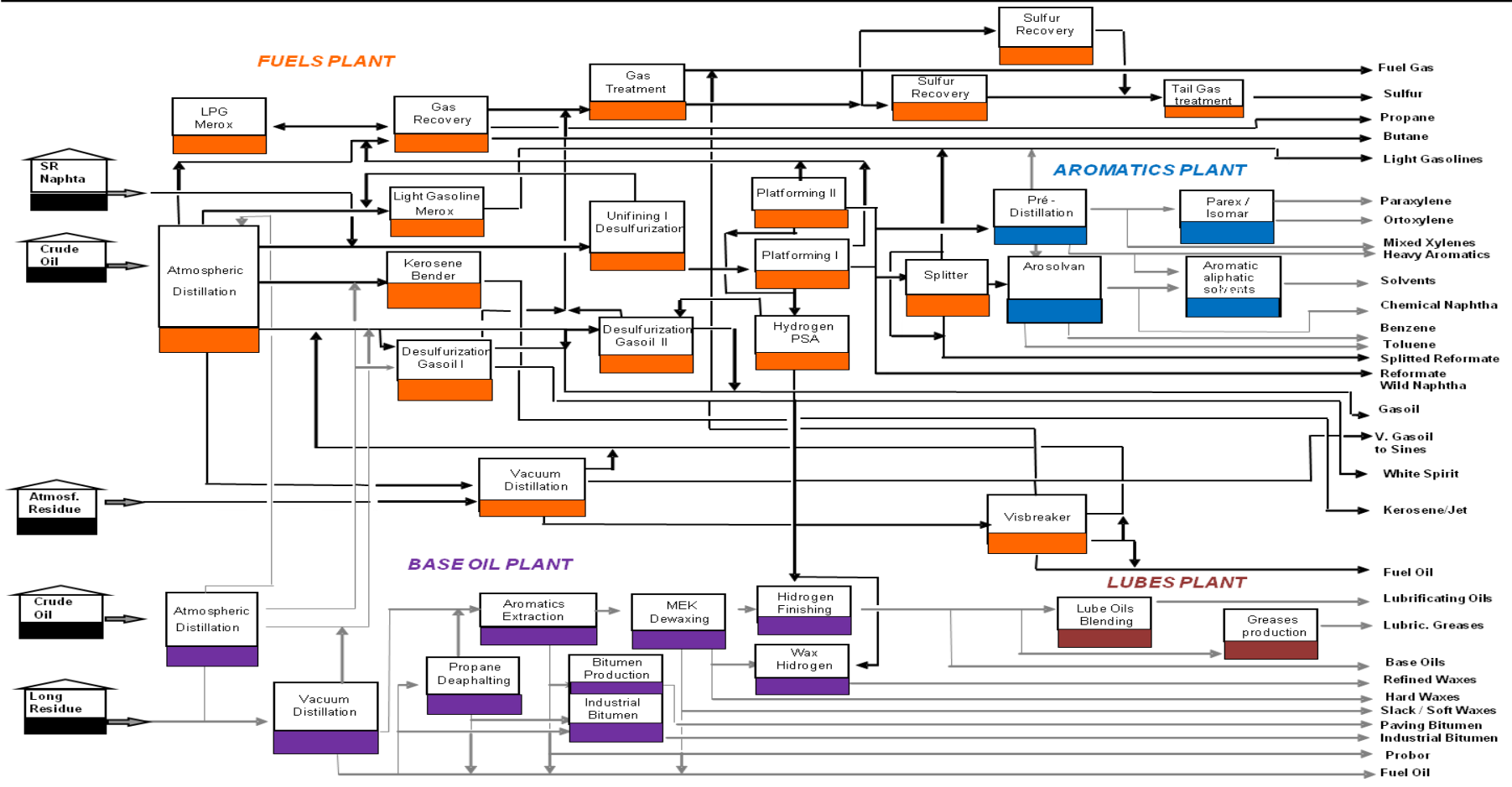


Figure 2.6: Flowsheet of Matosinhos refinery (Galp Energia, 2011c).

### 2.2.3. Fuel Plant

The fuel plant of Matosinhos started its activity in 1969. It has a distillation capacity of 12000 ton/day and it produces different fuels important in the market, raw materials for the aromatic and solvent plants and other products that are feedstock to the cracking unit of Sines refinery. These products are obtained due to the existence of a great variety of processes at the fuel plant, such as physical separation, chemical treatment and conversion processes as the conversion of the molecular structure of certain hydrocarbons families.

The fuel plant, shown in Figure 2.7, is composed of a large set of unit processes which uses crude oil as the main raw material (Galp Energia, 2011a). Initially, crude oil was subjected to an atmospheric distillation (3000 unit<sup>1</sup>) where the separation of the different hydrocarbons were promoted by heating, cooling and separation under specific temperature and pressure conditions. The atmospheric distillation was fed by a mixture of different crude oils with different characteristics with the idea to optimize and adjust the production to the market needs. As these obtained products do not meet the necessary specifications to be marketed they are subsequently treated in different units passing through separation and conversion processes until commercial products are obtained. After the atmospheric distillation, the intermediate products obtained were processed to obtain the following final products:

**1. Gases.** The light fraction resulting from the atmospheric distillation after treatment (3600 unit) is separated in a fuel gas and LPG stream. Before the production of fuel gas, the light fraction is sent to the 1500 unit for an acid gas treatment. This unit is responsible for the extraction of the sulfide gas ( $H_2S$ ), in several gas streams, by an absorption processes using diethanolamine (DEA). The obtained sulfide gas is then sent to the unit responsible for the sulfur recovery (3800/10800 unit) and the fuel gas is sent to the fuel gas system in the refinery. Relatively to the LPG stream it is separated in commercial propane and butane;

---

<sup>1</sup> Unit is referred as a unit process which comprises the fuels plant. As visualize in Figure 2.7 each process is characterized by a different number.

**2. Light gasoline.** The light gasoline is sent to the gas oil/ hydrodesulfurization unit (1400 unit), to reduce the sulfur content. In this unit (1400 unit), a treatment with hydrogen and in the presence of an appropriate catalyst is necessary for the extraction of these compounds. After the treatment, the light gasoline can be used in the production of other products;

**3. Heavy gasoline.** With the intention to remove the sulfur compounds, the heavy gasoline is sent to the 1200 unit. In this unit a catalytic conversion of some compounds as sulfur, oxygen, nitrogen and other contaminants occurs, after which they are removed. This only occurs under an atmosphere of hydrogen and in the presence of an appropriate catalyst. The products obtained in this unit are heavy gasoline desulfurated, the light gasoline and gas. The heavy gasoline goes to the platforming unit (1300 and 3300 unit), the light gasoline goes to the gas oil/ hydrodesulfurization (1400 unit) while the gas flows are sent to the treatment unit (1600 unit). In the platforming unit the idea is to increase the octane number of the heavy gasoline. For that, the heavy gasoline desulfurated is blended with a gas stream rich in hydrogen and passes, due to the existence of a catalyst, by different chemical reactions. A gas rich in hydrogen, a mixture of propane and butane and a gasoline with a high octane number (reformate) are obtained in this unit. The reformate is then used as feedstock to the aromatic plant and as blending in gasoline;

**4. Petroleum.** After being obtained, petroleum atmospheric derived from the atmospheric distillation and from storage is sent to the 1400 unit (gas oil/ petroleum hydrodesulfurization) for treatment. The main feedstock of this unit is a mixture of gas oil and petroleum which passes through a desulfurization process with hydrogen. The final product is used as blending for gas oil. Besides that, in this unit it is possible to treat the White-Spirit, a product from the base oil plant, that is widely used in other industries as those of paints, solvents and resins;



**5. Gas oil.** With the intention to reduce the quantity of sulfur, the gas oil is sent to the gas oil desulfurization unit (3700 unit). For that the gas oil passes through a hydrodesulfurization process in the presence of hydrogen and an appropriate catalyst. Although, gas oil can also be sent to the 1400 unit;

**6. Residue from the atmospheric distillation unit.** To increase the conversion of the heaviest fractions with lower commercial value, the residue from the atmospheric distillation unit is sent to the vacuum distillation unit (10000 unit). Here the atmospheric residue, which comes from the atmospheric distillation and from storage, is processed at a lower pressure than the atmospheric one. Heavy vacuum gas oil (HVGO), medium vacuum gas oil (MVGO), light vacuum gas oil (LVGO) and vacuum residue are the products obtained after the vacuum distillation. Normally, HVGO can be used as feedstock of a catalytic cracking or of a hydrocracking unit. The MVGO is mixed with HVGO and then sent to storage while the LVGO can be used as feedstock in the gas oil desulfurization unit (3700 unit). The residue from the vacuum distillation unit goes to the visbreaker unit (10100 unit). The visbreaker unit is used for increasing the conversion of the heavier fractions into light fractions such as gasoline and gas oil and to decrease the viscosity of the residue from the vacuum distillation unit. For that, a thermal cracking process taking into account the temperature and the residence time is used.

With the description of all processes necessary for the production of final products with commercial value it is possible to conclude that crude oil needs to pass through different physical-chemical operations, with high complexity, to be used by the end consumer. It is this complexity that gives a refinery its very own characteristics and potentialities.

From all the processes which were presented from the fuels plant, the vacuum (10000 unit) and visbreaker (10100 unit) units are considered the most important for this project as the main idea is to analyse the products obtained when upgrading the “bottom of barrel”.

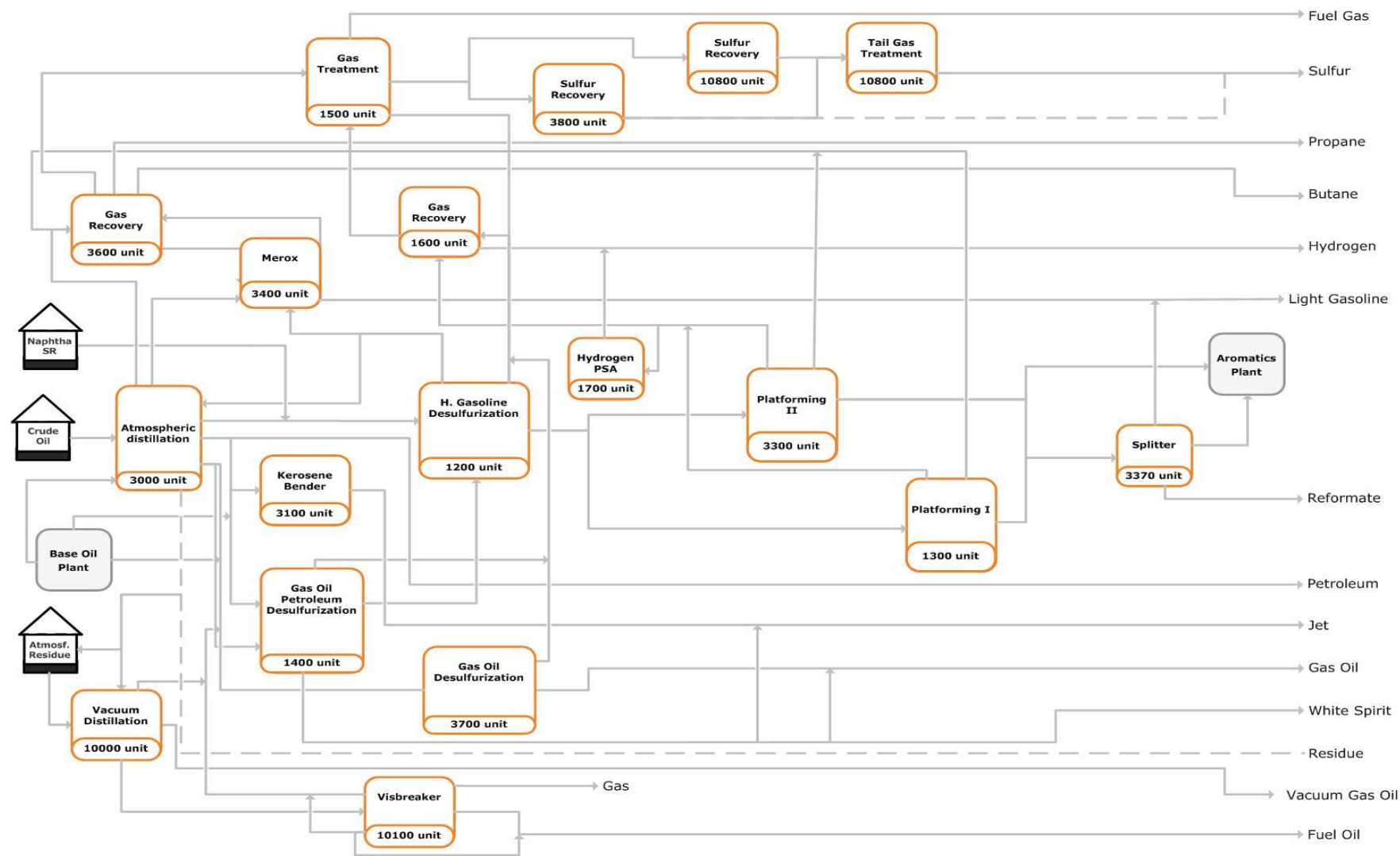


Figure 2.7: Flowsheet of the fuels plant of Matosinhos refinery (Galp Energia, 2011a).

## 2.3. Upgrading the “Bottom of Barrel”

Nowadays, the great challenge for the crude oil refining industry is to refine large quantities of heavy crude oils and to transform them into specific sets of refined products to meet the market demands (Wauquier, 1995). This means that all crude mixtures should be converted from heavy fractions to light distillate products in multiple complex refining steps, including distillation (atmospheric and vacuum distillation), catalytic cracking, hydrocracking, coking (Altgelt *et al.*, 1994) and other processes as previously stated. Vacuum distillation and visbreaking are processes used to convert the heavy fractions or residues to light distillate products. These heavy fractions, called bottom-of-the-barrel, gained significant importance due to the increase in competitiveness in the markets, the continual increase in the crude oil prices, the increase into the global energy consumption and the demand for transportation fuels.

### 2.3.1. Vacuum Distillation

The vacuum distillation unit receives the residue from the atmospheric distillation unit to recover additional distillates. This distillation occurs under high vacuum conditions in a vacuum column where very low pressure is guaranteed (Jones *et al.*, 2006). One of the main advantages of this distillation is that it allows the distillation of the residue using lower temperatures comparing with the atmospheric distillation. At these temperatures, the thermal cracking of the components does not occur as easily, although if occurring, it can lead to the formation of coke and olefinic products (Jones *et al.*, 2006). The presence of coke is undesirable due to coke deposits formed in the pieces of equipment and piping, which reduces the process unit run-time and increases the maintenance costs.

A typical vacuum distillation includes a preheating system, a furnace, a vacuum column, a top system, an amine absorber and a tempered water system (see Figure 2.8). With the preheating the residue from the atmospheric distillation unit reaches the adequate temperature before entering in the furnace. At the furnace and, before to be sent to the vacuum column, the heating and the vaporization of the residue occur. When the residue goes to the flash zone, in the column, to ensure maximum distillate yield, the temperature should be higher and the pressure lower (Speight, 2006). The acid gas is cooled to remove

the water from the condensate hydrocarbons and subsequently sent to the amine absorber to produce the off-gas stream which is then sent to the furnace. The obtained liquid fractions are the light vacuum gas oil (LVGO), medium vacuum gas oil (MVGGO), the heavy vacuum gas oil (HVGO) and the vacuum residue. In Matosinhos refinery the LVGO can be used as cutter stock of fuel oil or it can be sent to the gas oil desulfurization unit. The MVGO and HVGO are sent together to storage or can also be used as cutter stock of fuel oil. The residue from the vacuum distillation unit is the raw material of the visbreaker unit (Galp Energia, 2011a).

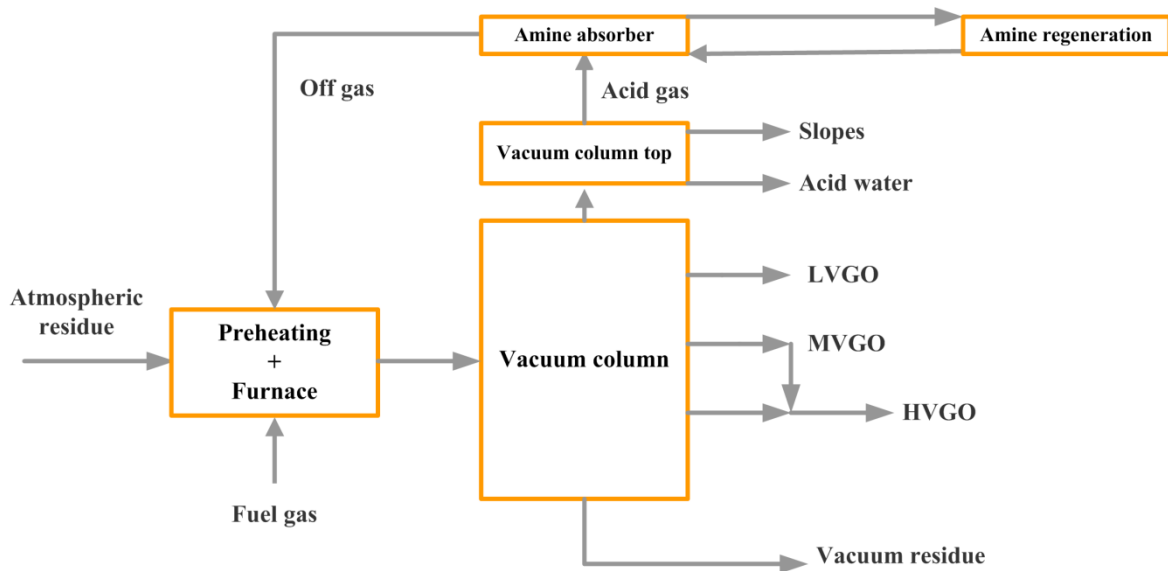


Figure 2.8: Example of a vacuum unit (Galp Energia, 2011a).

### 2.3.2. Visbreaker Unit

The visbreaking is an example of a thermal cracking process, with moderate severity, which contributes to the reduction of the pour point of a waxy residue and to the reduction of the viscosity of the residue from the vacuum distillation unit and other similar fractions for values that allows its use as a fuel oil component (Singh *et al.*, 1990; Speight, 2006). Besides residual fuel oil, gas, naphtha and gas oil are also formed with the thermal decomposition of hydrocarbons (Jones *et al.*, 2006; Joshi *et al.*, 2008).

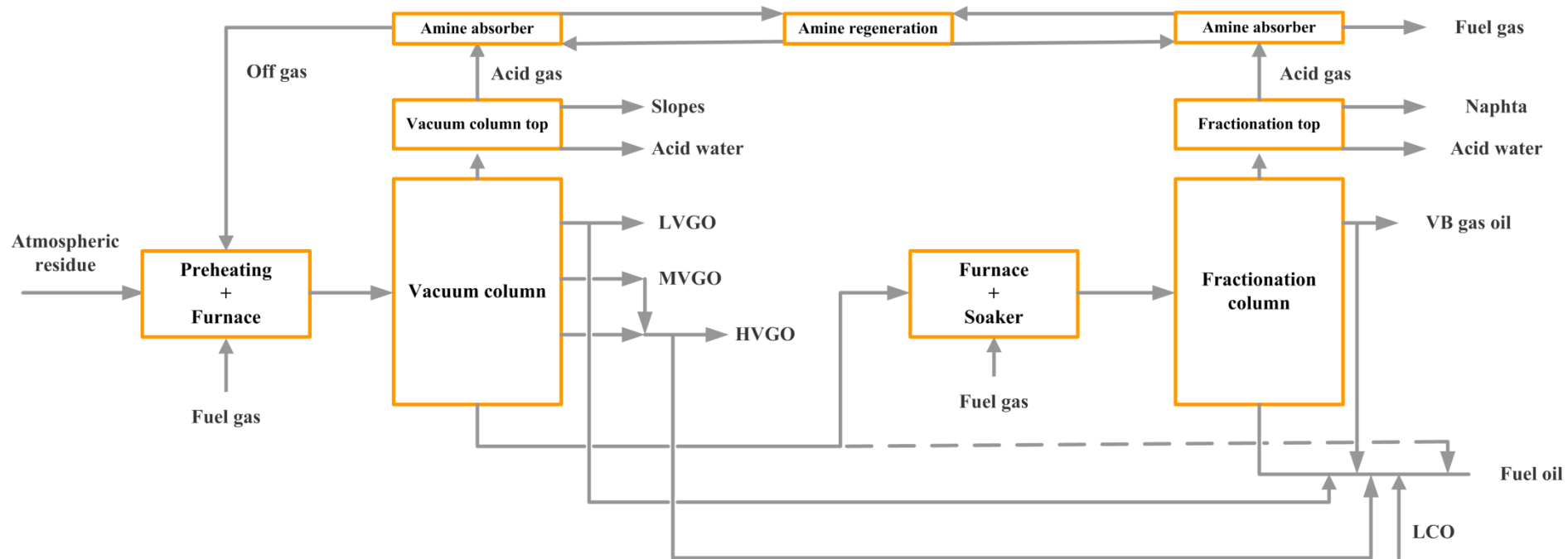
A typical visbreaker unit includes a furnace, a reactor (for example a “soaker” reactor) and a fractionation column (Meyers, 2004) (see Figure 2.9). The feed, the residue from the atmospheric or vacuum distillation units, is preheated and partially cracked in the visbreaker unit. Then, the feed is sent to the reaction section which is constituted by a heater and a soaking drum. In the heater a mild thermal cracking process occurs where the severity can be controlled by the temperature level. In the soaking drum, a reaction time is provided to obtain the desired conversion. Normally an extra residence time is provided with the intention to have a heater operating at lower temperatures, thereby saving on fuel consumption. However, it is important to have in attention that with these conditions a decoking operation may occur, here the solution is to have the appropriate equipment or a good practice in coke removal and handling. Afterwards, the effluent of the reaction section is sent to the fractionation column where the separation of the products occurs, taking into account its boiling point. Naphtha, gas, gas oil and residue are the products obtained in the visbreaker unit (Galp Energia, 2011a). Sometimes, when the idea is to reduce the quantity of lighter cutter stock which is mixed with the visbroken residue to meet viscosity specifications, instead of removing the gas oil as a product it can be integrated with the visbroken residue. Thereby, naphtha, gas and residue are the only products obtained from this unit. The visbroken residue is used to produce the fuel oil (Simanzhenkov *et al.*, 2003).

The visbreaker severity is determined by the stability of the fuel oil produced from the blending of cutter stock into the visbroken residue. The fuel oil stability can give an idea relatively to the blending of fuel oils without the occurrence of asphaltenes precipitation (Dente *et al.*, 1997). The stability and compatibility of the visbroken residue can be affected by the chemical nature of the residue from the vacuum distillation unit (Somov *et al.*, 1999). An aromatic residue from the vacuum distillation unit is preferred to a paraffinic one, once the aromatic ones have high solvent power and contribute to the stability of the visbroken residue.

During visbreaking process there are some reactions that may occur, namely, cracking reactions, dehydrogenation, cyclization, aromatization, condensation, among others (Meyers, 2004). With these reactions: i) the paraffin chain suffer C-C splitting reactions, which contributes to the reduction in their carbon number, normally from 50 to 30 carbons

in saturates, and in the formation of olefinic molecules; ii) some aromatics and resins compounds go through dealkylation reactions suffering a reduction in chain length and in the quantity of naphthenic rings, while other compounds undergo condensation reactions resulting in carbon number increasing; iii) it is possible that resins undergo substitutive addition reactions but when the residence time is shorter, in visbreaking, these reactions can be ignored; iv) asphaltenes undergo dealkylation and condensation reactions; and v) the quantity of corrosive compounds (carboxylic and naphthenic acids) is reduced due to the existence of decarboxylation reactions (Dente *et al.*, 1997; Joshi *et al.*, 2008; Kulkarni *et al.*, 2010; Zhang *et al.*, 2010). Summarizing, during these reactions there is a reduction in the saturate and resin contents, an increase in the aromatic content and an increase in asphaltene content (Somov *et al.*, 1999).

The reactions, previously referred, strongly depend on the main operative variables of the visbreaker unit as the temperature, pressure and the residence time. Normally small pressures and higher temperatures contribute to the formation of low molecular weight compounds, olefinic hydrocarbons with two to four carbons, due to the favoured reactions in gaseous phase. When using higher pressures and moderate temperatures the thermal decomposition is smaller and in these cases the formation of products presenting a distillation similar to the distillation of gasoline and gas oil is predominant. The existence of higher residence time contributes to the formation of coke due to the occurrence of condensation and polymerizations reactions. Moreover, using higher residence time and smaller temperatures also contributes to higher yields of coke. Relatively low temperatures and moderate residence times are very important to reduce coke formation, but essential to have the desired conversion and a unit working for a desired time (Speight, 2006). The most favourable is to have a visbreaker unit working with an adequate severity, this means, good conditions that do not compromise the stability of the visbroken residue, do not contribute to the formation and deposition of coke in pipelines and equipments and to obtain a significant amount of distillates.



**Figure 2.9:** Example of a vacuum – visbreaker unit (Galp Energia, 2011a).

## 2.4. Crude Oil Derivatives

Between all products obtained from the vacuum and visbreaker units, LVGO, HVGO and fuel oil are considered in this project (see Figure 2.10). Since fuel oil is a final product it is very important to control its characteristics and quality. Producing fuel oil with the correct specifications and without having product loss is one of the goals of the refinery. The characterization of LVGO and HVGO is considered very important to control the right operation of the vacuum column. Since the residue from the vacuum distillation unit cannot be sampled at the Matosinhos refinery, the analysis of these two streams will be very important to understand and control the unit operation.

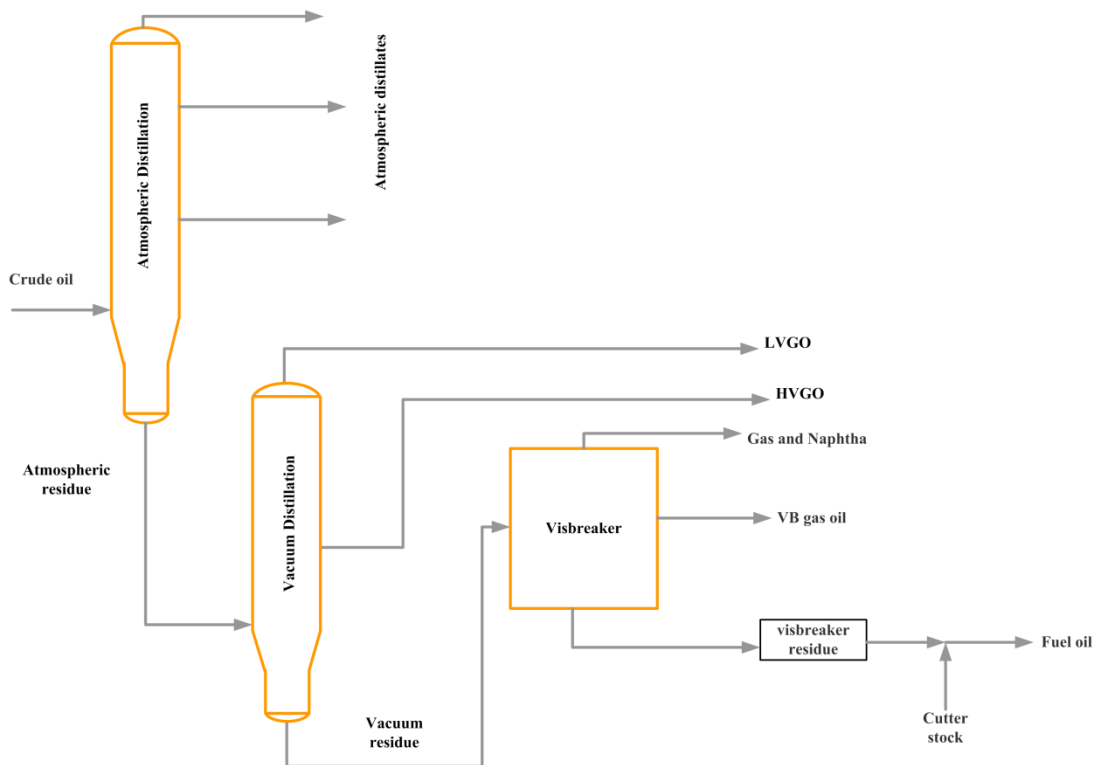


Figure 2.10: Flowsheet of the vacuum and visbreaker unit of Matosinhos refinery.



### 2.4.1. Vacuum Gas Oil (LVGO and HVGO)

Vacuum gas oil produced by conversion of residue from the atmospheric distillation unit includes light vacuum gas oil (LVGO), medium vacuum gas oil (MVGO) and heavy vacuum gas oil (HVGO). Vacuum gas oil has a boiling point over 345 °C with the LVGO presenting a boiling point of 345 to 400 °C, the MVGO of 400 to 430 °C and the HVGO of 430 to 540 °C (Simanzhenkov *et al.*, 2003). Normally, vacuum gas oil is used as feedstock to the catalytic cracking unit to be converted to a lower boiling point fractions as jet and diesel fuels (Kapur *et al.*, 2005; Speight, 2006).

**Light vacuum gas oils** is a mixture of 40-49 wt% of saturate compounds, typically 50 wt% of aromatic compounds and 1-10 wt% of polar compounds. The saturate compounds with a carbon number typically from C15 to C30 are mainly paraffins and alkylnaphthenes presenting 1 to 4 rings. The aromatic ones are essentially: i) alkylbenzenes, naphthalenes and phenanthrenes with or without naphthenic rings; and ii) aromatics with substantial amounts of heterocompounds with sulfur. Relatively to the polar compounds, they are referred as compounds presenting nitrogen or/and oxygen. LVGO is characterized as more aromatic, when compared with atmospheric gas oil, due to have a higher molecular weight and higher concentration of heteroatoms (Altgelt *et al.*, 1994).

**Heavy vacuum gas oils** presents a higher molecular weight and higher heteroatom content compared to LVGO. Normally, HVGO has a carbon number typically from C20 to C50, meaning higher carbon number and longer alkyl chains. HVGO is a mixture of 32 wt% of saturates, 58 wt% of aromatics, 10 wt% of polar compounds and small percentages of other derivatives. It contains: i) aromatic and polar compounds with 1-3 aromatic rings as well as naphthenic rings; ii) heterocompounds types with more aromatic rings; iii) heterocompounds bearing several functional groups, such as amides, sulfoxides, carboxylic acids and hydroxycarboxylic acids; and iv) some other nitrogen and oxygen compounds. Once again, comparing with LVGO, HVGO has an increase in aromatics and polar content and a reduction in saturate compounds (Altgelt *et al.*, 1994).

### 2.4.2. Fuel Oil

Fuel oils (also called No. 6 fuel oil, Bunker C oil or residual fuel oil) are produced from the distillation residues and/or visbroken residues by blending cutter stocks (Speight, 2006). Besides the residual fuel oil, there is another type of fuel oil, the distillate fuel oil. This type of fuel oil, which it is vaporized and condensed during a distillation process, is characterized for having a definite boiling point range and for the absence of high boiling point oils or asphaltic components. However, these two terms are no longer used and were replaced by terms referring to the use of fuel oil such as domestic fuel oils, diesel fuel oils and heavy fuel oils, once the production of fuel oils have now in attention specific uses and can be distillates or/and residuals. The domestic fuel oils, the most used at home, includes kerosene, stove oil and furnace fuel oil. Such fuel oils are example of distillate fuel oils. The diesel fuel oils, other example of distillate fuel oils, includes diesel oil for diesel compression ignition and light heating oil for industrial and commercial uses. Heavy fuel oils include all residual fuel oils blended with distillates such as some industrial oils and the bunker oils used to fuel ships (Speight, 2002; Speight, 2006).

Fuel oils are normally used for heating and processing in some power plants, marine vessels and industrial furnaces (Simanzhenkov *et al.*, 2003; Nielsen *et al.*, 2008), and in steam generation providing mechanical energy that can be used for electrical power generation and propulsion (Speight, 2002). They are characterized for presenting a very complex composition of different hydrocarbons rich in paraffinic, naphthenic, olefinic, and polycyclic hydrocarbons as asphaltenes and for presenting a higher quantity of impurities, compared with distillate fuels. It is possible to find in their composition other compounds as sulfur, nitrogen and heavy metals such as nickel and vanadium (Simanzhenkov *et al.*, 2003; Nielsen *et al.*, 2008). One of the problems associated with fuel oils is normally related with the specifications that it needed to respect, as example, sulfur content due to environmental implications, the viscosity due to transportation and storage and the carbon residue due to carbon formation and deposition under thermal conditions (Speight, 2002; Simanzhenkov *et al.*, 2003). The difficulty is to find out what are the best cutter stock that can be blended with the visbroken residue to obtain a homogeneous, stable residual fuel oil that respects the specifications. There are different cutter stocks that can be used, e.g. kerosene, light cycle oil (LCO), visbroken gas oil and LVGO, but it is important to have in

attention its nature. An aromatic cutter stock is preferred over a paraffinic one that decreases fuel oil stability.

### **2.4.3. Physical-Chemical Characterization of Petroleum Fractions**

As previously mentioned during the conversion of petroleum into petroleum fractions, LVGO, HVGO and fuel oil, there are many processes and reactions occurring. In detail, there are some large molecules breaking into smaller ones, there are heteroatoms that have to be removed and there are the hydrogenation of aromatic compounds and cracking reactions of naphthenes into paraffins and isoparaffins occurring. Concluding, there is a need to routine control composition of these heavy fractions to obtain knowledge of constituent components, information about the molecular level composition of the product and its effect on performance properties. This kind of information can improve refine operations, even contribute to the preservation of precious resources, minimize energy and contribute to the reduction of pollution problems. Besides the difficulty associated with the composition analysis, due to the complex and heterogeneous mixture that characterized these fractions, the potential benefits obtained are even greater. The main limitation associated with these heavy fractions, as previously mentioned, are the complexity of the mixture characterized for having thousands of compounds, hydrocarbons and heterocompounds, ranging in carbon number and relative quantity. For this reason, it is expected that each heavy fraction presents different compositions, and in consequence, different physical-chemical properties. Density, kinematic viscosity, pour point, carbon residue, sulfur content, among others, are some examples of properties that are important to describe, analyse and evaluate (Wauquier, 1995). Normally these properties are obtained taking into account ASTM methods and other traditional tests (Table 2.3).

**Table 2.3:** ASTM methods and other tests used in the characterization of petroleum fractions.

Petroleum Fraction	Properties	ASTM method
LVGO	Density (g/cm <sup>3</sup> )	ASTM D1298
	Sulfur content (% m/m)	IP 336
	Kinematic viscosity (cSt)	ASTM D445
	ASTM distillation (°C)	ASTM D86
	Flash point (°C)	ASTM D93
HVGO	Density (g/cm <sup>3</sup> )	ASTM D1298
	Sulfur content (% m/m)	IP 336
	Kinematic viscosity (cSt)	ASTM D445
	Carbon residue (% m/m)	ASTM D4530
	Vanadium	ASTM D1548
	Nickel	IP 285
	ASTM distillation (°C)	ASTM D1160
Flash point (°C)	ASTM D93	
Fuel Oil	Density (g/cm <sup>3</sup> )	ASTM D1298
	Kinematic viscosity (cSt)	ASTM D445
	Sulfur content (% m/m)	IP 336
	Flash point (°C)	ASTM D93
	P-value	SMS 1600
	Carbon residue (% m/m)	ASTM D4530

### 1. Kinematic viscosity

The kinematic viscosity in fluid, expressed in centistokes (cSt), is a measure of its resistance to flow. The kinematic viscosity depends on the temperature and when the temperature increases the kinematic viscosity decreases. The decrease in kinematic viscosity, when the temperature increases, may be related to the energy that the fluid molecules have to create a hole in the liquid contributing to their movement in or out of the liquid. The kinematic viscosity also depends on the composition of the sample in analysis, e.g. depends on the heteroatom content, the aromatic ring condensation and the interactions that may occur. As example, a paraffinic fraction normally is characterized for having low viscosity due to the lower interactions occurring even in fractions with a high boiling point (Altgelt *et al.*, 1994; Simanzhenkov *et al.*, 2003). The analysis of the kinematic viscosity data can give important information relatively to fuel transfer, storage and other that contributes to an efficient combustion (Speight, 2002).

## **2. Density**

Density is a very important property used when it is desirable to make conversions between mass (weight) and volume measurements. This property, very easy to measure, can be a good indicator of fuel quality and, moreover, it can be correlated with aromaticity, naphthenicity and paraffinity. On the other hand, this propriety is applied in product control, e.g. this is not a property that is used to evaluate the performance of the product but it is used in the product control and in the weight-volume relationship. Normally, this property, expressed in  $\text{g/cm}^3$ , is determined at standard temperatures, 15 or 20 °C (Speight, 2002).

## **3. Carbon residue**

Carbon residue is defined as an indicator of coking propensity. It provides information relatively to the tendency of hydrocarbons that have difficult combustion characteristics, e.g. the amount of coke or carbonaceous type deposits that can be formed during exposure to thermal effects. Normally there is an increase in the amount of deposits during combustion process in cases where the quantity of carbon is higher and consequently its burning is hard. There also exist a correlation between carbon residue and H/C ratio. The smaller the H/C ratio the higher the percentage of carbon residue (Simanzhenkov *et al.*, 2003).

## **4. Flash point**

Flash point can be defined as the temperature needed for the vapors of a fuel ignite when a test flame is applied. This property gives important information relatively to safety during storage, transportation and contamination by more volatile products. The presence of contamination normally leads to a lower flash point than expected. For fuel oil, the flash point has to be more than 60 °C (Speight, 2002).

### **5. Sulfur content**

Sulfur content normally is related with crude oil origin and with the refining process. During combustion process and in the presence of an excess of air, temperature and pressure, sulfur is converted into sulfur oxides as sulfur dioxide ( $\text{SO}_2$ ) and sulfur trioxide ( $\text{SO}_3$ ). These sulfur oxides in the presence of humidity are responsible for the formation of sulfurous acid ( $\text{H}_2\text{SO}_3$ ) and sulfuric acid ( $\text{H}_2\text{SO}_4$ ). However, these acids are undesirable due to corrosion problems. For these reasons, the use of fuels with high sulfur content can contribute to the development of corrosion problems as equipment damages, air pollution and acid water due to the emission of  $\text{SO}_x$  in the atmosphere, contamination of the catalyst and are toxic (Speight, 2002; Simanzhenkov *et al.*, 2003). These problems were the reasons for the development of new processes and technologies to decrease the sulfur content and sulfur emissions.

### **6. Metal content**

Normally, vanadium and nickel are the most commonly found in some heavy fuel oil molecules such as in the asphaltenes. With combustion processes it is possible to find some vanadiumoxides. The existences of double oxides/sulfates with sodium are the most critical problem. Currently, there are some additives and treatments that can be used to prevent the existence of these compounds in high quantities (Simanzhenkov *et al.*, 2003).

### **7. Pour point**

Pour point is described as the lowest temperature at which an oil will continue to flow when it is cooled under specified standard conditions. In the indication of the pour point of an oil, it is very important to know the temperature that is achieved to transfer the oil from one place to another and even in storage conditions. When the oil is kept at a temperature below of its pour point it is possible that wax crystallizes. The presence of wax can lead to the obstruction of filters and can be deposited on heat exchangers. However, even when the pour point specification is met, it is possible that wax settling occur. For this reason and

when there are large changes in temperature it is very important to pay attention to diesels with a higher heavier n-paraffins content (Speight, 2002; Simanzhenkov *et al.*, 2003).

### **8. Peptization value (P-value)**

The P-value method describes a way for quantifying the intrinsic stability of the asphaltenes in a fuel oil. The stability of asphaltenes depends on its own nature and on its chemical environment such as on the aromatic content and on the molecular mass of the asphaltenes. Incompatibility and instability of fuel oils usually results from the precipitation of the asphaltenes. Asphaltenes are the molecules in crude oil containing higher concentrations of detrimental compounds, such as, vanadium, nickel, sulfur and nitrogen. They have a negative effect in the production, transportation and oil processing, and consequently the elimination of these molecules is one of the main objectives of the refineries. This method is not only useful to analyse the stability/instability between fuel oils, as it is a way to analyse the effect of the addition of cutter stocks to visbroken residues in order to reduce the viscosity. An aromatic cutter stock that contributes to keep the asphaltenes dispersed provides a good stability reserve.

### **9. Distillation**

The distillation is one of the most important separation processes considered as an alternative way to determine boiling ranges of refined products using smaller samples and smaller operation times. It consists in the separation of individual hydrocarbons taking into account differences in their boiling point (Altgelt *et al.*, 1994). Distillation characteristics of hydrocarbons give important information relatively to their safety and performance, especially in the case of fuels and solvents. Information on the boiling point range is very useful to understand the composition and behaviour of the samples during storage and use, while the rate of evaporation plays an important role in the application of different solvents (ISO Standards 3405, 2011).

## **2.5. Final Remarks**

The implementation of new processes to increase the production of products with high commercial value and to take advantage of the heaviest part of crude oil, the residuum, contribute to the formation of vacuum gas oil (HVGO and LVGO) and fuel oil. The vacuum gas oil is used as feedstock to other process units to be converted to a lower boiling point fraction, as jet and diesel fuels, while the fuel oil is a final product used for heating and processing in some power plants, marine vessels and industrial furnaces and in steam generation providing mechanical energy that can be used for electrical power generation and propulsion. The characterization of these heavy fractions is a challenge due to its complex mixture rich in thousands of compounds. Furthermore, the knowledge of such composition can be very useful in controlling the process and in the production of final products with desirable characteristics. For such characterization different ASTM methods and other traditional tests are used.



# 3

## CHROMATOGRAPHIC AND SPECTROSCOPIC ANALYSIS OF PETROLEUM FRACTIONS WITH EMPHASIS IN NMR<sup>2</sup>

The characterization of crude oil fractions can be done by different techniques, such as gas chromatography (GC), high performance liquid chromatography (HPLC), thin layer chromatography (TLC), infrared (IR) spectroscopy, Raman spectroscopy, nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS). Nuclear magnetic resonance spectroscopy is the chosen technique due to its capability to provide information in the chemical nature of individual types of hydrogen and carbon atoms in different and complex mixtures of crude oils. The progress made in the interpretation of the NMR spectra with the development of new NMR techniques can give relevant information about the identification and characterization of hydrocarbons and their physical-chemical properties. These progresses can improve the refining industries operation as a result of better knowledge on the petroleum composition that is fed in the refining process, as well as in the prediction of better operating conditions to obtain refined products with desired specifications and in quantities desirable to meet the markets demands. The improvement in the refining operation conditions are reflected in economic benefits.

---

<sup>2</sup> The contents of this chapter were adapted from Silva, S.L., Silva, A.M.S., Ribeiro, J.C., Martins, F.G., Da Silva, F.A., Silva, C.M., **2011**. Chromatographic and spectroscopic analysis of heavy crude oil mixtures with emphasis in nuclear magnetic resonance spectroscopy: A review. *Analytica Chimica Acta* 707 (1–2), 18-37.

### **3. CHROMATOGRAPHIC AND SPECTROSCOPIC ANALYSIS OF PETROLEUM FRACTIONS WITH EMPHASIS IN NMR**

3.1. Chromatographic and Spectroscopy Characterization of Petroleum Fractions.....	45
3.1.1. Gas Chromatography (GC).....	45
3.1.2. High Performance Liquid Chromatography (HPLC) .....	47
3.1.3. Thin Layer Chromatography (TLC) .....	48
3.1.4. Infrared (IR) Spectroscopy .....	49
3.1.5. Raman Spectroscopy.....	51
3.1.6. Mass Spectrometry (MS) .....	52
3.1.7. Nuclear Magnetic Resonance (NMR) Spectroscopy .....	54
3.2. NMR Spectroscopy - Overview .....	56
3.2.1. $^1\text{H}$ NMR Spectroscopy.....	65
3.2.2. $^{13}\text{C}$ NMR Spectroscopy.....	67
3.2.3. NMR Spectrometer .....	68
3.3. Applications.....	70
3.3.1. $^1\text{H}$ NMR Spectroscopy in the Analysis of Petroleum Fractions .....	70
3.3.2. $^{13}\text{C}$ NMR Spectroscopy in the Analysis of Petroleum Fractions.....	74
3.3.3. $^1\text{H}$ and $^{13}\text{C}$ NMR Spectroscopy in the Analysis of Petroleum Samples .....	82
3.3.4. NMR Spectroscopy in Asphaltenes Characterization.....	86
3.3.5. Spectral editing NMR Techniques.....	89
3.3.6. NMR Relaxation Effects.....	98
3.4. Final Remarks.....	99

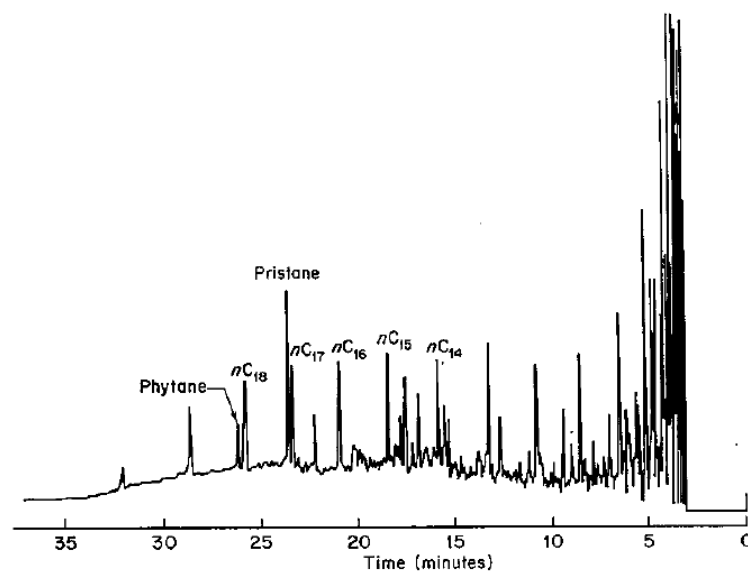
### **3.1. Chromatographic and Spectroscopy Characterization of Petroleum Fractions**

There are different methods classifying the chemical compounds present in petroleum which allows to draw some conclusions concerning their properties. Until now most of the methods used to analyse petroleum fractions are chromatographic methods, like gas chromatography (Beens *et al.*, 2000), high-performance liquid chromatography (Boduszynski, 1988), thin-layer chromatography, and some spectroscopic techniques like infrared (Sastry *et al.*, 1996), Raman (Ahmadjian *et al.*, 1976), and NMR spectroscopy and also mass spectrometry (Boduszynski, 1987).

#### **3.1.1. Gas Chromatography (GC)**

GC is used in the separation and analysis of complex mixtures of many components that can be vaporized without decomposition. In the GC technique the sample is carried through the column by the moving phase (a gas). The rate taken by the chemical constituents of the sample to pass through the column depends on their physical-chemical properties and on the interaction with the stationary phase (liquid or solid) of the column. The time taken for each compound to leave the column is called the retention time. The quantity of separated substances which coming out from the column are detected and represented by an electronic signal. These signals are detected by some detectors as the flame ionization detector (FIDr), flame photometric detector (FPD) and the thermal conductivity detector (TCD), having different sensitivity (smallest quantity of compounds in analysis) and selectivity (type of compound). FIDr is very sensitive and selective just for some compounds, used in the analysis of organic substances as benzene in gasoline according to the ASTM D3606 method (Simanzhenkov *et al.*, 2003). The FPD is used in the identification of organic compounds containing sulfur or phosphorus as heavily biodegraded spill samples (Butt *et al.*, 1986). TCD is not very sensitive but very selective in detecting everything (Speight, 2001) while FIDr is preferred in the analyses of crude oil fractions or products (Simanzhenkov *et al.*, 2003). GC studies, using TCD detector, have proven that the mixture components are not destroyed during the analysis and just a small amount of the material is needed, while in the case of FIDr and FPD detectors the

compounds are destroyed (Kenkel, 1994). Beyond FIDr, FPD and TCD there are other detectors, as electron trap detector and nitrogen/phosphorous thermo ionic detector, which can be used depending on the exactly compound under investigation. In spite of the advantages that can be achieved with the GC technique, due to the short time needed to make an analysis and the required small amount of sample, it is possible to conclude that this method is not widely used for the analysis of heavy crude fractions since all compounds analysed by GC must be in the gaseous phase and need to have a boiling temperature less than 350 °C (Altgelt *et al.*, 1994; Simanzhenkov *et al.*, 2003). The problem is that only the light fractions, as gasoline, kerosene and gas oil, have a boiling point smaller than 350 °C. For the heaviest ones, as residue, the distillation temperature is much higher, over 350 °C. This method is rather used when in presence of light (Figure 3.1) and middle distillates (de Andrade *et al.*, 2010). Moreover, it is also applied to obtain the true boiling point distribution of distillates and crude oils. This simulated distillation has been developed to simulate the time-consuming true boiling point distillation relating the retention time with the distillation temperatures of the hydrocarbon components. Example of two standardized methods of simulated distillation are the ASTM D2887 for the analysis of gaseous boiling range (ASTM Standards D2887, 2013) and the ASTM D5307 for the determination of boiling range distribution of crude petroleum (ASTM Standards D5307, 2007). Other examples on the application of GC for the crude oil characterization are the ASTM D2163 for the determination of hydrocarbons in liquefied petroleum gases and propane mixtures (ASTM Standards D2163, 2007), ASTM D2427 method to determine the composition of gasoline (ASTM Standards D2427, 2011) and ASTM D3606 method for the aviation gasoline (ASTM Standards D3606, 2010).



**Figure 3.1:** Example of a GC chromatogram of a light Nigerian crude (fraction with a boiling temperature smaller than 343 °C). Reprinted with permission from Butt *et al.* (1986). Copyright John Wiley and Sons.

### 3.1.2. High Performance Liquid Chromatography (HPLC)

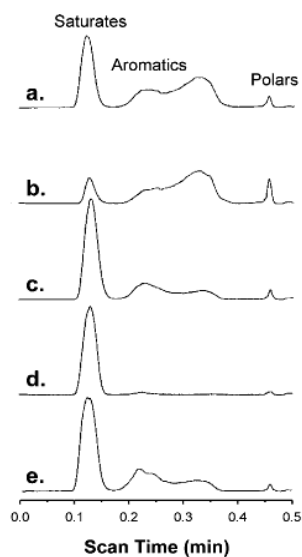
HPLC is characterized by using high pressure to force chemical compounds to pass the column (metal tube) containing a stationary phase. This column has very thin particles that contribute to an efficient separation. HPLC is an important tool for the analysis of compounds that do not present enough volatility to be analysed by GC. Besides the similarities between HPLC and GC, there are compounds analysed by HPLC but not by GC. A crude with 80% of heavy fractions needs temperatures above 350 °C to vaporise, thus is more efficiently analysed with HPLC than with GC. The possibility to obtain a good and efficient separation of heavy oil fractions and a very precise analysis in a very short time are the main advantages of HPLC. The main disadvantage of HPLC is that depending on the used detector [e.g. UV absorption (UV), refractive index (RI) or fluorescence (F)] there are some compounds which are impossible to distinguish. UV and F spectrophotometries only detect some species, F only detects fluorescent species while RI has the ability to analyse all compounds. However, the RI detectors are less sensitive than the UV detectors and strongly dependent on the temperature of the sample (Kenkel, 1994). HPLC is a chromatography technique scarcely used in the analysis of crude oils since the objective is using one technique that provides precise analysis of every component in a very short time, like minutes, if possible. When the objective is analysing

group of compounds, such as paraffinic, naphthenic and aromatic compounds, HPLC continues to be a possibility, but to analyse individual compounds of crude oil the use of HPLC is inadequate. Besides that, HPLC does not give precise results when analysing hydrotreated and hydrocracked compounds, which present identical boiling point range. However, HPLC was used in the identification of molecular types in non-volatile feedstocks, in the study of asphaltene fractions aiming to identify molecular species (Speight, 2001) and in the identification of aromatic groups (mono, di, tri- aromatics) (Pasadakis *et al.*, 2001). HPLC is very important in fingerprinting oils and used in the identification of vanadyl compounds (Fish *et al.*, 1984).

### **3.1.3. Thin Layer Chromatography (TLC)**

TLC like the other chromatographic techniques, GC and HPLC, also has a stationary and a mobile phase. Here the stationary phase is silica or alumina placed on a glass, aluminium or plastic plate unlike the GC and HPLC where this phase is located in the column. The mobile phase is a solvent or a mixture of solvents that are needed to help in the separation of compounds. Hexane, tetrachloromethane, benzene, dichloromethane, chloroform, ethyl acetate, dioxane, acetone, isobutanol, isopropanol, ethanol and methanol are example of some used solvents. The use of mixture of solvents is desirable to ensure the properties needed for the chromatographic separation. TLC is an analytical method that can be used with other chromatographic and spectroscopic techniques to provide more information and analytical data on the isolated compounds. The TLC is used in the separation of mixtures and is proven to be faster and to provide better separations. It is also a simpler and an inexpensive technique, when compared with the HPLC and GC. It is recommended in the analysis of some organic substances. However, TLC is more often used in qualitative analysis, and when necessary to identify isolated substances and the structure of the compound other precise methods are required. Up to now, precise quantitative analysis were not obtained by TLC (Simanzhenkov *et al.*, 2003). TLC is used in the analysis of heavy petroleum fractions, applied in the characterization of semi-volatile and non-volatile products and in the separation of compounds, such as asphaltenes, which present a high boiling point, proving to be simpler and quicker. Advances in TLC as TLC with flame-ionization detection (TLC-FIDr) have been used when the purpose was to make a

distinction between hydrocarbon groups, such as saturates, aromatics and polars ones, in heavy oil as shown in Figure 3.2. Although, TLC presents some limitations as the impossibility to distinguish between diesel and jet fuel, two very similar products (Speight, 2006) and to give precise and accurate results.

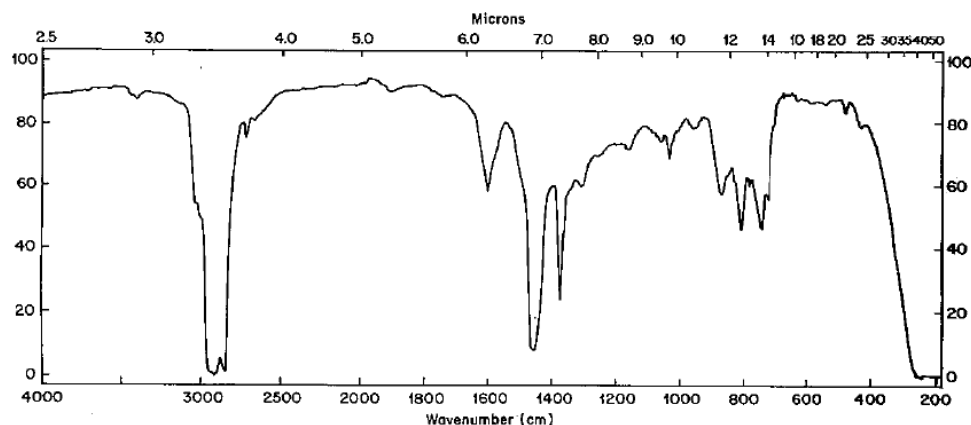


**Figure 3.2:** Example of TLC-FIDr chromatograms of 500 N VGO cut and its fractions. Chromatogram a) VGO, b) extract, c) raffinate, d) wax and e) base oil. Reprinted with permission from Barman (2005). Copyright 2005 American Chemical Society.

### 3.1.4. Infrared (IR) Spectroscopy

IR spectroscopy is one of the most important techniques that can provide miscellaneous information of complex mixtures of compounds, such as information about hydrocarbon skeleton and functional groups (e.g. hydroxyl and carbonyl groups). It allows measuring a great number of structural parameters like paraffinic and naphthenic character, aromatic hydrocarbons and methyl group content and gives information about functional features of various petroleum constituents, nature of polymethylene chains and the nature of polynuclear aromatic systems. It also contributes to the aging determination of oils related to the oxidation of carboxylic acids (Gautam *et al.*, 1998). The main advantage of this technique is the possibility to analyse the hydrogen bonding in the crude oil mixture. It is possible to say that the IR can give more qualitative rather than quantitative information

making impossible to detect some compounds. There is more limitations associate with this technique like the overlap that occurs between frequency ranges, which can be overcome by the use of NMR spectroscopy. There are some studies in the application of IR in the analysis of crude oil fractions (as Figure 3.3), but the implementation of this technique is more related with the study of middle distillates. An example is the determination of fatty methyl esters (FAME) content in middle distillates (European Standard EN 14078, 2009) and of benzene content in motor and aviation gasoline (ASTM Standards D4053, 2009), the method of carbon type analysis in lubricating oils (FORD Standards EU-AJ 051-01, 2001) and even the identification of chemical species as contaminated species (by comparison IR spectra). Another example is the use of IR in the study of spilled oils (derivatives from crude), but only in the definition of the chemical classes to which they belong, since the use of IR in the characterization and differentiation of different classes of hydrocarbons, from different heavy products, is difficult. In this case, once again, IR should be combine with other spectroscopic techniques to predict such hydrocarbons (Butt *et al.*, 1986). On the other hand, it is possible to use multivariate statistical methods, such as the partial least squares regression (PLS) and principal component analysis (PCA) to predict properties of the sample in analysis (Aske *et al.*, 2002) and with this establish correlations, which could be an improvement for online IR. As it will be demonstrated in Chapter 5, these multivariate statistical methods were used in combination with other spectroscopic technique to predict some petroleum fractions properties.

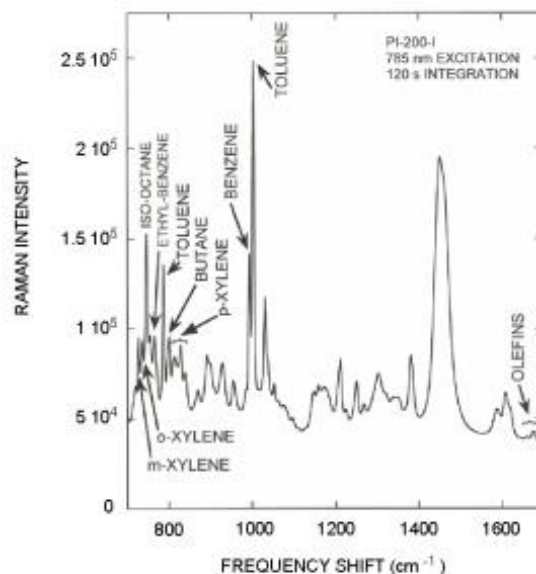


**Figure 3.3:** Example of an IR spectrum of heavy fuel oil. Reprinted with permission from Butt *et al.* (1986). Copyright John Wiley and Sons.



### 3.1.5. Raman Spectroscopy

Raman spectroscopy is used to obtain information about the vibrations of molecular bonds in the sample and to measure the energy needed to change the ground state of a chemical compound. The Raman effect occurs when there is a change in the molecular polarization, and can give more complete information than infrared spectroscopy. These two techniques are very similar in terms of results and can be used in combination to obtain more information about the analysis of petroleum samples (Simanzhenkov *et al.*, 2003). Raman spectroscopy is used in the analysis of aromatic and olefinic compounds in hydrocarbon mixtures, it does not need a sample preparation and is quicker. This technique is not considered the best choice for the characterization of crude oil, due to the weak signal obtained in the Raman spectrum. Another disadvantage of using Raman spectroscopy in the analysis of crude oil may be related with the natural fluorescence, that leads to the impossibility of identifying any signals when hydrocarbon components are irradiated by visible lasers (Simanzhenkov *et al.*, 2003). Details of these methods are mostly found when applied to light elements as gasoline (as for example registered in Figure 3.4), gas oil, jet and kerosene and also when Raman is used in combination with Fourier transform (FT). There are some studies on the use of FT-Raman spectroscopy in the determination of the octane number (RON), the motor octane number (MON), the pump octane number (pump) and the Reid vapour pressure (RVP) (Cooper *et al.*, 1995).

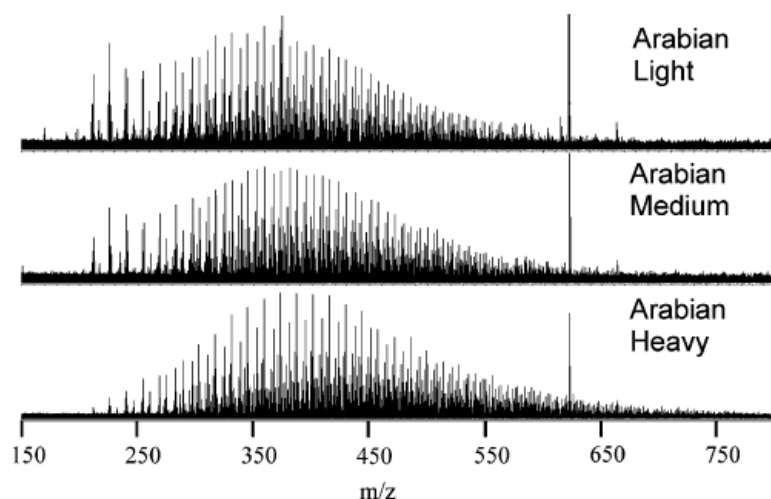


**Figure 3.4:** Example of a Raman spectrum of a gasoline sample. Reprinted with permission from Process Instruments (2001). Copyright Process Instrument, Inc..

### 3.1.6. Mass Spectrometry (MS)

MS can give information on the structure of some compounds, for example assigning their molecular formula based on the molecular weight, contributing to the identification of some compounds or to have an idea about new compounds. There are different MS techniques that can be used, such as electron impact (EI), chemical ionization (CI), field ionization (FI), fast atom bombardment (FAB), among others. EI-MS (electron impact – mass spectrometry) is a technique of major importance and is characterized by the existence of a set of electrons with high energy that will contact with the molecules and will be responsible for the ionization and fragmentation of these molecules (Kenkel, 1994). This technique gives a fragmentation pattern exhibiting both parent ion peaks and fragmentation ion peaks, characteristic of each molecular type. The fragmentation is normally used to distinguish between isomers of pure compounds and of molecules in simple mixtures. For example, in alkyl chains the most pronounced fragments are caused by the loss of  $\text{CH}_3$ ,  $\text{C}_2\text{H}_5$  or  $\text{C}_2\text{H}_4$ . Four carbons atoms are an example of a fragment that is most common in ions of paraffinic chains. Other examples of fragments are those of

having C<sub>3</sub>, C<sub>2</sub> and C<sub>5</sub>. The monoalkylbenzenes are characterized by the fragment of C<sub>6</sub>H<sub>5</sub>. Besides these examples, there are others related to monoalkylnaphthalenes, monoalkylphenanthrenes, monocyclic alkanes and other molecules (Altgelt *et al.*, 1994). Since heavy petroleum fractions are rich in thousands of compounds, it is expected that due to such complexity and closely related compounds that the fragmentation patterns became crowded and impossible to distinguish. With this, EI-MS is not frequently used in the analysis of heavy petroleum fractions. Therefore, the most common type of MS techniques used in the analysis of heavy crude oil is the non-fragmenting (NF)-MS, also called “soft ionization”. These techniques produce simpler spectra compared with those resulting from the fragmentation (Skoog *et al.*, 1997). For example, the hard fragmentation of aliphatic hydrocarbons does not follow a given pattern, making very difficult to identify these compounds. Sulfur and nitrogen compounds and species with the same chemical formula are also difficult to analyse, while other compounds like the aromatics has a good behaviour in the final spectra and are easy to identify, especially with “soft-ionization” techniques. Recently, the combination of some “soft-ionization” techniques as the low-voltage electron ionization, electrospray ionization (ESI), field desorption ionization (FDI) and atmospheric pressure photoionization (APPI) contributed to the development of FT-ICR-MS (Fourier transform ion cyclotron resonance mass spectrometry), which was used in the analysis of thousands of chemical constituents in heavy petroleum fractions. This new technique has already been used, for example, in the analysis of polycyclic aromatic sulfur heterocycles in different Arabian crude oils, as shown in Figure 3.5 (Panda *et al.*, 2007). When the boiling point increases it becomes more difficult to use the MS in sample analysis due to the increase in the number of types of compounds and the decrease in the concentration of these compounds (Behera *et al.*, 2008). Concluding, aromatic hydrocarbons, sulfur and nitrogen compounds are examples of compounds complicated to be analysed by MS, and thus, to obtain information about all heavy crude oil fractions it is necessary to use this technique conjugated with other complementary techniques. There is a possibility to use MS hyphenated techniques, like GC-MS or HPLC-MS. In other words a spectral method can be combined with chromatographic methods to exploit the advantages of both and obtain better information about the sample in analysis. Besides the advantageous that the MS hyphenated techniques can bring, normally a great diversity of standards is required, which makes it a very expensive method.



**Figure 3.5:** Example of an ESI FT-ICR mass spectra of condensed thiophenes in different Arabian crude oils. Reprinted with permission from Panda *et al.* (2007). Copyright 2005 American Chemical Society.

### 3.1.7. Nuclear Magnetic Resonance (NMR) Spectroscopy

The NMR spectroscopy has been proven to be a very important technique in the study of heavy petroleum fractions, considered increasingly as a potential alternative to some conventional methods used in laboratories. The problem with the laboratory methods, based on traditional techniques, are normally time consuming, quite elaborated and expensive if using expensive solvents and instruments (Bansal *et al.*, 1998; Molina *et al.*, 2007). Regarding these facts, and taking into account that the determination of petroleum fraction properties are very important in process monitoring and in controlling the refining operation, there is a need to use known technologies to characterize petroleum fractions and to especially contribute in the reduction of the time and cost of analysis and also to give online information.

The NMR spectroscopy has gained a prominent place in the study of the chemical composition of petroleum and in the characterisation of petroleum fractions. It has been demonstrated that its use can provide an analysis in only a few minutes per stream, which can be performed in a continuous and online way and has the ability to analyse dark and opaque samples, such as crude. It is a rapid and non-invasive procedure, offering great analyses in the chemical nature of individual types of hydrogen and carbon atoms, in different and complex mixtures of petroleum and in the final products obtained by refining

processes (Gautam *et al.*, 1998). The analysis of different types of carbon atoms and groups, like CH<sub>3</sub>, CH<sub>2</sub>, aromatic and aliphatic CH, aliphatic C and other C (Altgelt *et al.*, 1994) are possible. Besides the previously mentioned advantages there are other associated with this technique including the improvement in the analysis, reduction in the number of analysis, reduction in maintenance costs and unique ability to provide all control properties with one analysis.

As mentioned by the *Mesures Magazine* (2002) the main advantages are the direct relation between the measured signal and the total proton concentration and the superior resolution of chemical information contained in the signals. In the NMR spectra the functional groups of aromatics, aliphatics and olefins are well distinguished, in consequence it is possible to distinguish a polynuclear aromatic from a mononuclear aromatic and to identify the physical properties by the concentration of aromatics and the chain length of aliphatics. Besides these advantages there also exist commercial and operational benefits associated to the NMR spectroscopy. The commercial benefits are related to the optimization in the feed of refining processes to adjust the production taking into account the market necessity, to benefit the attractive price of some crudes and in saving money by using complete systems of acquisition analysis in a short time. The operational benefits are to provide a higher value in product yields, maximize unit recovery of potential lost, due to instability, and improve the performance of downstream units (The Qualion Company, 2006). On the other hand, there are some weaknesses associated with this spectroscopic technique like the high cost, the risk of magnetic disturbances, requiring magnetic shielding, and the overlap of frequency ranges. The overlap continues to make the analysis of the spectra a very difficult task. However improvements have been made to oppose these weaknesses as the use of some spectral editing techniques (DEPT, GASPE, COSY, etc.). Also the very complex information given by NMR spectroscopy requires a statistical approach to correlate the spectral data with the characterization of petroleum fractions. The combination of NMR spectroscopy with multivariate data methods will be discussed in Chapter 5.

It is possible to conclude that there are some restrictions associate with all chromatographic and spectroscopic techniques discussed, making it unrealizable to analyse the heavy petroleum fractions. However, the NMR has been suggested as one of the most

powerful methods for industrial analysis. Until now, NMR has been used in some refinery laboratories with the purpose to determine: i) the content of oil in paraffins and aromatic carbon contents of hydrocarbon oils by high resolution NMR (ASTM Standards D5292, 1999); ii) the aromatic hydrogen and carbon content by high resolution NMR (IP Standards 392, 1990); iii) the hydrogen content of middle distillate petroleum products by low-resolution pulsed NMR (ASTM Standards D7171, 2005); and iv) the aromatic carbon content of lubricant mineral base oils and middle distillate petroleum fractions by  $^{13}\text{C}$  NMR spectroscopy (IP Standards 499, 2011). Besides the laboratory applications, online NMR systems have been used in many refineries around the world applied on the feed and products units for control and optimization (Bakeev, 2010). Examples of some applications are: i) NMR for gasoline analysis in Australia Caltex Brisbane refinery; ii) NMR to the optimization of a crude unit in *Petróleos De Venezuela S.A.* (PDVSA) - Isla refinery; iii) NMR for monitoring the feed, intermediate and final products of a lubricant oil manufacturing plant in Livorno, Italy; and iv) most frequently in many other refineries, NMR is used in the analysis of feed and cracked products such as light cycle oil on fluid catalytic cracking units (Bakeev, 2010). Examples of other applications are related to the use of NMR well logging for real-time analysis of rock and fluid properties (Hirasaki *et al.*, 2003; Ramos *et al.*, 2009). Here, instead of using the  $^1\text{H}$  or  $^{13}\text{C}$  NMR spectra, the NMR well logging measured both  $^1\text{H}$  relaxation times and diffusion coefficient. This information can be useful to distinguish between water, oil and gas content in reservoir fluids as well as to analyse the diffusion coefficients, porosity and estimate the oil viscosity, establishing differences between light to extra heavy crude oils. Crude oils are characterized for having broad relaxation time and diffusivity distributions. The ones presenting lower relaxation time are normally associated to a higher viscosity (Hirasaki *et al.*, 2003; Ramos *et al.*, 2009).

### **3.2. NMR Spectroscopy - Overview**

The NMR spectroscopy is a method based on the magnetic properties of nuclei and their interactions with applied magnetic fields (Gil *et al.*, 2002; Bakeev, 2010). The information given by the NMR spectroscopy, for each type of observed nuclei, depends either on the electronic environment where the nuclei are immersed and on the positions of the nuclei in

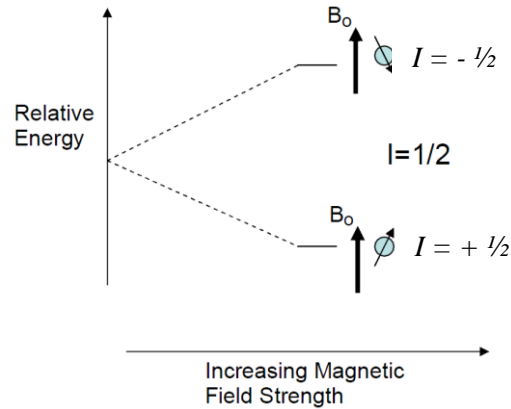
the molecules. Therefore, the NMR data is very important to understand the molecular structure of the sample as well as the internal and global dynamics of the molecules in the sample (Gil *et al.*, 2002).

The nuclei observable by NMR are those possessing nuclear spin. Protons and neutrons, quantum mechanically subatomic particles, have spin. When the number of protons and neutrons is equal, it means that when in some nuclei these spins are paired and cancel each other out, the nucleus of the atom has no overall spin. Nuclei with an even atomic number and even mass number, such as the isotopes  $^{12}\text{C}$ ,  $^{16}\text{O}$  and  $^{32}\text{S}$ , are example of some nuclei with a spin number equal to zero. However, the number of protons and neutrons is not always equal and in many cases the sum of the number of protons and neutrons is an odd number, thus contributing to nuclei with half-integer spins. On the other hand, if the number of protons and the number of neutrons are both odd, then it gives rise to what is called the integer spin. Thus, the spin quantum number,  $I$ , can be  $1/2$ ,  $1$ ,  $3/2$  and  $5/2$ , this is, it can be multiples of  $1/2$ . The proton ( $^1\text{H}$ ) and carbon ( $^{13}\text{C}$ ) are examples of nuclei with a spin  $I = 1/2$  (Pavia *et al.*, 1996; Gil *et al.*, 2002; Bakeev, 2010). When the nuclei are spinning they possess angular momentum,  $V$ , and a nuclear charge which gives rise to an associated magnetic moment,  $\mu$  (Equation 3.1), such as:

$$\mu = \gamma V \tag{3.1}$$

where  $\gamma$  presents the magnetogyric ratio, a constant for particular nuclei that represents the strength of the magnetic moment. When the nuclei, without preferred orientation, are placed under the influence of an external magnetic field of strength  $B_0$ , applied in a defined direction, an interaction, commonly called the Zeeman interaction, between the magnetic moment of the nucleus and the applied magnetic field occurs (Becker, 2000). Consequently, the magnetic moments will align themselves relative to the external magnetic field into  $2I + 1$  energy states. Thus, the proton ( $^1\text{H}$ ) and all other nuclei having a spin of  $I = 1/2$ , when placed in a magnetic field, will have two allowed energy states that equate to aligning with or against the applied magnetic field (Figure 3.6). However, from these two allowed energy states the nucleus can only adopt one, taking up one of the two possible orientations with respect to the applied external magnetic field, parallel

( $I = +1/2$ ) or anti-parallel ( $I = -1/2$ ) (Christian *et al.*, 1986; Gil *et al.*, 2002). The parallel orientation, the one towards the magnetic field, is more stable while the anti-parallel orientation, opposed to the applied field, is of higher energy.



**Figure 3.6:** Representation of the possible energy levels for nuclei with spin of  $I = 1/2$ . Reprinted with permission from Bakeev (2010). Copyright 2010 John Wiley and Sons.

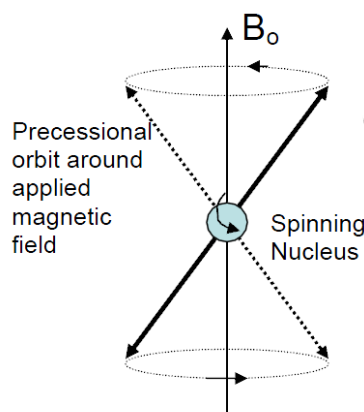
The magnetic moments cannot align directly along the axis of the applied magnetic field due to the torque imposed by the applied field on the magnetic moments. The existence of such torque contributes to the circular path of the magnetic moments around the applied field direction in a motion referred to as Larmor precession (Jacobsen, 2007). The rate of this precession depends on the magnetogyric ratio of the nucleus, which leads to the precessing of different nuclei around the applied magnetic field at an angle with different angular velocities or Larmor frequencies (Figure 3.7). The rate of the precession can be defined by the angular velocity,  $\omega_o$  ( $\text{rads}^{-1}$ ) (Equation 3.2), or frequency,  $\nu$  (Hz) (Equation 3.3), expressed as:

$$\omega_o = -\gamma B_o \quad (3.2)$$

$$\nu = \frac{\omega_o}{2\pi} \quad (3.3)$$

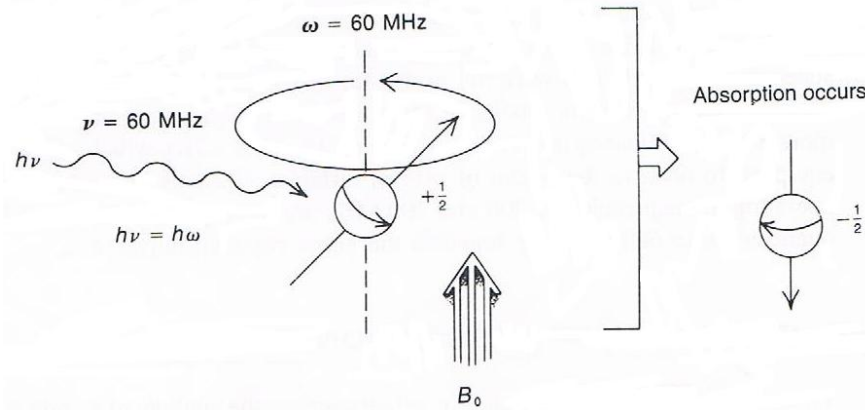


where  $\gamma$  is the strength of the nuclear magnetic moment and  $B_o$  represents the strength of the external magnetic field. Consequently, the frequency at which a proton precesses is proportional to the strength of the nuclear magnetic moment as well as to the strength of the external magnetic field. The precession frequency increases with increase of the strength of the external magnetic field (Pavia *et al.*, 1996; Gil *et al.*, 2002; Jacobsen, 2007; Bakeev, 2010).



**Figure 3.7:** Precession of a single proton around the external magnetic field. Reprinted with permission from Bakeev (2010). Copyright 2010 John Wiley and Sons.

When the precession process occurs it is generated an oscillating electric field with the same frequency  $\omega_o$ . If radiofrequency waves of this frequency are required to the precessing proton, the energy can be absorbed and the proton promoted to the less favorable higher energy state. This energy transferred from the incoming radiation to the nucleus can only be performed when the frequency of the electromagnetic radiation applied matches the frequency generated by the precessing nucleus, the Larmor precession. When the two fields, the frequency of the applied radiation and of the precession, coincide or resonate and the energy can be transferred, the nucleus changes its spin state and consequently the nuclear magnetic resonance process occurs (Figure 3.8) (Pavia *et al.*, 1996; Claridge, 1999).



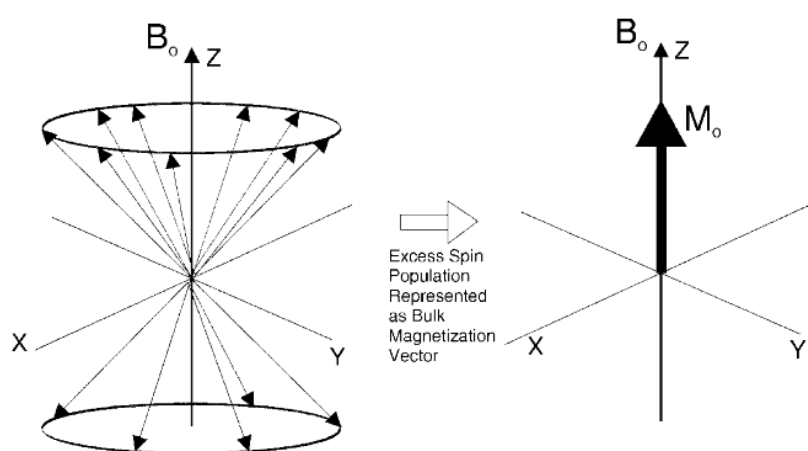
**Figure 3.8:** The nuclear magnetic resonance process (Pavia *et al.*, 1996).

When considering a collection of similar spin-half nuclei in the applied magnetic field, it is verified that, in equilibrium, the distribution of spin population between the two possible energy states, the parallel ( $\alpha$ ) and the anti-parallel ( $\beta$ ) to the applied magnetic field, is not the same. A slight excess of spin population in the parallel direction is found due to this level being energetically more favorable. The distribution of the spin population between the two energy levels is defined by the Boltzmann distribution as given by Equation 3.4 (Claridge, 1999; Becker, 2000):

$$\frac{N_{\alpha}}{N_{\beta}} = \exp\left(\frac{\Delta E}{kT}\right) \quad (3.4)$$

with  $N_{\alpha,\beta}$  representing the populations of nuclear spins in the parallel ( $N_{\alpha}$ ) and anti-parallel levels ( $N_{\beta}$ ),  $k$  the Boltzmann constant and  $T$  the absolute temperature. Since the energy separation of the spins states is comparatively small and consequently the corresponding population differences between both levels are also small, NMR is considered to be an insensitive technique relative to other techniques, such as IR and UV (Christian *et al.*, 1986). However, the insensitivity of the NMR instrument can be overcome with the increase of the operation frequency of the NMR instruments. Thus contributing to a larger separation between the energy levels and consequently to higher

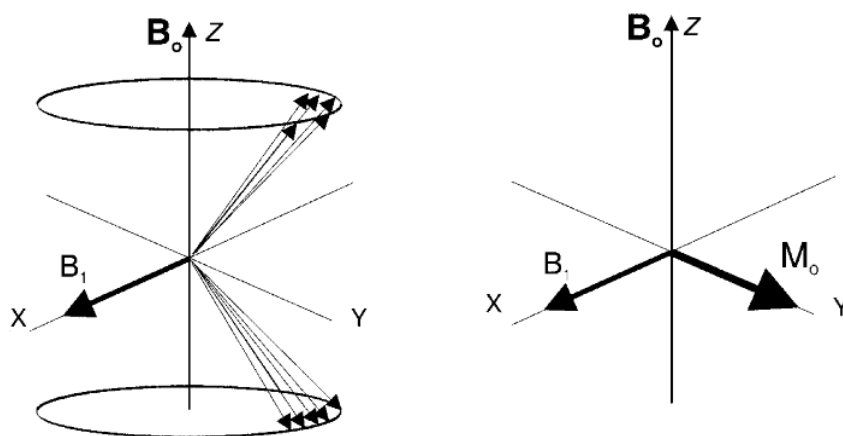
population difference. These improvements were very important especially to observe less sensitive nuclei, such as  $^{13}\text{C}$ , that is not very abundant, and to improve the signal-to-noise ratio of the spectrum (Pavia *et al.*, 1996). The small excess of nuclear spins in the parallel state is measured in NMR and is responsible for the NMR signal. The excess of population distributed randomly about the precessional cone and parallel to the Z-axis is summed together, after the opposing nuclear spins in the parallel and anti-parallel direction cancel each other out, in a bulk magnetization vector along the magnetic field  $\mathbf{M}_0$  (see Figure 3.9).



**Figure 3.9:** Formation of the bulk magnetization vector from the excess spin population. Reprinted with permission from Bakeev (2010). Copyright 2010 John Wiley and Sons.

When the sample is irradiated by an orthogonal radiofrequency field,  $\mathbf{B}_1$ , all of the magnetic nuclei in the molecule are simultaneously excited. Consequently, the spins, oscillating at the Larmor frequency, start to precess around the orthogonal  $\mathbf{B}_1$  field and, after a few microseconds, the net magnetization vector  $\mathbf{M}_0$  will rotate into the transverse (x-y) plane, where it can be electronically detected. This radiofrequency pulse is transmitted through a coil surrounding the sample, whose geometry dictates that the  $\mathbf{B}_1$  field exists in the transverse plane, perpendicular to the static magnetic field  $\mathbf{B}_0$  (Pavia *et al.*, 1996; Claridge, 1999; Bakeev, 2010). The amplitude and the duration of the radiofrequency pulse are responsible for the angle through which the net magnetization vector  $\mathbf{M}_0$  turns. Thus, the magnetization vector  $\mathbf{M}_0$  reaches the y axis after a  $90^\circ$  pulse (Figure 3.10) and consequently, no magnetization vector will be observable along de z

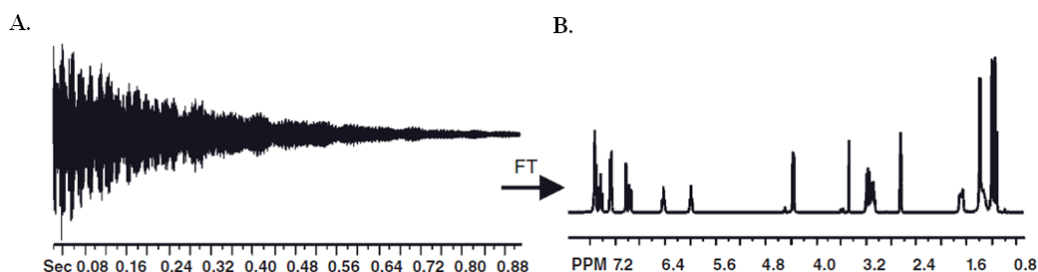
axis. With this, after the application of the  $90^\circ$  pulse, the populations of the  $\alpha$  and  $\beta$  states are equalized, since there is no magnetization along the  $z$  axis. However, the net magnetization in the  $x$ - $y$  plane results from the bunching of the individual spins, which are in a phase coherent manner, caused by the radiofrequency pulse applied. On the other hand, if the radiofrequency pulse is left on for twice as long, the magnetization vector  $\mathbf{M}_0$  becomes oriented along the  $-z$  axis, representing a  $180^\circ$  pulse. With the  $180^\circ$  pulse the populations of the spin states are inverted and more spins are found in the  $\beta$  state than in the  $\alpha$  state, to place the bulk vector anti-parallel to the static field. When the magnetization vector  $\mathbf{M}_0$  is oriented along the  $-z$  axis no signal is detected. Only magnetization vector in the  $x$ - $y$  plane is able to induce a signal in the detection coil (Claridge, 1999).



**Figure 3.10:** Representation of the net magnetization vector  $\mathbf{M}_0$  in the transverse plane after the application of the  $90^\circ$  radiofrequency pulse. Reprinted with permission from Bakeev (2010). Copyright 2010 John Wiley and Sons.

As soon as the radiofrequency pulse is turned off, the system will adjust to re-establish the Boltzman equilibrium, and the bulk magnetization vector begins to precess around  $\mathbf{B}_0$  reducing its intensity along the  $y$  axis and simultaneously increasing in intensity along the  $z$  axis toward its initial value and direction ( $\mathbf{M}_0$ ). This return to equilibrium referred to as the relaxation is responsible for the NMR signal to decay with time. The decaying signal is called as free induction decay (FID) and it is the basis for all NMR data (Claridge, 1999). The Fourier transformation (FT) of the FID, of the time domain data, is used to separate each of the individual components of the decay signal and convert them to frequencies,

giving rise to a frequency domain spectrum (common NMR spectrum) as visualized in Figure 3.11 (Pavia *et al.*, 1996).



**Figure 3.11:** Application of the Fourier transformation on the FID. (A) Time domain data, FID, (B) Frequency domain spectrum. Reprinted with permission from Bakeev (2010) Copyright 2010 John Wiley and Sons.

The relaxation process can be divided in two categories, the spin-lattice or longitudinal relaxation (relaxation time called  $T_1$ ) and the spin-spin or transverse relaxation (relaxation time called  $T_2$ ). The longitudinal relaxation (recovery of magnetization along the z axis) corresponds to the re-establishment of the equilibrium of populations and consequently to the energy exchanges occurring between the spins and the surroundings (lattice). The transverse relaxation (decay of magnetization in the x-y plane) involves energy transfer between the magnetized spins. These two relaxation process described are assumed to follow an exponential behaviour, occurring simultaneously but as completely separate phenomena. This means, any nonequilibrium longitudinal component will approach  $M_0$  exponentially with time  $T_1$  while any transverse component of the magnetization experience an exponential decay with time  $T_2$  (Christian *et al.*, 1986; Claridge, 1999; Jacobsen, 2007).

After obtaining the NMR spectra, there are some useful parameters that can be extracted in order to provide important information for molecular structure characterization. For instance in the  $^1\text{H}$  NMR spectra these parameters include: i) chemical shift ( $\delta$ , expressed in part per million, ppm); ii) spin-spin coupling pattern (multiplicity); iii) coupling constant ( $J$ ); and iv) signal intensity.

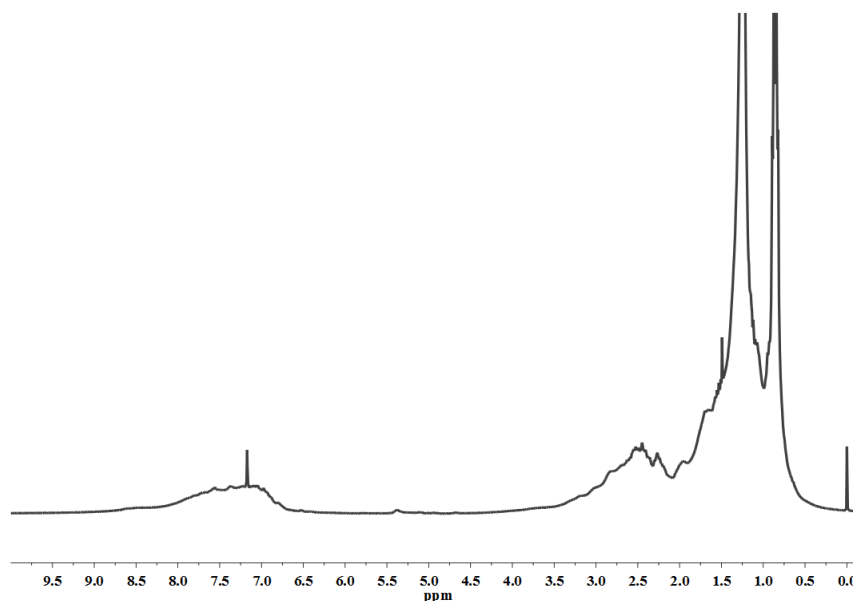
Any nucleus in a molecule has an electronic environment which depends on where the nucleus is in the molecule and what is attached to it. When the electrons, that surround the nuclei, are subject to a strong static magnetic field, the electrons are induced to circulate around the applied field. The circulation of the electrons is responsible for the production of its own magnetic field that acts in opposition to the applied field. Thus, the effective magnetic field experienced by the nucleus results from the effect of the magnetic field and the induced magnetic fields generated by the circulation of the surrounding electrons. These induced magnetic fields from the electrons are responsible for shielding the nucleus from the applied magnetic field. Since, the shielding effect depends on the electron density around the nucleus, then the effective magnetic field that each individual nuclei experience depends on the chemistry surrounding the nucleus under observation. Thus, in the case of proton observation, protons at different sites in the molecule are magnetically shielded to different extents depending on their chemical environment and location (neighbouring atoms and type of chemical bond). Consequently, each site experiences a different magnetic field and has a different position on the NMR spectrum. The position of NMR peaks with each peak corresponding to a unique chemical environment is measured by its resonance frequency (expressed in Hz) and it depends on the magnetic field strength and on the operating frequency of the spectrometer. Consequently, the position of a NMR signal vary from one spectrometer to another as the magnetic field strength changes. However, in order to allow consistent comparison of the NMR data, the obtained resonance frequencies are: i) normalized using a reference compound (set to 0.0 ppm); and ii) divided by the resonance frequency of the spectrometer (in MHz). The resulting dimensionless quantity obtained is termed chemical shift ( $\delta$ ) and expressed in parts per million (ppm). This chemical shift is independent of the operating frequency of the spectrometer (Bakeev, 2010).

The interaction that occurs between the neighbouring magnetic nuclei with a molecule is commonly referred as spin-spin coupling or scalar coupling. This interaction is responsible for the splitting of NMR signals into multiplets. The multiplicity of a signal, referred as the number of peaks in the split signal, is expressed as  $2nI + 1$ , where  $n$  represents the number of neighbouring equivalent nuclei and  $I$  the spin quantum number.

For  $^1\text{H}$  NMR spectra, the peak area (integrated peak area or integral) of the signals is directly proportional to the number of equivalent nuclei represented by that peak (Jacobsen, 2007). In this context, for quantitative NMR acquisition conditions, it is important to allow the spins to fully relax between pulses in order that the observed signal intensities will not be diminished by incomplete recovery of the magnetization between acquisitions. The pre-processing of the NMR spectra is also a fundamental requirement to enhance the quantification results. In the next chapter all the pre-processing steps used are described. Besides the  $^1\text{H}$  NMR spectra, the most commonly used for quantitative analysis of complex mixtures, quantitative  $^{13}\text{C}$  NMR may also be used in an attempt to obtain complementary structural information. Moreover, the use of different nuclides (nitrogen, sulfur and oxygen) may be a good choice when the concentration of acidic OH and basic N groups are desirable. It will be possible to identify some compounds as ethers, secondary amines, thioethers and thiophenes (Snape, 1986). The identification of ethers is a way to identify oxygen bonds that may have some impact in some processes, such as desulfurization and denitrogenation. From all nuclei presented,  $^1\text{H}$  and  $^{13}\text{C}$  are the most commonly used in the analysis of petroleum fractions. For this reason they were studied and applied in this work.

### 3.2.1. $^1\text{H}$ NMR Spectroscopy

The natural abundance of  $^1\text{H}$  is 99%, one of the main reasons which allows a proton spectrum to be obtained in a few seconds. Generally the proton resonates over a range of 15 ppm; using a reference compound such as tetramethylsilane (TMS) its chemical shift goes from 0 ppm to 15 ppm. The chemical shift of 0 ppm is assigned to TMS. An example of a  $^1\text{H}$  NMR spectrum for a fuel oil is presented in Figure 3.12.



**Figure 3.12:**  $^1\text{H}$  NMR spectrum of a fuel oil.

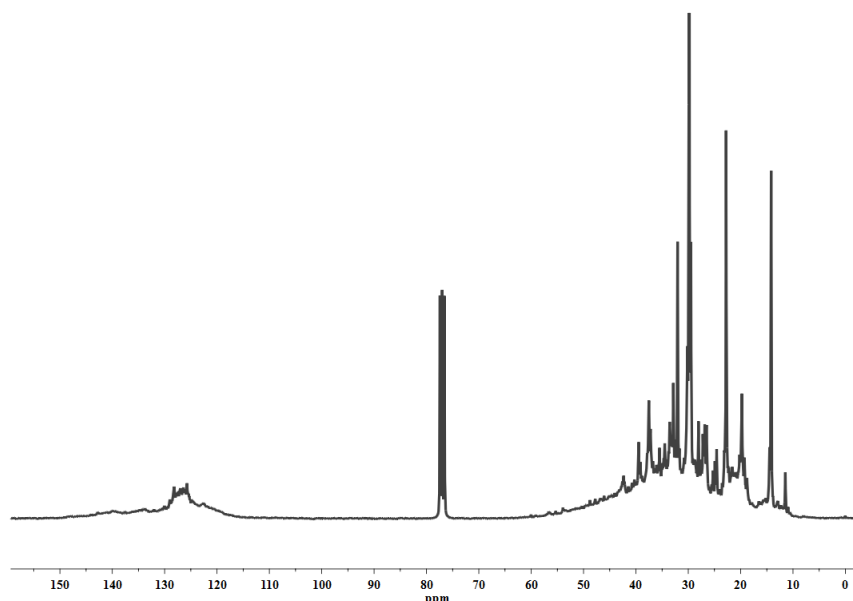
In a  $^1\text{H}$  NMR spectrum it is possible to identify different types of hydrogens. In a petroleum or even in a petroleum fraction these hydrogens can be differentiated in four different groups: i) aromatic hydrogen – hydrogens attached to carbons on aromatic rings; ii) hydrogens attached to carbons in the  $\alpha$  position on the aromatic rings ( $H_\alpha$ ); iii) hydrogens of methyl groups ( $\text{CH}_3$ ) located in the  $\gamma$  position and beyond relative to the aromatic rings ( $H_\gamma$ ); and iv) all other hydrogens –  $\text{CH}$ ,  $\text{CH}_2$  of chains in the  $\beta$  position and beyond, and the  $\text{CH}_3$  hydrogens in the  $\beta$  position ( $H_\beta$ ) (Wauquier, 1995; Speight, 2002).

The percentage of hydrogen for each one of the above groups can easily be determined due to the fact that the signal's area are directly proportional to the number of protons and related to the percentage of atoms. This is one of the main advantages of  $^1\text{H}$  NMR compared with other spectral analysis methods. With the information that can be taken from the  $^1\text{H}$  NMR spectrum it is possible to affirm that the  $^1\text{H}$  NMR spectroscopy is a technique which contributes to obtain information about the structural characteristics and to estimate the aromaticity factor, molecular weight, average length of some chains and other important parameters.



### 3.2.2. $^{13}\text{C}$ NMR Spectroscopy

$^{13}\text{C}$  resonates over about 220 ppm. The  $^{13}\text{C}$  NMR is less sensitive than  $^1\text{H}$  NMR due to the lower natural abundance (1.1%) and to the lower gyromagnetic ratio of the C-13 isotope. In addition, the resonance of  $^{13}\text{C}$  nuclei is the most difficult to observe and the carbon nuclei have a longer relaxation time comparing to the protons. Even using the Fourier transform data acquisition, the acquisition of a  $^{13}\text{C}$  NMR spectrum can be long and it can take hours to be obtained (Pavia *et al.*, 1996). This can be explained by the fact that  $^{13}\text{C}$  nuclei relaxes slower than protons, making a very long relaxation delay between repetitive pulses necessary. Moreover, the  $^{13}\text{C}$  NMR spectroscopy is characterized for presenting a very complex spectrum with high multiplicity of the signals. To eliminate these high multiplicity that results from the spin-spin splitting due to the protons attached to each carbon, the decoupling of the proton must be used. With the decoupling, where all the protons in the molecule are irradiated with continuous low power radio frequency energy at the proton resonance frequency, the  $^{13}\text{C}$  nucleus appears as a singlet independently of the number of protons attached. This occurs since with the decoupling the  $^{13}\text{C}$  nucleus only sees the average of each proton between the two energy levels. The final spectrum will be composed by a set of peaks, one for each carbon, indicating especially the number of carbons in the sample (Morrison *et al.*, 1996; Jacobsen, 2007). Figure 3.13 is an example of a decoupling  $^{13}\text{C}$  NMR spectrum of a crude oil.

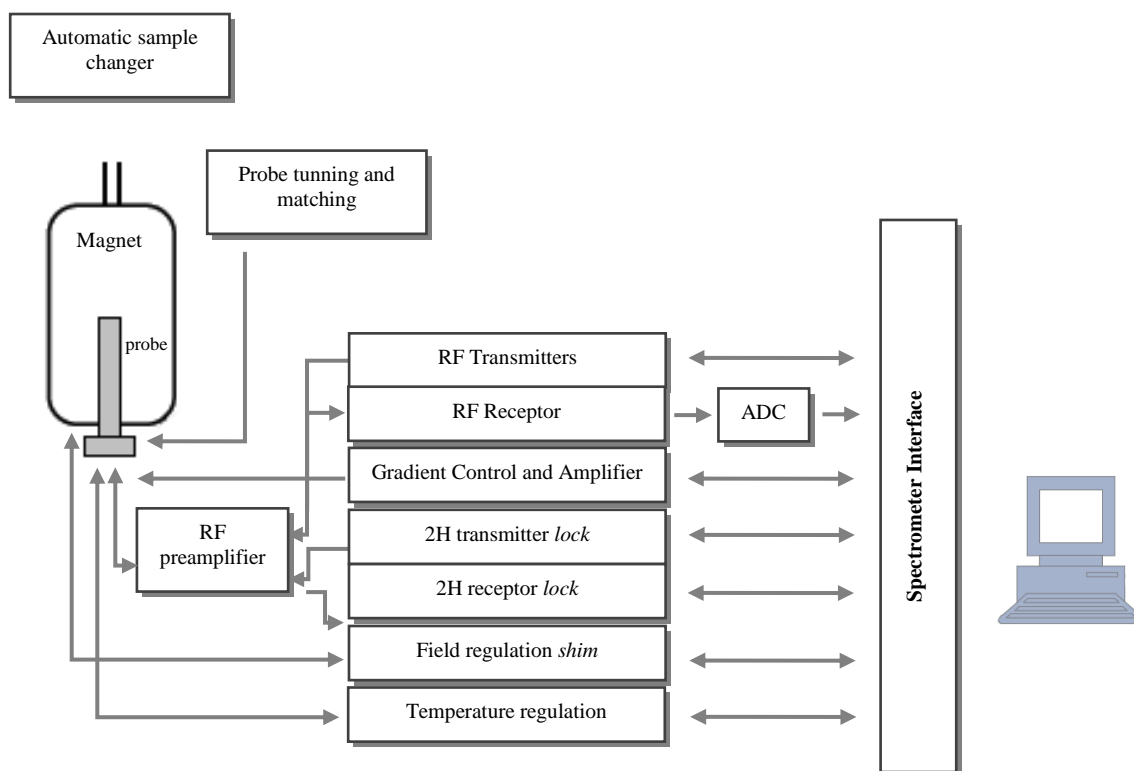


**Figure 3.13:** Inverse gate decoupling  $^{13}\text{C}$  NMR spectrum of a crude oil (Albecore Crude Oil).

The  $^{13}\text{C}$  NMR spectrum is produced basically by the same way that the  $^1\text{H}$  NMR spectrum and the same basic principles are applied. The information given by the  $^{13}\text{C}$  NMR spectra is generically similar as that of the  $^1\text{H}$  NMR spectra. The  $^{13}\text{C}$  NMR gives information related to: i) aromatic CH groups; ii) aromatic carbons substituted by an alkyl chain; and iii) condensed aromatic carbons belonging to two attached rings (Wauquier, 1995). However, the normally used  $^{13}\text{C}$  NMR spectrum is not quantitative, this is, the peak areas are not proportional to their concentration especially due to the relaxation phenomena (Christian *et al.*, 1986). It is for this reason that  $^1\text{H}$  NMR spectroscopy is preferred over the  $^{13}\text{C}$  spectroscopy in the analysis of mixtures. Although it is possible to acquire a quantitative  $^{13}\text{C}$  NMR spectrum which, especially in the case of polymers and complex mixtures, it takes an extremely long time. To obtain a quantitative  $^{13}\text{C}$  NMR spectrum, the different relaxation times of  $^{13}\text{C}$  nuclei in different chemical groups, especially for aliphatic versus aromatic C, and the nuclear Overhauser enhancement (NOE) effect, must be suppressed (Altgelt *et al.*, 1994). The NOE effect is related to the increase degree in the signal intensity, with a weak radio frequency signal, when C-H coupled protons are saturated by the decoupling field. However, the increase peak intensity is not always the same and due to such variability, the NOE must be avoided for quantitative measurements (Pavia *et al.*, 1996). To reduce the long relaxation times of some carbons a paramagnetic relaxation reagent, such as trisacetylacetonatochromium(III) [ $\text{Cr}(\text{acac})_3$ ], is needed to be added to the sample. Using this reagent the long relaxation time of some carbons is reduced and the relaxation mechanism is changed (Altgelt *et al.*, 1994). Besides all this, the  $^{13}\text{C}$  NMR spectroscopy is used to obtain information relatively to the molecular carbon skeleton and to the chemical composition of a sample, also considered as an important method for the structure elucidation.

### 3.2.3. NMR Spectrometer

Figure 3.14 represents a schematic illustration of a high resolution NMR spectrometer. By analysis of the Figure 3.14 it is possible to identify some indispensable parts that form the NMR spectrometer such as the magnet, the probe containing the sample, the radio frequency transmitter to pulse and receive the emitted response, the preamplifiers and the computer to collect, digitize and perform the Fourier transform of the signals.



**Figure 3.14:** An example of a modern NMR spectrometer. Adapted from Claridge (1999).

The superconducting magnet is one of the main reasons why NMR spectrometer is considered a complex and expensive equipment. To operate it requires to be in a bath of liquid helium, responsible for cooling the system down to 4 K, and surrounded by a bath of liquid nitrogen also responsible for cooling the instrument down to about 77 K avoiding helium evaporation. There are different types of magnets such as the permanent magnet, an electromagnet or a superconducting solenoid (“supercon”), but all of them are responsible for the generation of a very strong, stable and homogeneous magnetic field. It is in the central bore of the magnet that many shim coils (or electrical coils) are housed. These shim coils with a small magnetic field are responsible for the *shimming*, a process that contributes for optimizing the magnetic field homogeneity. Inside the magnet, the probe alternately transmits and receives radio frequency signals. The probe which is like a coil of wire positioned around the sample receives higher power and very short pulse duration of radio frequency, emitted by a transmitter. After that and to obtain a digital FID signal, the electrical analogue signal received is amplified, converted to a digital format and sampled at regular intervals of time by the computer via the analogue-to-digital converter (ADC). A

final spectrum, on the computer monitor, is obtained after a Fourier transform is applied (Christian *et al.*, 1986; Claridge, 1999; Jacobsen, 2007).

### **3.3. Applications**

Many researches using NMR spectroscopy can be found in the literature, however, the vast majority of these applications can be divided in different main headings: i) determination of structures of pure compounds; ii) quantitative determination of mixtures; iii) determination of diffusion coefficients and porosity; and iv) measurements of relaxation times. NMR spectroscopy combined with other major techniques is very useful for structural determination, while NMR spectroscopy combined with multivariate statistical methods is important for quantitative determinations.

Relatively to the analysis of petroleum fractions, few applications have been described in the literature. NMR has been already used in this field for a long time, being first published by Friedel (1959), “*Absorption Spectra and Magnetic Resonance Spectra of Asphaltene*”. Since then, some researchers have been publishing the use of NMR in the analysis of petroleum fractions. The main purpose was to help refineries to refine heavy crude oils in more efficient ways and at lower processing costs. Examples of the applications of  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectroscopy,  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra of petroleum samples, asphaltenes characterization, NMR techniques and NMR relaxation effects are going to be presented. The use of multivariate statistical methods of NMR spectra is going to be presented in another chapter (see Chapter 5).

#### **3.3.1. $^1\text{H}$ NMR Spectroscopy in the Analysis of Petroleum Fractions**

$^1\text{H}$  NMR spectroscopy was used to determine the aromatic content of petroleum products such as base oils (Sarpal *et al.*, 1998). Some methods have been developed for such purpose as the one presented by Brown and Ladner (Sarpal *et al.*, 1998). However, since these methods required the use of other techniques such as C and H elemental analysis and other parameters not directly estimated, new alternatives were required. The objective of modifying the Brown-Ladner equation (represented by the Equation 3.5) was to quantify

the factors needed to determine the aromatic content, like the elemental H and C content of the sample ( $H/C$ ), the average number of hydrogen per  $\alpha$ -alkyl substituents, the atomic  $H/C$  ratio for non-aromatic groups ( $Z_a$ ) in the sample, the aromatic protons ( $H_a$ ), among others.

$$C_a = 100 - \left[ \frac{(100 - H_a)(H/C)}{Z_a} \right] \quad (3.5)$$

Sarpal *et al.* (1998) developed a direct and quick method with the objective to create a new equation to estimate the aromatic content ( $C_a$ ) of base oils using  $^1\text{H}$  NMR spectroscopy. From the interpretation of the  $^1\text{H}$  NMR spectra and the use of two-dimensional heteronuclear correlation (2D HETCOR) technique it was possible to calculate the aromatic carbon content by changing the Brown-Ladner equation (Equation 3.5), as well as to estimate the bridgehead carbons ( $C_b$ ) and the quaternary aromatic carbons ( $C_q$ ) content. The new equation (Equation 3.6) does not need the determination of such factors but just the directly use of one and two dimensional NMR spectra to estimate the number of protonated aromatic carbons ( $C_{ah}$ ), the substituted aromatic carbons ( $C_s$ ), the contribution of bridgehead carbons ( $C_b$ ) and the total relative number of carbons ( $C_T$ ).

$$C_a = \left[ \frac{C_{ah} + C_s + C_b}{C_T} \right] \times 100 \quad (3.6)$$

This method gives similar results as the Equation 3.5 being more advantageous due to the possibility to be used in all types of base oils as well as to the quick results which are obtained.

Since the reduction of the aromatic content is one of the future objectives of the refineries it is important to quantify the total amount of aromatics in the petroleum fractions. Bansal

*et al.* (1998) started to use  $^1\text{H}$  NMR spectroscopy to estimate the content of total aromatics and their distribution as mono and polynuclear aromatics. These authors (Bansal *et al.*, 1998) started to analyse the  $^1\text{H}$  NMR spectra and the 2D HETCOR ( $^{13}\text{C}$ - $^1\text{H}$ ) NMR spectra of a representative diesel sample and of the aromatic fraction of a diesel sample, respectively. They estimated some structural parameters like the number of substituted aromatic carbons ( $C_s$ ; Equation 3.7), the average of alkyl chain length ( $n_a$ ) of aromatic substituents (Equation 3.8), the number of bridgehead carbons ( $C_b$ ; Equation 3.9), the total aromatic content of the sample ( $T_A$ ; Equation 3.10) and the estimation of mono-ring ( $M_A$ ; Equation 3.11) and global di-plus-ring aromatics ( $P_A$ ; Equation 3.12):

$$C_s = \left(\frac{B}{2}\right) + \left(\frac{C_r}{3}\right) \quad (3.7)$$

$$n_a = \frac{\left[\left(\frac{I_{3.5-2.4}}{2}\right) + \left(\frac{I_{2.0-1.0}}{2}\right) + \left(\frac{I_{1.0-0.5}}{3}\right)\right]}{\left(\frac{I_{3.5-2.4}}{2}\right)} \quad (3.8)$$

$$C_b = 0.68 I_d \quad (3.9)$$

$$T_A = \left(\frac{A_W}{T_W}\right) \times 100 \quad (3.10)$$

$$M_A = T_A - P_A \quad (3.11)$$

$$P_A = \frac{(C_b \times 100)}{21.7} \quad (3.12)$$

with  $B$  representing the 2.4-3.5 ppm region;  $C_r$  the 2.0-2.4 ppm region;  $I_x$  the integral intensity of the  $x$  ppm region;  $I_d$  the integral intensity in the 7.5-10.0 ppm region;  $A_w$  the molecular weight of the aromatics; and finally  $T_w$  the total group molecular weight. 2D HETCOR spectra were used to improve the analysis and the signals assignment in the  $^1\text{H}$  NMR spectra, due to the overlapping. It allowed to conclude that only exist  $-\text{CH}_3$  groups between 0.5-1.0 ppm and did not exist any overlap; while some  $\alpha\text{-CH}_3$  ( $\alpha$ -substitution in aromatic carbons) groups were found between 2.4-2.6 ppm. It was possible, as well, to find a relationship between the  $-\text{CH}_2$  carbons, 25-40 ppm, and the protons that exist in the 2.4-3.5 ppm region. With these data, it was concluded that some  $\alpha\text{-CH}_2$ ,  $\alpha\text{-CH}_3$  substituents on the aromatics and also some  $\alpha\text{-CH}$  protons exist in the range 2.4-3.5 ppm. In the  $^{13}\text{C}$  NMR spectra, the overlap achieved between the 18-20 ppm may be due to the methyl substituents of the aromatic rings.

From the interpretation of the  $^1\text{H}$  NMR spectra of heavy crude oils developed by Behera *et al.* (2008), Bansal *et al.* (1998), and Ali *et al.* (2005) it is possible to assign a range of chemical shifts to different types of protons (Table 3.1).

**Table 3.1:** Chemical shift regions of various types of protons (solvent: deuteriochloroform).

	Chemical shift range (ppm)	Type of protons
<b>Aromatic region</b>	6.0-9.0	Total aromatic proton
	6.0-7.2	Mono-aromatic proton
	7.2-7.6	Di-aromatic molecules
	7.2-8.0 or > 8.0	Poly-aromatics molecules
<b>Aliphatic region</b>	2.05-4.5	H- $\alpha$ to aromatic ring
	1.1-2.05	H- $\beta$ to aromatic ring in paraffinic CH and CH <sub>2</sub>
	0.4-1.1	H- $\gamma$ to aromatic ring/terminal CH <sub>3</sub>
	1.3	CH <sub>2</sub> in long alkyl chains
	0.5-4.5	Total aliphatic proton
	0.9	CH <sub>3</sub> in long alkyl chains
	2.1	CH <sub>2</sub> in aromatic rings
	2.6	CH <sub>3</sub> in aromatic rings

### 3.3.2. $^{13}\text{C}$ NMR Spectroscopy in the Analysis of Petroleum Fractions

Yoshida *et al.* (1980) developed a work on some characteristic compounds in the  $^{13}\text{C}$  NMR spectra of the ring-type fractions of coal-derived oil aiming to determine the chemical shift ranges of protonated, bridgehead and substituted carbons of this type of sample and with that determine the equation for structural parameters. From the results of the Yoshida work (Yoshida *et al.*, 1980) it was established that usually aromatic carbons appear in the order of protonated, bridgehead and substituted carbons from low to high frequency values. Normally, the chemical shifts of the protonated carbon appear at 115.0-129.2 ppm, the bridgehead carbons at 129.2-132.5 ppm, and the substituted carbons at 132.5-149.2 ppm. However, the presence of the phenolic OH and the amino groups influence the chemical shift of the neighbor protonated carbon shifting them to lower frequency values. Some of the bridgehead carbons were extended to a field lower than 132.5 ppm and others continued to appear in the range of 129.2-132.5 ppm, depending if these carbons are naphthalenic and methylated or just naphthenonaphthalenic, respectively. On the other hand, the chemical shift of the substituted carbons depends on the type of aliphatic carbons bonded to aromatic rings. In the subsequent  $^{13}\text{C}$  NMR spectra analysis it was possible to assign some signals to chemical shifts of model compounds, for example, the signals at 132.5-137.2 ppm were due to substituted aromatic carbons bonded to methyl groups or to cycloparaffinic  $\alpha\text{-CH}_2$ . The signals at 137.2-149.2 ppm corresponded to the resonance of substituted aromatic carbons bonded to some alkyl groups (not include methyl and methylene bridge between aromatic rings) and that at 149.2-158 ppm to aromatic carbon attached to OH groups (phenolic OH). Other important aspects established in the Yoshida work (Yoshida *et al.*, 1980) was that the coal-derived oil contain a lot of alkyl substituents and naphthenic rings. It was also possible to reach this conclusion through the analysis of some structural parameters like the aromaticity ( $f_a$ ; Equation 3.13), the ratio of non-bridgehead aromatic carbon to total carbon ( $H_{au}/C_a$ ; Equation 3.14), and the degree of aromatic ring substitution ( $\sigma$ ; Equation 3.15). These structural parameters were defined by the following equations:



$$f_a = \frac{C_a}{C_T} \quad (3.13)$$

$$\frac{Hau}{Ca} = \frac{Ca - C_b}{Ca} \quad (3.14)$$

$$\sigma = \frac{C_s}{Ca} - C_b \quad (3.15)$$

with  $C_a$  indicating the amount of the aromatic carbon,  $C_T$  the amount of total carbon,  $C_b$  the amount of the bridgehead carbon, and  $C_s$  the amount of substituted carbon.

Cookson and Smith (1985) studied a total of nine diesel samples, six of which were petroleum and three of them were synfuels (fuels from nonpetroleum sources), with two kerosene samples, aiming to obtain a better understanding on the differences in the composition, structural characteristics, and in the properties of each fuel using the  $^{13}\text{C}$  NMR spectroscopy. They have used the gated spin echo technique (GASPE) with the objective to minimize the overlap of resonances due to C, CH,  $\text{CH}_2$  and  $\text{CH}_3$  groups that usually occurred in a  $^{13}\text{C}$  NMR spectrum of complex mixtures. From the analysis of the resonances appearing in the spectra, they could identify a possible structure corresponding to an n-alkane formalized as  $(\alpha\text{CH}_3)_2(\beta\text{CH}_2)_2(\gamma\text{CH}_2)_2(\delta\text{CH}_2)_2(\varepsilon\text{CH}_2)_y$ , and therefore calculate some parameters as the  $N_C$  (average n-alkane chain length; Equation 3.16),  $BS$  (average number of branching sites per molecule; Equation 3.17),  $N_B$  (average number of branches per molecule; Equation 3.18) and  $N_R$  (average number of rings per molecule; Equation 3.19). It was also demonstrated that all the diesel samples have an average chain length of 15-16 carbons while the kerosene examined has about 12 carbons chain length. However, firstly it is necessary to estimate the average number of carbons per molecule ( $NC$ ) that depends on the sample used.  $NC$  is not always equal to the average n-alkane chain length ( $N_C$ ), once n-alkanes, branched saturated and cyclic saturated structures may have different average number of carbon atoms. Besides  $NC$ , it is also required to

determine the total  $^{13}\text{C}$  NMR intensity due to the  $\text{CH}_2$  and  $\text{CH}_3$  groups and to know the C, CH and  $\text{CH}_3$  fractions of carbon atoms present as C, CH and  $\text{CH}_3$  groups, respectively:

$$N_c = \frac{3 I(\text{CH}_2) + 2 I(\text{CH}_3)}{I(\text{CH}_3)} \quad (3.16)$$

$$BS = NC(C + CH) \quad (3.17)$$

$$N_B = NC(2C + CH) \quad (3.18)$$

$$N_R = 0.5 \times NC(C + CH - \text{CH}_3) + 1 \quad (3.19)$$

The same authors, Cookson and Smith (1990), compared 50 jet and diesel fuels using  $^{13}\text{C}$  NMR spectroscopy, to understand some fuel properties and to define its composition (like aromaticity information as well as n-alkane abundance). The main objective of this analysis was to provide better information about the composition of this type of samples needed for the preparation of better quality refining products. Using earlier published works (Cookson *et al.*, 1985; Cookson *et al.*, 1987), Cookson and Smith (1990) developed a new approach (Equation 3.20) that listed a fraction relation of the total  $^{13}\text{C}$  NMR spectral intensity due to a long-chain n-alkyl carbon ( $C_n$ ) atoms with the fraction of the total  $^{13}\text{C}$  NMR spectral intensity due to aromatic carbon ( $C_a$ ) atoms, such as the following one:

$$P = b_1 C_n + b_2 C_a + c \quad (3.20)$$

In Equation 3.20,  $b_1$ ,  $b_2$  and  $c$  were determined by a multiple linear regression and  $P$  was the property value. Cookson and Smith (1990) only considered the  $^{13}\text{C}$  NMR spectroscopy

to determine the samples composition, since it was considered a good technique to provide aromatic carbons identification. They had also proven that this approach was more efficient than other complex equations, better to distinguish a large quantity of fuels from different sources and available to determine their properties.

Sarpal *et al.* (1997) began to develop some experiments with some base oils aiming to estimate the content of some structural parameters, like the n-paraffin (*NP*) and isoparaffin (*IP*) contents, the number of branching sites (*BS*) and the average alkyl chain length (*C\**). These base oils, produced from different treatments like hydrocracking and wax isomerisation, were hydrofinished-solvent-refined oils (HF) of high viscosity index (HVI), several hydrofinished (SHF) oils of very high viscosity index (VHVI) and hydrotreated base oils (HT). It was verified that although the similarity between the types of branched structures produced by the hydrocracking and wax isomerisation, there were some differences, especially in the obtained quantities. For example, the HT base oils presented more branched structures bearing methyl groups located at the fifth (or higher) carbon away from the end of the chain, but it did not have the contribution of some branched structures, like the S-8, and had a greater amount of S-3, S-4 and S-5 structures (Table 3.2, page 80). The HF and the SHF have more S-8 type branched structures and reduced amounts of S-3, S-4 and S-5 (Table 3.2). From the analysis of the <sup>13</sup>C NMR spectra, Sarpal *et al.* (1997) could estimate the percentage of n-paraffin (Equation 3.21) and isoparaffin (Equation 3.22) contents by using the following expressions:

$$NP = 100 \times \left[ \frac{3I_{31.9} + I_{29.5-30.7}}{I_T} \right] \quad (3.21)$$

$$\frac{NP}{IP} = \frac{I_{31.9}}{I_{10.0-15.0} - I_{31.9}} \quad (3.22)$$

where,  $I_T$  represents the total integral intensity in the region 5-160 ppm and  $I_x$  the integral area at the indicated  $x$  ppm. Sarpal *et al.* (1997) concluded that the HT and SHF base oils

presented an IP of 45-60%, whereas the HF base oils presented a smaller amount of isoparaffins (30-40%). As it was intended, other structural parameters were also calculated such as the  $C^*$  (Equation 3.23),  $BS$  (Equation 3.24),  $t - Me$  (percentage of terminal methyl; Equation 3.25) and the  $b - Me$  (percentage of branched methyl; Equation 3.26). However, to determine such parameters, the integral area at the indicated ppm ( $I_x$ ) and the percentage of branched paraffinic carbons ( $C_p$ ) were required.

$$C^* = 2 \left( \frac{C_p}{t - Me} \right) \quad (3.23)$$

$$BS = \frac{b - Me \times C^*}{IP} \quad (3.24)$$

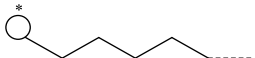
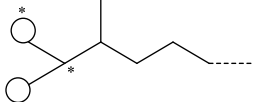
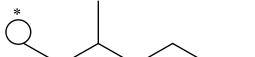
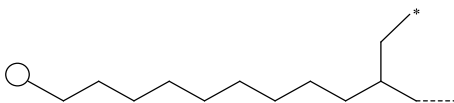
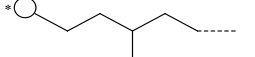
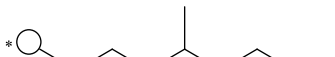
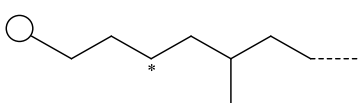
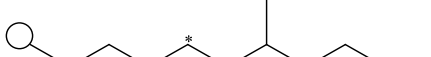
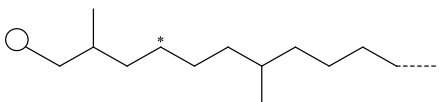
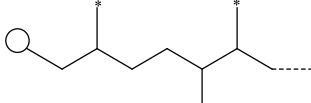
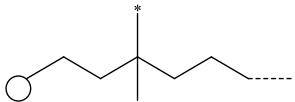
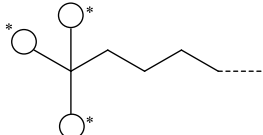
$$t - Me = 100 \times \left( \frac{I_{10.0-15.0}}{I_T} \right) \quad (3.25)$$

$$b - Me = 100 \times \left( \frac{I_{15.0-21.0} + I_{28.2}}{I_T} \right) \quad (3.26)$$

From these results it was possible to conclude that the HT and SHF base oils have a less average number of branching sites ( $BS$ ) (3.0-3.5) compared with the HF base oil (3.5-5.5). It was also found that the HT base oils presented more branched paraffinic structures than branching near terminal methyl groups (S-2, S-3 and S-4) compared with other oils (Table 3.2). It also had more predominance of S-6 and S-7 and less quantity of S-8 structures (Table 3.2). These last structures were found in a higher percentage in the HF and SHF oils. HT oils presented high quantity of branched structures, high quantity of isoparaffins, low quantity of  $BS$ , a lower pour point and a higher viscosity index. These results allowed to conclude that oils having structures with higher quantity of branching near the end of the

chain present a lower pour point. It was also concluded that the high viscosity index (HVI) and very high viscosity index (VHVI) base oils presented a higher quantity of branched structures near (S-2, S-3) and far (S-6, S-7) from the terminal carbon (Table 3.2). The VHVI oils had high quantity of NP and more S-6 and S-7 structures (Table 3.2) that contribute to the higher viscosity index. The LVI base oils presented a higher number of branching sites, a higher pour point and a lower viscosity index.

**Table 3.2:** Some branched structures identified in base oils and their characteristic  $^{13}\text{C}$  NMR chemical shifts. P=–CH<sub>3</sub>, S=–CH<sub>2</sub> and \*=carbons for which chemical shifts are reported (Sarpal *et al.*, 1997).

Structure	Identification	$\delta$ (ppm)
	S-1	14.0 (P)
	S-2	22.7 (P), 28.2 (S)**
	S-3	11.4 (P)
	S-3'	10.7 (P)
	S-4	14.2 (P)
	S-5	14.3 (P)
	S-6	27.0 (S)
	S-7	27.0 (S)
	S-8	24.0-25.6 (S)
	S-9	16.2 (P)
	S-10	28.6 (P), 27.4 (P)
	S-11	30.8 (P)

\*\*As the authors referred, the contribution of iso-methyl carbons of S-2 may be explained from the signal at 28.2 ppm corresponding to a –CH carbon.

Very recently, Verdier *et al.* (2009) used the  $^{13}\text{C}$  NMR data of a broad variety of samples produced from different vacuum gas oils to analyse the molecules with high viscosity indexes (VI). They concluded that oils presenting molecules with long alkyl chains, low aromatic contents, ethyl branching and tertiary carbons presented high VI. Theoretically, the VI has been considered an important way to predict the influence of the temperature on the viscosity, since the viscosity of oils with a high VI are slightly influenced by temperature. However, these authors (Verdier *et al.*, 2009) also indicate that it is not always correct to correlate the viscosity index with the influence of the temperature on the oils viscosity, being also important to analyse the effect of the activation energy. They had also announced that it is preferred to use the flow activation energy rather than the viscosity index to analyse the influence of the temperature on the oils viscosity.

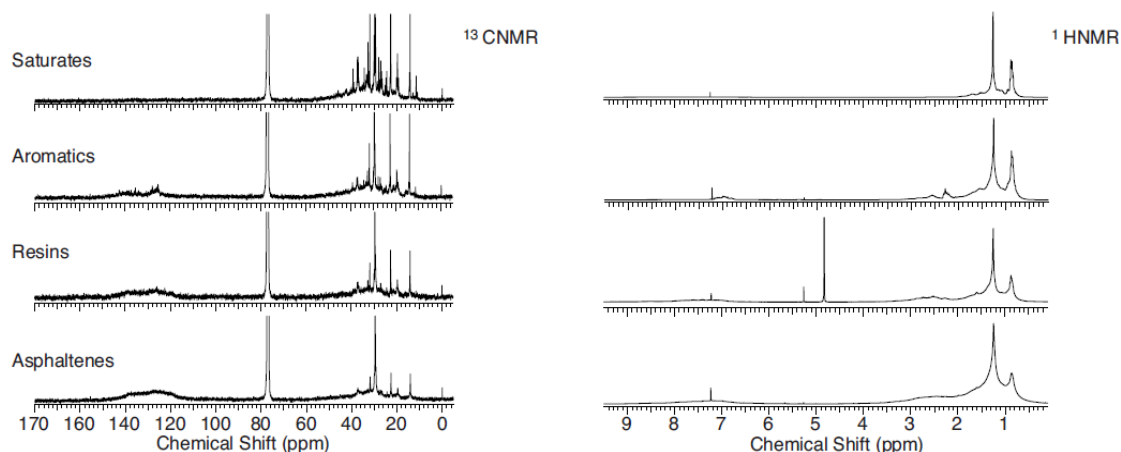
Based on the results obtained by Yoshida *et al.* (1980) and Behera *et al.* (2008) it was possible to assign a range of chemical shifts to resonances of different carbon signals in the  $^{13}\text{C}$  NMR spectra of crude oils as are shown in Table 3.3.

**Table 3.3:** Assignment of  $^{13}\text{C}$  NMR chemical shifts to different types of carbon atoms (solvent: deuteriochloroform) (Yoshida *et al.*, 1980; Behera *et al.*, 2008).

Chemical shift range (ppm)	Type of carbon atoms
202.0-220.0	Ketone carbonyl carbon
182.0-170.0	Acid carboxyl carbon
165.0-175.0	Ester or amide carboxyl carbon
100.0-160.0	Total aromatic carbon
149.2-158.0	Aromatic carbons attached to heteroatoms
137.2-149.2	Substituted carbons bonded to some alkyl groups (not include methyl and methylene bridge between aromatic rings)
132.5-137.2	Substituted carbons bonded to methyl group or to cycloparaffinic $\alpha\text{-CH}_2$
129.2-132.5	Bridgehead aromatic carbons
115.0-129.2	Protonated aromatic carbon
128.5-160.0	Quaternary aromatic carbons without triple-bridged aromatic carbons
100.0-128.5	Tertiary aromatic and triple-bridged aromatic carbons
50.0-10.0	Total aliphatic carbon
25.0-60.0	Naphthenic carbon
29.1-31.5	Carbons in alkyl chains ( $n > 6$ )
27.6-28.6	Carbons in branching position of a terminal <i>iso</i> -propyl group
17.6-20.4	Carbons in $\text{CH}_3$ branches
13.7-15.5	Carbons in terminal position of n-alkyl chains ( $n > 6$ )
0.0-20.5	Total carbon in $\text{CH}_3$ groups

### 3.3.3. $^1\text{H}$ and $^{13}\text{C}$ NMR Spectroscopy in the Analysis of Petroleum Samples

Woods *et al.* (2008) used NMR spectroscopy to characterize the SARA fractions (saturated, aromatics, resins and asphaltenes) separated from different Canadian crude oils. The assignment of different types of proton and carbon atoms in the  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra (Figure 3.15) of SARA fractions led the authors (Woods *et al.*, 2008) to conclude that the aromaticity region (110-150 ppm) increases from saturated to asphaltenes fractions. This was confirmed by the  $H/C$  ratio decreasing with the increasing of aromaticity. On the other hand, the saturated fraction was characterized by an aliphatic nature represented with a high  $H/C$  ratio. It was also concluded that the  $H/C$  ratio decreases from saturated > aromatics > resins > asphaltenes; becoming the asphaltenes rich in long chain aliphatic compounds as can see in Figure 3.15.



**Figure 3.15:** Typical  $^{13}\text{C}$  and  $^1\text{H}$  NMR spectra for SARA fractions, obtained in a Varian Unity Inova 400 MHz spectrometer with a 5 mm 4 nucleus probe. Reprinted with permission from Woods *et al.* (2008). Copyright Oil & Gas Science and Technology-Revue d'IFP Energies nouvelles.

Cookson and Smith (1987) studied the monoaromatic fractions obtained from petroleum, coal and shale-derived distillate fuels with the purpose to get new structural information. With the obtained data it was possible to analyse some average structure parameters, such as  $NC$  (average number of carbon atoms per molecule; Equation 3.27),  $N(\text{CH}_n)$  (average number of  $\text{CH}_n$  groups; Equation 3.28),  $N_S$  (number side chains per molecule; Equation 3.29),  $L_S$  (average side chain length; Equation 3.30),  $N_{SB}$  (number of side chain branches



per molecule; Equation 3.31),  $N_R$  (number of rings; Equation 3.32),  $H_{ar}$  ( $^1\text{H}$  NMR aromaticity; Equation 3.33):

$$NC = \frac{6}{f_a} \quad (3.27)$$

$$N(CH_n) = [f(CH_n)]NC \quad (3.28)$$

$$N_S = N(C^{ar}) \quad (3.29)$$

$$L_S = \frac{(1 - f_a)NC}{N_S} \quad (3.30)$$

$$N_{SB} = N(CH^{al}) + 2N(C^{al}) \quad (3.31)$$

$$N_R = 0.5 \left( N_S + N(CH^{al}) + 2N(C^{al}) - N(CH_3^{al}) \right) \quad (3.32)$$

$$H_{ar} = \frac{f(CH^{ar})}{\sum_{n=0}^3 [nf(CH_n)]} \quad (3.33)$$

where  $f_a$  represents the fraction of  $^{13}\text{C}$  NMR intensity in the aromatic region;  $f(CH_n)$  the fraction of  $^{13}\text{C}$  NMR intensity due to the various  $\text{CH}_n$  groups;  $N(C^{ar})$  the number of nonprotonated aromatic carbon atoms;  $N(CH^{al})$  number of aliphatic CH groups;  $N(CH_3^{al})$  number of aliphatic  $\text{CH}_3$  groups; and  $N(C^{al})$  number of aliphatic C groups. With the obtained results the authors (Cookson *et al.*, 1987) could then concluded that all analysed

diesel fuels had an average of 13-15 carbon atoms per molecule whereas the kerosene aromatics had a lower value (11-12 carbon atoms). Other similarities between all the seven diesel fractions were seen, such as the same number of side chains per molecule ( $N_S = 3$ ), the same average side chain length ( $L_S \approx 2.8$ ) (in only six of the samples) and the absence of quaternary carbon branching sites. One of the main differences between the samples was related to the number of ring structures involving aliphatic carbon atoms. The number of rings ( $N_R$ ) observed for a specific fraction depends of the original sample; when a fraction is a product of hydrotreatment it has a large quantity of saturated ring structures. Kerosene samples compared with the diesel fractions presented a lower number of saturated rings per molecule and lower number of carbon atoms per molecule ( $NC$ ). This could be due to the lower number of side chains and to the lower side chain length. From the  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra it was also possible to determine the average number of specific  $\text{CH}_3$  groups ( $\alpha\text{-CH}_3$  and  $\gamma\text{-CH}_3$ ) per molecule. The assignment of proton resonances indicated that normally the intensity near 2.25 ppm region was due to  $\alpha\text{-CH}_3$ , while for some samples that signal could be at 2.1-2.4 ppm. The intensity of these signals generally yields 0.1 or 0.14  $\alpha\text{-CH}_3$  groups per molecule, depending on the spectra and on the occurring overlapping. In the  $^1\text{H}$  NMR spectra, the signal at around 1 ppm was assumed to be due to the  $\gamma\text{-CH}_3$  groups. In this work (Cookson *et al.*, 1987) it was also possible to identify some sub-structures, like methyl, ethyl, propyl and butyl groups, tetralin, indane and 1-methylindane ring side chains in the analysed samples of shale oil diesel.

Unlike the work of Cookson and Smith (1987) and Bansal *et al.* (1998), Lee and Glavincevski (1999) decided to analyse the content of aromatics in middle distillate oils and have developed a NMR method for the determination of the molar content of aromatics. A new expression (Equation 3.34):

$$fA_m = \frac{1}{3} \left[ \frac{H_\alpha}{\left(\frac{H}{C}\right)_\alpha} - 0.5C_{ar} + H_\alpha \right] \quad (3.34)$$

was developed to estimate the molar content of aromatic groups by using the aromatic  $\alpha$ -hydrogen to  $\alpha$ -carbon atom ratio  $(H/C)_\alpha$ , the number of  $\alpha$ -hydrogen ( $H_\alpha$ ), the number of aromatic ring carbons ( $C_{ar}$ ) and the number of aromatic ring hydrogens ( $H_a$ ). However, these authors (Lee *et al.*, 1999) have concluded that the presence of olefinic protons should be taken into account, therefore the following equation was used in the aromaticity calculation:

$$f_a(\text{corrected}) = f_a - \frac{fH_n^o \left(\frac{H}{C}\right)_t}{\left(\frac{H}{C}\right)_0} \quad (3.35)$$

where  $fH_n^o$  is the molar fraction of olefinic protons,  $(H/C)_t$  the total hydrogen to carbon atom ratio, the  $(H/C)_0$  the hydrogen to carbon atom ratio of the olefinic bond structure and  $f_a$  the fraction of aromatic carbon (aromaticity).

Sergeant *et al.* (1995) used  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectroscopy to analyse different products from shale deposits in Australia in order to obtain information about the diversity of structures composition of lubricating oils, of maltenes and of distillates. From the differences between these products, it is noteworthy that lubricating oils were characterized by the absence of olefinic structures, the Rundle distillate samples by the presence of monoaromatic with some polyaromatic and naphthenic structures, while the maltenes were characterized by the existences of some olefinic structures and aliphatic chains that were less substituted.

Recently, de Andrade *et al.* (2010) presented a review with 62 references on the determination of conjugated dienes in petroleum products using different chemical and instrumental techniques, including NMR spectroscopy. Among the implementation of different techniques analysis, the authors indicated NMR spectroscopy as one of the techniques chosen to characterize naphtha or gasoline samples and to characterize conjugated dienes in the samples. NMR was considered one of the techniques which

contributed to obtain rapid answers and information about the conjugated dienes composition.

### 3.3.4. NMR Spectroscopy in Asphaltenes Characterization

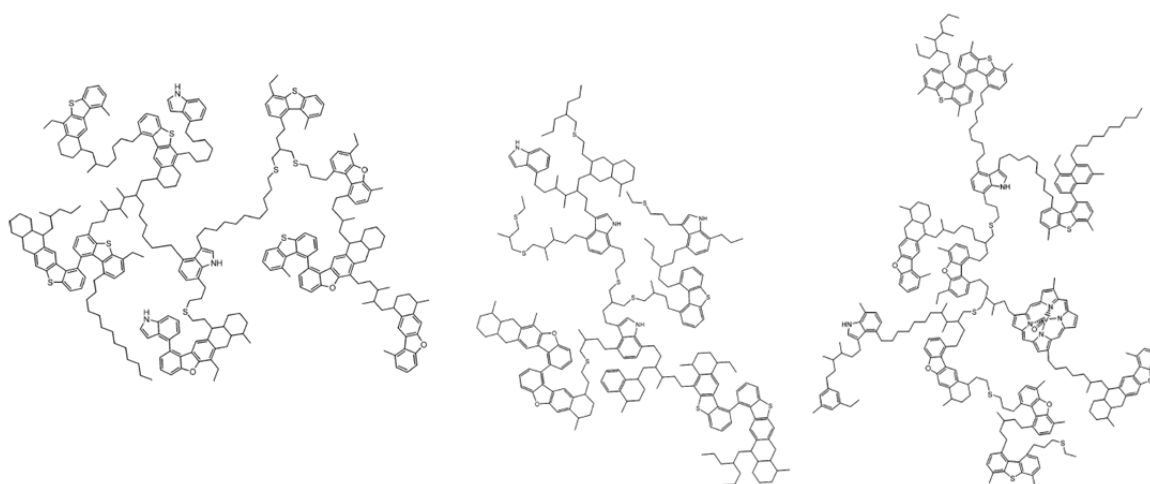
Asphaltenes present a dark colour and are defined as the fraction from crude oil that is soluble in toluene and insoluble in heptane. They are known to be the most complicated compounds in crude oil and are very difficult to analyse due to their tendency to associate, to the high molecular weight (approximately  $1000 \text{ g}\cdot\text{mol}^{-1}$ ) (Aske *et al.*, 2002) and to the paramagnetism (Simanzhenkov *et al.*, 2003). Due to their complex nature it has not been found a specific molecular structure or a specific family compounds of asphaltenes (University of Illinois at Chicago, 2001), however there are different ideas relatively to their composition. Asphaltenes are characterized to be complex mixtures containing high quantity of heteroatoms (like nitrogen, oxygen, sulfur and metals such as vanadium, nickel or iron) and condensed aromatic rings, aliphatic chains and naphthenic rings (see Table 2.1 – Chapter 2 in page 12 and 13).

The objective of several research groups is to gain more information about asphaltenes and to further understand their influence in thermal and catalytic processes.

Calemma *et al.* (1995) compared the content of asphaltenes present in heavy crude oils of seven different sources and reported structural characterization of asphaltenes using NMR and other spectroscopic techniques. NMR techniques ( $^{13}\text{C}$  NMR GASPE) allowed the identification of some CH groups in the aliphatic chains, some condensed alicyclic structures, especially in the more aliphatic asphaltene molecules, and the aliphatic  $\text{CH}_3/\text{CH}$  ratio. Furthermore it was possible to characterize the maturity of asphaltenes that is related to their capability to aggregate through  $\pi$ - $\pi$  aromatic interactions. With the analysis of some molecular parameters from the NMR spectra, like the aromatic carbon fraction ( $f_a$ ), an average number of carbon per alkyl side chain ( $n_n$ ) and the percentage substitution of aromatic rings ( $A_5$ ) it was possible to draw some conclusions. Almost all analysed asphaltenes presented comparable average aromatic condensations and average numbers of polycondensed aromatic rings in the range of 5-7. It was possible to observe that asphaltenes presented different behaviours with the increase in the carbon content. As

consequence, when the carbon content of the asphaltene samples increases, the aromaticity and the average aromatic core size increase, the average length of alkyl side chains decrease and the heteroatom content and the average molecular weights also decrease. Asphaltenes with smaller molecular weight presented more aromatic carbons, less heteroatoms, slightly sulfur content, and smaller aliphatic side chains.

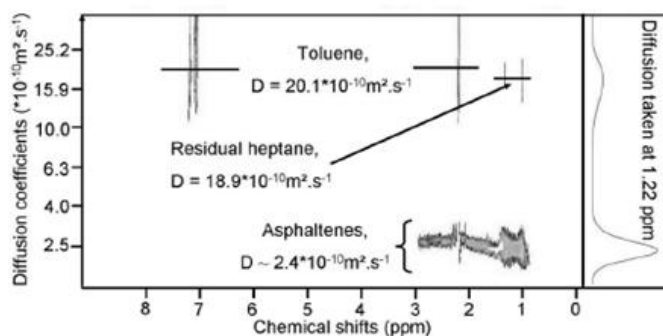
In the same path, Sheremata *et al.* (2004) described a quantitative molecular representation of Athabasca asphaltenes, where  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectroscopy was used to provide data from elemental analysis. The authors have pointed out that the combination of  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectroscopy were very useful in the analysis of asphaltene chemistry and for an additional and better chemical detailed molecular structure description of these compounds. This study provided more information about aromatic and aliphatic primary structures, the concentration of quaternary and protonated aromatic carbons, the naphthenic content and the concentration of aliphatic CH and  $\beta\text{-CH}_3$  groups. Finally, three possible structures of asphaltenes, with multiple bridges between aromatic group were represented (Figure 3.16). However it was stated that these structures, as well as those denominated asphaltenes as archipelago structures, were not rigid, and some changes can occur depending on the solvent environment. It is possible to occur interactions between the porous surface of asphaltenes and the surrounding solvent which has the capacity to entrain in the reticulated asphaltenes microstructure and change its geometry.



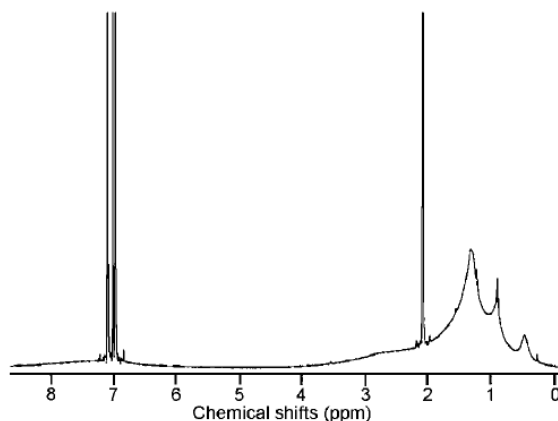
**Figure 3.16:** Example of some asphaltene structures generated using Monte Carlo method (Sheremata *et al.*, 2004).

Douda *et al.* (2008) analysed some asphaltene mixtures by comparing the asphaltene fractions obtained after pyrolysis (at different temperatures) with the original ones, by using  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectroscopy. The large bands of aromatic and saturated regions in the  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra were thought to be due to the high values of different structures of asphaltene compounds. From the obtained results it was concluded that pyrolysis processes at high temperatures were responsible for a higher formation of asphaltene. This knowledge about the effect of the temperature in the asphaltene modification is very important in the prevention of some problems during the crude oil refining, like the deposition of asphaltene in cooking, cracking and in the distillation.

Durand *et al.* (2008) presented for the first time a  $^1\text{H}$  DOSY NMR spectrum of asphaltene (Figure 3.17), from which it was possible to conclude that asphaltene has more aliphatic than aromatic structures. Although in this spectrum no aromatic protons were detected due to the influence of the toluene signals that have a high signal-to-noise ratio compared to the sample aromatic protons. However, a separation in the aromatic region exists and there are some aromatic protons, although these aromatic rings are very substituted by alkyl chains. Furthermore, it was calculated the diffusion coefficient of asphaltene and toluene using the modified Stokes-Einstein equation and then compared them with the sample concentration. The intention was to analyse the intermolecular interactions between solvent and solute and their dependence with the solute concentration and also to analyse the behaviour of the diffusion coefficients, while increasing the sample concentration, which can give some information about the nature of the mixture constituents. The separation in the diffusion coefficient between residual heptane and asphaltene was demonstrated (Figure 3.17), and it was concluded that the diffusion coefficient remains constant at low concentrations, probably due to the predominance of the solute-solvent interactions. From the signals in the spectrum (Figure 3.18) it was concluded that terminal  $\text{CH}_3$  groups of long alkyl chains were represented by signals at 0.9 ppm while at 1-2 ppm some  $\text{CH}_2$  groups were found and peaks at 2.1-3 ppm correspond to  $\text{CH}$ ,  $\text{CH}_2$  and  $\text{CH}_3$  in polycyclic aromatic hydrocarbons. It was also calculated the average molecular weight of asphaltene ( $M \approx 3450 \text{ g}\cdot\text{mol}^{-1}$ ), being this result higher than expected when comparing with the published value (approximately  $1000 \text{ g}\cdot\text{mol}^{-1}$ ) (Aske *et al.*, 2002).



**Figure 3.17:** Example of a  $^1\text{H}$  DOSY NMR spectrum of a base oil. Reprinted with permission from Durand *et al.* (2008). Copyright American Chemical Society.



**Figure 3.18:** Asphaltenes  $^1\text{H}$  NMR spectrum at 1 wt% in toluene at 20 °C. Reprinted with permission from Durand *et al.* (2008). Copyright American Chemical Society.

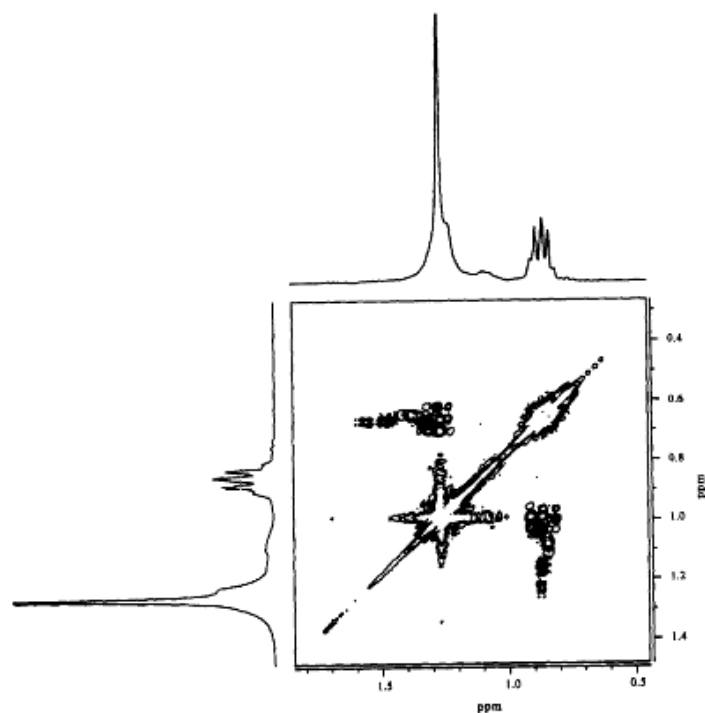
### 3.3.5. Spectral editing NMR Techniques

Since crude oil is a complex mixture of compounds it is important to use different techniques to improve the quality of the NMR spectrum and to obtain better quantitative information about the sample in analysis. For this, beyond the use of 1D NMR spectroscopy, the typical  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra, it is beneficial to use the “spectral editing” techniques and the 2D NMR experiments (see Table 3.4, page 96 and 97). Although, the advantages of the “spectral editing” techniques and the 2D NMR experiments were known, they were not applied during this work. The motivation for not

applying these techniques were related with the type of NMR spectrometer available for online analysis, since the idea was evaluating the possibility of using an online NMR spectrometer to process stream monitoring. Moreover, the experimental time required to obtain the final spectra when using these techniques was much larger than when using 1D  $^1\text{H}$  NMR spectroscopy, were the spectra data obtained could be recorded in only a few minutes. The time required to obtain the required information was very important to improve the NMR potential for routine application in industry. However, in the literature, depending on the object of research, there are many works where 1D and 2D NMR spectra are used in combination, not referring only to studies about heavy crude oil but in other different areas such as the Organic Chemistry and Biochemistry, in mixture analysis of natural products (Exarchou *et al.*, 2000; Lewis *et al.*, 2007; Kontogianni *et al.*, 2009) where the use of 2D NMR spectroscopy has demonstrated great potentialities in structural elucidation and quantification of complex mixtures.

The “spectral editing” (Altgelt *et al.*, 1994) techniques can be used to improve the analysis of the  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra due to the overlapping of signals, previously mentioned, which make difficult the identification of some structures. The aim of these “spectral editing” techniques is to contribute to the separation of primary ( $\text{CH}_3$ ), secondary ( $\text{CH}_2$ ), tertiary ( $\text{CH}$ ) and quaternary ( $\text{C}_q$ ) carbons and to a sensitivity improvement. Examples of “spectral editing” techniques are the INEPT (insensitive enhancement by polarization transfer), DEPT (distortionless enhancement by polarization transfer), GASPE (gated spin echo), PCSE (part-coupled spin echo) and QUAT (quaternary-only carbon spectra). INEPT and DEPT are examples of methods known as polarization transfer methods. The aim of these methods is to transfer the large excess of population (polarization) of the  $^1\text{H}$  to the insensitive nuclei before its perturbation (in the present case to  $^{13}\text{C}$  nuclei). On the other hand, the 2D NMR spectrum depend on the coupling between two nuclei. It is possible to have a 2D homonuclear spectrum that results from the coupling between nuclei from the same type and the 2D heteronuclear spectrum between nuclei from a different type. 2D COSY is an example of a homonuclear spectrum (Figure 3.19). In this spectrum it is correlated one proton to another proton, depending on the coupling constant value, which leads to identify resonances that are coupled each other. Cookson and Smith (1987), Sarpal *et al.* (1996), and Behera *et al.* (2008) are example of works, presented below, where this technique was used in the analysis of crude oils.

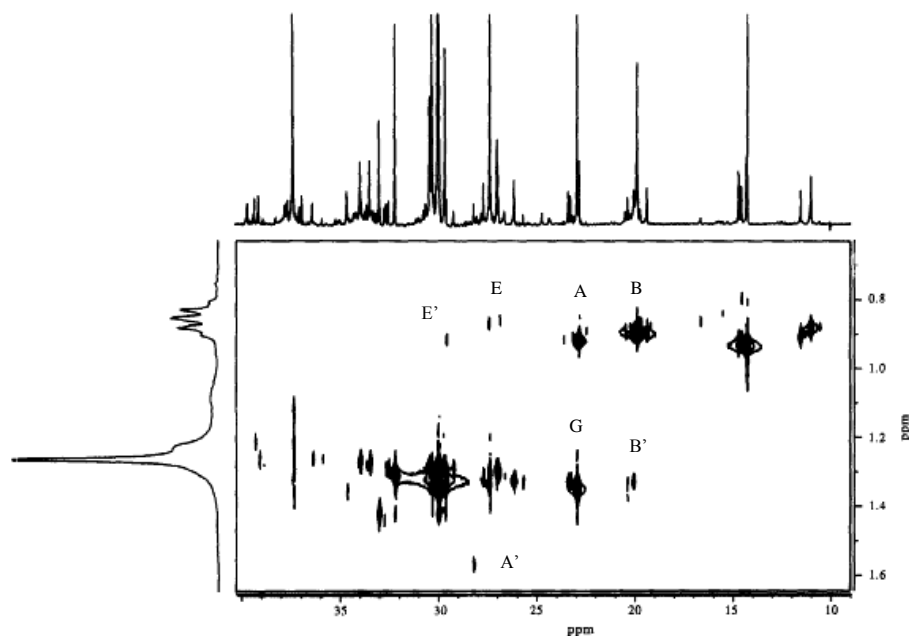




**Figure 3.19:** Example of a COSY spectrum of a base oil. Obtained using a 300 MHz NMR spectrometer with a sweep width of 3016 Hz and 4 K data points. Reprinted with permission from Sarpal *et al.* (1996). Copyright Elsevier.

TOCSY (total correlation spectroscopy), another 2D homonuclear spectrum, is very useful in the analysis of complex mixtures helping in the characterization of the molecule structure. In this case, it is established a correlation between all the spins in a set of coupled spins in a molecule. Doan *et al.* (1995) has used this technique in the analysis of polyaromatics in crude gas oil mixtures.

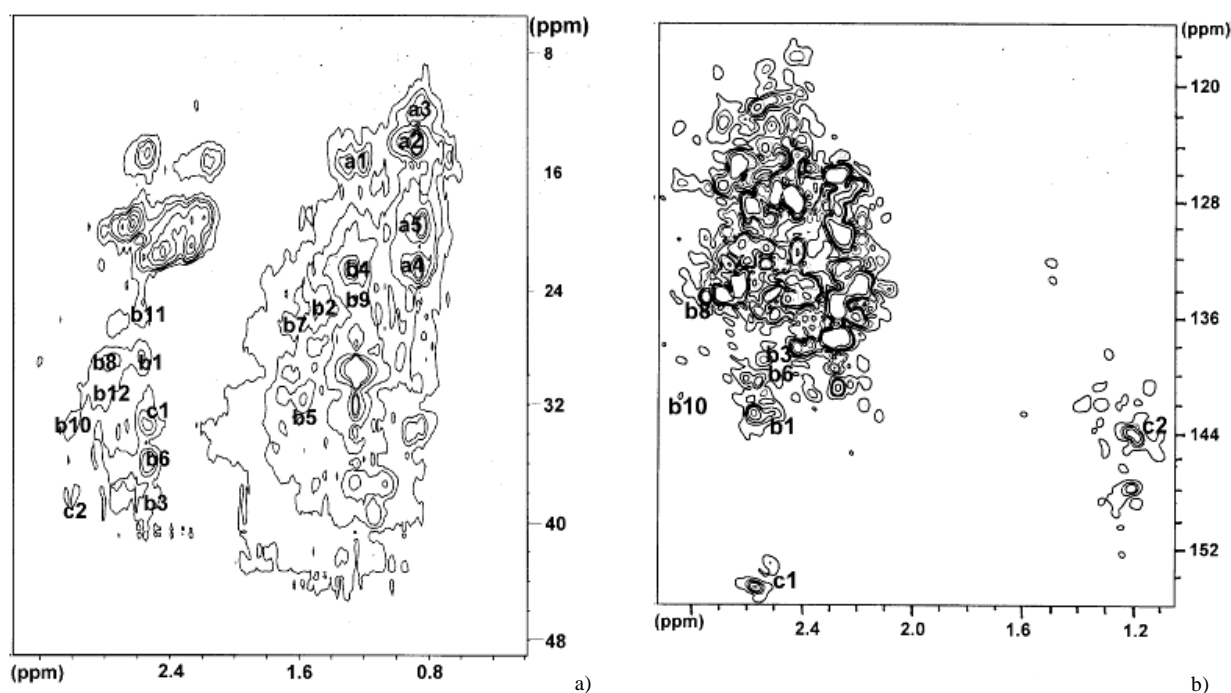
HETCOR is a 2D heteronuclear spectrum (see Figure 3.20) where it is possible to obtain a correlation between coupled heteronuclear spins across a single bond. Cookson and Smith (1987), Sarpal *et al.* (1996) and Behera *et al.* (2008) are examples of some works using the HETCOR technique in their research.



**Figure 3.20:** Example of an HETCOR spectrum of a base oil, obtained in a NMR spectrometer operating a 300 MHz for  $^1\text{H}$  and 75 MHz for  $^{13}\text{C}$ . A = isopropyl methyl groups (0.93 and 22.7 ppm) of S-1 and S-2 structures. A' = -CH group resonance (S-2) at  $\delta$  28.2. B and B' = -CH<sub>2</sub> carbon of S-4 at  $\delta$  20.3 correlated with a methyl group (19 – 21 ppm). E (0.9 and 28.6, 27.4 ppm) and E' (0.99 and 30.2 ppm) = branched methyl of S-10 and S-11. G =  $\beta$  – methylene groups (1.3 and 22.4 ppm) of S-2. With S-1, S-2, S-4 and S-10 from Table 3.2. Reprinted with permission from Sarpal *et al.* (1996). Copyright Elsevier.

Besides the referred homo- and heteronuclear spectra there are the 2D inverse detected homo- and heteronuclear experiments, although only the use in crude oil area will be described, such as the heteronuclear HSQC (heteronuclear single quantum correlation) and HMBC (heteronuclear multiple quantum correlation) spectra (Figure 3.21). In these techniques the  $^1\text{H}$  magnetization is directly detected while the  $^{13}\text{C}$  magnetization is indirectly detected. For this reason, the  $^{13}\text{C}$  magnetization is known as the “inverse” experiment. While the HETCOR, known as a heteronuclear experiment, is the “normal” one, because in this case, the  $^1\text{H}$  magnetizations are indirectly detected and the  $^{13}\text{C}$  directly detected. The “inverse” experiments are preferable over HETCOR, since the latter has lower sensitivity due to the detection of the lower- $\gamma$  nuclide, typically  $^{13}\text{C}$ . In addition, the “inverse” experiments are proton detected as the proton are much more sensitive nucleus compared with the  $^{13}\text{C}$  (Claridge, 1999). HSQC correlates coupled heteronuclear spins across a single bond while HMBC correlates coupled spins across multiple bonds. Some workers have demonstrated that the combination of HSQC and HMBC spectra is a very

good choice to get information about the carbon skeleton. Kontogianni *et al.* (2009), Exarchou *et al.* (2000) and Behera *et al.* (2008) are representative works using these techniques in the analysis of crude oils.



**Figure 3.21:** Example of an a) HSQC and an b) HMBC spectra of a high boiling point fraction VGO, identifying CH<sub>3</sub> (a), CH<sub>2</sub> (b) and CH (c) groups in different side chains - ethyl (a<sub>1</sub>, b<sub>1</sub>), propyl (a<sub>2</sub>, b<sub>2</sub>b<sub>3</sub>), butyl (a<sub>3</sub>, b<sub>4</sub>b<sub>5</sub>b<sub>6</sub>), isopropyl (a<sub>4</sub>, c<sub>1</sub>), tetraline (b<sub>7</sub>b<sub>8</sub>), indane (b<sub>9</sub>b<sub>10</sub>) and 1-methylindane (a<sub>5</sub>, b<sub>11</sub>b<sub>12</sub>c<sub>2</sub>). Obtained in a Bruker Avance 400 MHz NMR spectrometer equipped with an inverse detecting probe, using a  $\pi/2$  pulses of 13.3 and 9.6  $\mu$ s for <sup>1</sup>H and <sup>13</sup>C, respectively, a recycle delay of 2 s and three sine gradients of 1.5 ms with 8:3:2 ratio for HSQC and 5:3:4 ratio for HMBC. Reprinted with permission from Behera *et al.* (2008). Copyright Elsevier.

The DOSY (Diffusion Ordered Spectroscopy) spectrum (Figure 3.17) is very useful to analyse the composition of mixtures based on the differences in diffusion coefficients of individual components. The obtained 2D spectrum is characterized for presenting, in one direction, the signals dispersed according to chemical shift while in the other the signals are dispersed taking into account the diffusion constant. Since this technique measures the self-diffusion coefficients it contributes to obtain information on structural, molecular size, shape, aggregation states and composition of complex mixtures. One of the great advantages of this technique is that it doesn't need a prior separation of individual

components to separate, assign and identify the different signals of a crowded spectrum (Claridge, 1999; Kapur *et al.*, 2000; Durand *et al.*, 2008).

Cookson and Smith (1987) used homonuclear 2D COSY spectra, coupled and decoupled 2D HETCOR spectra, and a RELAY 2D NMR method to provide a better analysis of the  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra and thus making a better structural assignment.

Sarpal *et al.* (1996) used GASPE, PCSE, INEPT, DEPT, QUAT, 2D COSY and HETCOR to classify the composition of some different base oils in terms of the hydrocarbon types, especially the isoparaffins. 2D COSY and HETCOR helped to specifically improve the identification of some branching structures, since these spectra allow clarifying the existing information inside the overlapping signals. The use of DEPT also contributed to the improvement of the signals overlapping, especially in the region 28-40 ppm, as well as contributing to the separation of some carbons of normal paraffins, but did not give any idea about the existence of quaternary carbons, which could be due to the size of these structures. The use of 2D COSY and HETCOR contributed to the improvement in the identification and quantification of some types of branched structures (about eleven types of different branched structures).

Besides the techniques used by Sarpal *et al.* (1996), it was already demonstrated that other techniques could contribute to the separation, assignment and identification of various signals in a crowded spectrum. Kapur *et al.* (2000) showed that the 2D DOSY NMR spectroscopy is very useful in the interpretation of the  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra of such complex mixtures such as crude oil, which are characterized by a great signals overlapping making difficult the identification of some components. In the  $^1\text{H}$  NMR spectra the overlapping occurs in the 0.5-2.0 ppm range due to the saturated components, whereas in the  $^{13}\text{C}$  NMR the overlapping was stronger in the 10-45 ppm range. Using the DOSY spectra the characteristic overlapping of both  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra was reduced allowing a better resolution of the signals and a better structure elucidation. This technique demonstrated to be very helpful in the analysis of heavy petroleum fractions.

After Kapur *et al.* (2000) used a DOSY technique in the analysis of an aromatic petroleum fraction, in the distillate boiling range, Durant *et al.* (2008) presented a DOSY spectrum of a diesel sample and for the first time a  $^1\text{H}$  DOSY spectrum of asphaltenes. These authors

(Durand *et al.*, 2008) confirmed that DOSY NMR technique is very important in the analysis of complex samples, such as crude oil, that could provided information on size, molecular weight, aggregation state and the composition of the mixture. The main advantage of this technique is the potential to give physical-chemical information all at once when compared to the pulse field gradient spin-echo NMR diffusion sequences (PFGSE NMR).

More recently, Behera *et al.* (2008) used some NMR techniques, such as the gated-decoupled  $^{13}\text{C}$ , DEPT, 2D ( $^1\text{H}$ - $^{13}\text{C}$ ) HETCOR and other 2D NMR methods for better identification of some structures in the very crowded NMR spectra of heavy petroleum fractions. For example, these authors (Behera *et al.*, 2008) showed that HSQC and HMBC spectra were proven to be better techniques in the identification of side chain structures of aromatics.

**Table 3.4:** Spectral editing NMR techniques.

		<b>Advantages/Drawbacks</b>
GASPE	Gated Spin Echo	Information about carbon type distribution, distinguishing and isolating C, CH, CH <sub>2</sub> and CH <sub>3</sub> sub-spectra.
PCSE	Part-Coupled Spin Echo	Information about carbon type distribution, distinguishing and isolating C, CH, CH <sub>2</sub> and CH <sub>3</sub> sub-spectra (practically identical to GASPE).
DEPT	Distortionless Enhancement by Polarization Transfer	Better than GASPE and PCSE due to a precise and quick analysis of the CH and CH <sub>3</sub> sub-spectra. Quantifies CH, CH <sub>2</sub> and CH <sub>3</sub> carbons but not quaternary carbons.
QUAT	Quaternary-only Carbon Spectra	Used to determine quaternary carbons.
DOSY	Diffusion-Ordered Spectroscopy	One of the most commonly employed methods for identifying compounds in complex mixtures of petroleum fractions depending on different diffusion coefficients. Allows a better resolution of the signals and a better structure elucidation in both <sup>1</sup> H and <sup>13</sup> C NMR spectra when comparing with the 1D <sup>1</sup> H and <sup>13</sup> C NMR spectra. Although, when comparing with a <sup>1</sup> H- <sup>13</sup> C HSQC or HMBC it presents very low resolution in the diffusion axis and therefore considered as an inferior method.
INEPT	Insensitive Enhancement by Polarization Transfer	Results of the excitation of <sup>1</sup> H and polarization transfer to <sup>13</sup> C by H- <sup>13</sup> C coupling which contributes to the identification of CH <sub>n</sub> groups but quaternary carbons are not identify in the INEPT spectra. The principal drawback is that the signal from each carbon is spread with many components resulting in the overlap of other carbon nuclei (Becker, 2000).
2D HETCOR	Heteronuclear Correlation Spectroscopy	Correlation of the <sup>1</sup> H and <sup>13</sup> C NMR spectra, allowing a better interpretation of these spectra due to the identification of peak position. Is called a “direct” experiment because it is a way of doing <sup>1</sup> H- <sup>13</sup> C correlation by direct detection of <sup>13</sup> C (Jacobsen, 2007).
2D COSY	Homonuclear Correlation Spectroscopy	Correlation between coupled protons NMR spectrum, allowing a better interpretation of the proton NMR spectrum due to the identification of peak position.

**Table 3.4:** Spectral editing NMR techniques (continuation).

		<b>Advantages/Drawbacks</b>
2D TOCSY	Homonuclear Total Correlation Spectroscopy	Correlation between all spins in a set of coupled spins of a molecule and contributes to the structure characterization of a molecule even in such complex mixtures as crude oils.
HSQC	Heteronuclear Single Quantum Correlation	Technique similar to HETCOR, called an “inverse” experiment since $^1\text{H}$ is directly detected and $^{13}\text{C}$ indirectly detected. This technique identifies the heteroatoms to which the protons are directly attached and can be used to detect carbon nuclei indirectly from more sensitive protons, to assign side chain structures to aromatics (Günther, 1995) and to locate the position of a functional group within a known carbon skeleton (Jacobsen, 2007).
HMBC	Heteronuclear Multiple Quantum Correlation	Technique also called an “inverse” experiment. This type of technique known as a long range heteronuclear correlation over typically two or three bonds, can be used to obtain information about the location of carbon-carbon or carbon-heteroatoms bonds. It is also used to detect indirectly quaternary and protonated carbons coupled to protons and provide unique information about the skeleton of a molecule (Claridge, 1999; Jacobsen, 2007).
PFGSE	Pulse Field Gradient Spin-Echo	Characterized by introducing a perturbation in the magnetic field that propagates to the nearby protons. This perturbation only leaves the peak of interest and the rest of the peaks are destroyed (Jacobsen, 2007). This technique allows the measurement of asphaltenes diffusion coefficients.
RELAY	Heteronuclear Relayed Coherence Transfer	Gives information about coupled $\text{H}_n\text{-H}_m$ groups and correlates $^{13}\text{C}$ resonances associated with the $\text{CH}_n$ and $\text{CH}_m$ groups (Cookson <i>et al.</i> , 1987).

The combination of HSQC and HMBC is, today, a powerful method used in tracing the carbon skeleton of a compound (Jacobsen, 2007).

### 3.3.6. NMR Relaxation Effects

The spin nuclear relaxation process is a very important way to obtain information about stereochemistry and molecular dynamics. This nuclear relaxation phenomenon is related with the structure, flexibility and molecular mobility that is responsible for the existent of different magnetic fields. The relaxation rate can be determined experimentally using the relationship between the relaxation times,  $T_1$  and  $T_2$ , and the correlation time for molecular tumbling. The NMR techniques based on the analysis of  $T_1$  (*longitudinal or spin-lattice relaxation time*) and  $T_2$  (*transverse relaxation time*) decay curves can be obtained using the low-field  $^1\text{H}$  NMR spectroscopy and correlated and supplemented with information taken with the high-field nuclear magnetic resonance spectrometer.

The relaxation times are important tools in the comprehension of the molecular structure. For example, when comparing two molecules if the relaxation time decreases it can be due to the existence of intramolecular hydrogen bonds that decrease the mobility. The low field  $^1\text{H}$  NMR relaxation method has been used for a long time especially in petrophysical analysis and well logging in petroleum exploration (Yao *et al.*, 2010). Since 1990s, this method has been used in the characterization of reservoir properties (porosity, permeability, viscosity and saturation of oil and gas) (Yao *et al.*, 2010).

The correlation of relaxation times ( $T_1$  and  $T_2$ ) with viscosity/temperature and Larmor frequency was analysed by Hirasaki *et al.* (2003). These authors demonstrated that relaxation times  $T_1$  and  $T_2$  are equal, in crude oils, at low viscosity and can be correlated with the ratio of viscosity/temperature. However, for high viscosity crude oils  $T_1$  and  $T_2$  are different.  $T_1$  is a function of the Larmor frequency while  $T_2$  follow the viscosity/temperature correlation. They concluded that the temperature influences the viscosity as well as the relaxation time of heavy oils. In Hirasaki *et al.* (2003) work NMR was used to analyse the presence of water, oil and gas in the reservoirs of petroleum, taking into account the relationship between diffusivity and relaxation time. From the interpretation of the NMR logs it was concluded that: i) water normally presented a large relaxation time in rocks due to the surface relaxation on the pore walls with different sizes; ii) the fluids presented different diffusivity, which is a way to distinguish one fluid from others; and iii) crude oil presented a huge relaxation time and a large diffusivity distribution due to its rich composition in different compounds.



Ramos *et al.* (2009) used  $T_2$  of protons in the viscosity prediction of crude oils. The analysis of the expansion of the  $^1\text{H}$  raw  $T_2$  relaxation curves of four Brazilian crude oil samples with different viscosities has demonstrated that there was a decrease in the relaxation decay curve when the oil viscosity increased. Therefore, a shorter  $T_2$  was produced for the sample with high viscosity. As a result, more viscous oil represents a lower relaxation time and less viscous oil has a larger relaxation time. In crude oil, saturated fractions and resins are responsible for a higher contribution to the  $T_2$  spectra, aromatics are responsible for a lower contribution while the asphaltene protons are less expressive in the  $T_2$  spectra.

### **3.4. Final Remarks**

The preference of a method to analyse heavy crude oil depends especially on the nature of the sample and on the analysis purpose. NMR was chosen as the main technique to be used in crude oil refining industries that want to characterize the “messy” that describes the crude oil fractions and the obtained refined products. With all analysed references, NMR have been proven to be a potential technique in providing information about the physical-chemical properties and in the structure and chemical composition of crude oils.



# 4

## **EXPERIMENTAL PROCEDURES**

This chapter is focused in the description of the physical-chemical techniques and nuclear magnetic resonance spectroscopy used to characterize LVGO, HVGO and fuel oil. The parameters described are important for process monitoring and during refining operation. Experimental setup and careful planning of the samples preparation are of utmost importance when the objective is to obtain the best possible data for analysis and interpretation. Some pre-processing and data manipulations of the NMR spectra employed to improve the performance of the subsequent multivariate analysis are also described.

## 4. EXPERIMENTAL PROCEDURES

4.1. Samples.....	103
4.2. Physical-Chemical Characterization.....	104
4.2.1. Kinematic Viscosity (ASTM D445) .....	104
4.2.2. Density (ASTM D4052 and ASTM D5002).....	109
4.2.3. Carbon Residue (ASTM D4530) .....	113
4.2.4. Sulfur Content (IP 336).....	116
4.2.5. Flash Point (ASTM D93).....	119
4.2.6. P-Value (SMS 1600).....	121
4.2.7. Distillation at Atmospheric Pressure (ISO 3405) .....	123
4.2.8. Distillation at Reduced Pressure (ASTM D1160) .....	124
4.3. NMR Experiments.....	127
4.3.1. <sup>1</sup> H NMR Experiments .....	127
4.3.2. <sup>13</sup> C NMR Experiments .....	129
4.3.3. Pre-Measuring Steps .....	131
4.3.4. Data Pre-Processing .....	133
4.3.4.1. Phase and Baseline Correction .....	134
4.3.4.2. Chemical Shift Alignment.....	135
4.3.4.3. Data Reduction .....	137
4.3.4.4. Binning .....	137
4.3.4.5. Normalization .....	138
4.3.4.6. Intensity Scaling .....	138
4.3.4.7. Creation of Data Set for Multivariate Analysis.....	139
4.4. Final Remarks.....	140

## 4.1. Samples

LVGO, HVGO and fuel oil are different products presenting different and complex chemical compositions and consequently different physical-chemical properties. There are some factors affecting the chemical composition of these fractions such as unit operation (changes in temperature, pressure and residence time) and processed crude oil, which is the feedstock from the entire process. The knowledge of the composition and the physical-chemical properties is very important to control the quality and the characteristics of finished products and to optimize unit operation. The determination of the kinematic viscosity, density, sulfur content, carbon residue, flash point, P-value and atmospheric and vacuum distillation are examples of some properties that can be determined with the goal to improve product quality parameters or plant control. The determination of such properties is normally made taking into account some tests following the ASTM (American Society for Testing and Materials) norms, the IP (Institute of Petroleum) norms and other traditional tests. Apart from the samples analysed by these techniques the samples undergo an analysis using the NMR spectrometer. All ASTM, IP and other test procedures used in this work to characterize the different samples analysed are presented in Table 4.1.

With the exception of the LVGO samples, HVGO and fuel oil samples were heated before use. A short heating with a lower temperature (no more than one hour with a temperature higher than 200 °C) was desirable only to ensure that the samples could be handled. An excessive heating could lead to the oxidation of some compounds and to some changes in the structure and composition of the samples. The heating, when necessary, and the homogenization of the samples were the only treatment needed before starting the analysis by the standard procedures.

**Table 4.1:** Identification of samples for properties that were analysed.

Properties	LVGO	HVGO	Fuel Oil
Kinematic viscosity 50 °C (ASTM D445)			✓
Kinematic viscosity 100 °C (ASTM D445)		✓	✓
Density (ASTM D4052)	✓		
Density (ASTM D5002)		✓	✓
Carbon residue (ASTM D4530)		✓	✓
P-value (SMS 1600)			✓
Sulfur content (IP 336)			✓
Distillation (ISO 3405)	✓		
Distillation (ASTM D1160)		✓	
Flash point (ASTM D93)			✓
<sup>1</sup> H NMR spectra	✓	✓	✓
<sup>13</sup> C NMR spectra			✓

## 4.2. Physical-Chemical Characterization

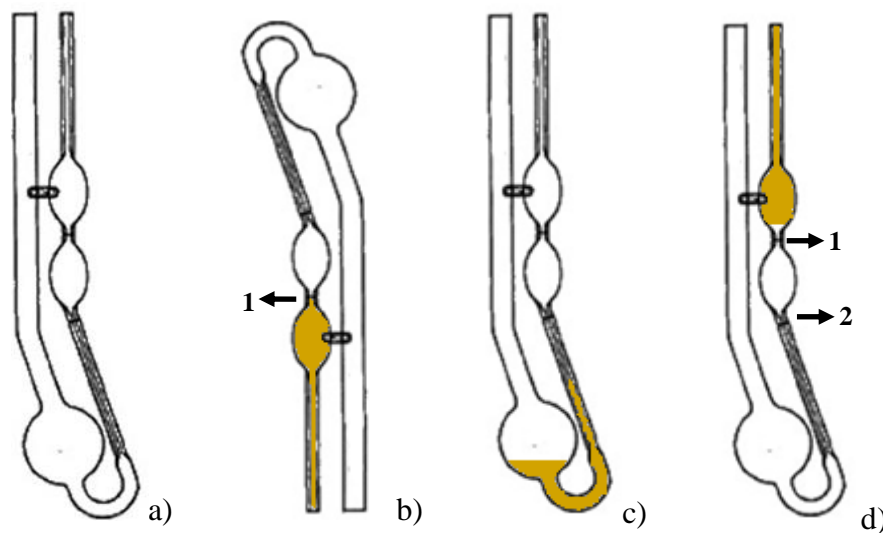
### 4.2.1. Kinematic Viscosity (ASTM D445)

The standard test method for kinematic viscosity of transparent and opaque liquids (ASTM Standards D445, 2009) is a very important procedure in the determination of the kinematic viscosity of petroleum fuels. The kinematic viscosity is a measure of liquids resistance to flow. In the determination of this property, the time for a volume of liquid to flow under gravity through a calibrated glass capillary viscometer under a reproducible driving head and at a known temperature is measured (Wauquier, 1995; Simanzhenkov *et al.*, 2003; ASTM Standards D445, 2009). In this work the kinematic viscosity of fuel oil and HVGO samples was determined. The kinematic viscosity of LVGO samples was not measured since the main objective of analysing the LVGO stream was to evaluate its fractionation. For that, other properties were considered more important.

#### Procedure for transparent liquids, such as HVGO

Prior to measure the kinematic viscosity of any sample it was necessary to select the right viscometer with an appropriate capillary having a diameter that let the sample flow freely.

The configuration and size of the capillary were calibrated and tested to give a constant value which was specific for each viscometer and helped to differentiate one viscometer from the other. Normally, for the HVGO samples the most common viscometer used was a 150 capillary viscometer *Cannon-Fenske*. After choosing the viscometer (Figure 4.1, a), it was filled, using suction, with the HVGO sample up to the filling mark (1) (Figure 4.1, b) and then inserted into the viscosity Tamson bath (Figure 4.2) at  $100.00 \pm 0.02$  °C. The sample was left in the bath during 30 minutes in order to achieve thermal equilibrium (Figure 4.1, c). Using suction, the sample was adjusted to above the upper timing mark 1 (Figure 4.1, d) and when it arrived at the first timing mark (1) by gravity, the time required for the sample to flow from the first (1) to the second timing mark (2) was measured in seconds ( $t_1$ ). The procedure was repeated ( $t_{1,2}$ ) and the kinematic viscosity ( $\nu$ ) was determined considering the average of the two viscosity values calculated ( $\nu_{1,2}$ ), as will be explained posteriorly.



**Figure 4.1:** Example of a Cannon-Fenske viscometer for transparent liquids.

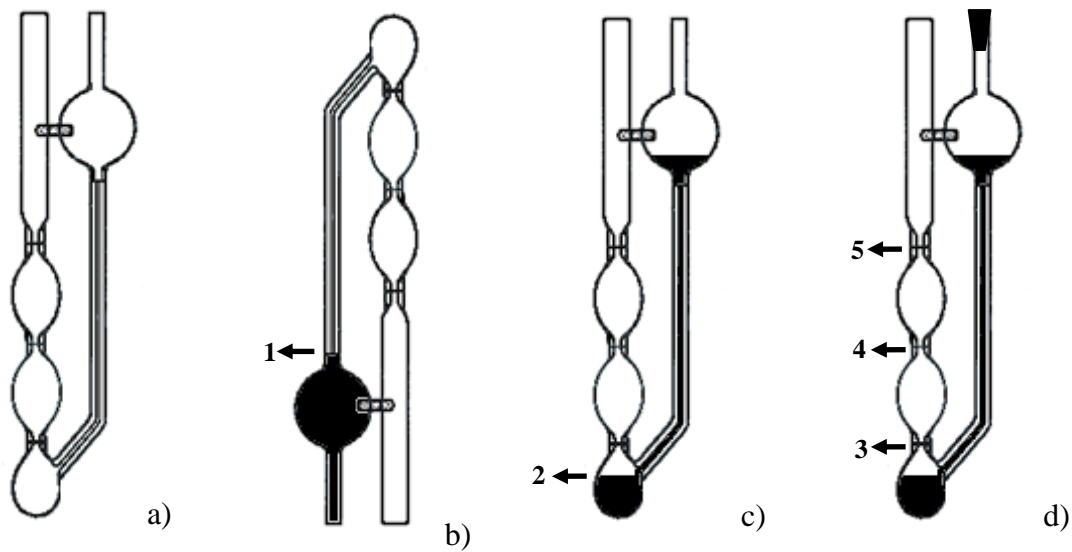


**Figure 4.2:** Viscosity Tamson bath with a viscometer for transparent liquids (ASTM Standards D445, 2009).

### **Procedure for opaque liquids, such as fuel oil**

The viscometer mostly used for the fuel oil samples was the 200 capillary viscometer *Cannon-Fenske* for a determination at 100 °C and the 450 capillary viscometer *Cannon-Fenske* for a determination at 50 °C (Figure 4.3, a). The procedure was similar with that used in the determination of the kinematic viscosity of transparent liquids but in this case the samples need to be filtered. The viscometer was also filled, using suction, with the fuel oil sample until the filling mark (1) (Figure 4.3, b), and then the viscometer was inserted into the bath at 100.00±0.02 °C (Figure 4.2) or 50.00±0.02 °C, letting the sample flow by gravity until it filled more than half of the first bulb (2) (Figure 4.3, c). When the sample was at a rich point a stopper was placed on top at one side of the viscometer (Figure 4.3, d). To achieve thermal equilibrium, the sample was left in the bath for 30 minutes. After which, the time required for the sample to advance between the first mark (3) to the second one (4) ( $t_1$ ), and from the second to the third mark (5) ( $t_2$ ) was measured in seconds and recorded for the kinematic viscosity determination.





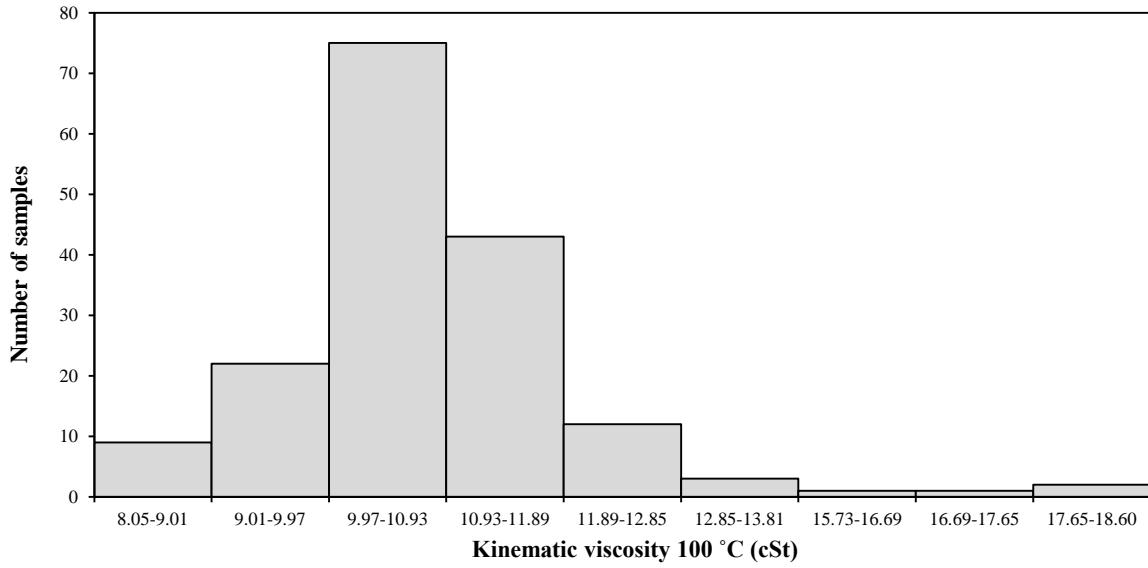
**Figure 4.3:** Example of a Cannon-Fenske viscometer for opaque liquids.

Independently of having an opaque or transparent liquid the way to determine the kinematic viscosity values,  $\nu_1$  and  $\nu_2$ , was the same. Each flow time,  $t_1$  and  $t_2$ , measured was multiplied by the respectively viscometer constant,  $C_v$ , by means of the following equation:

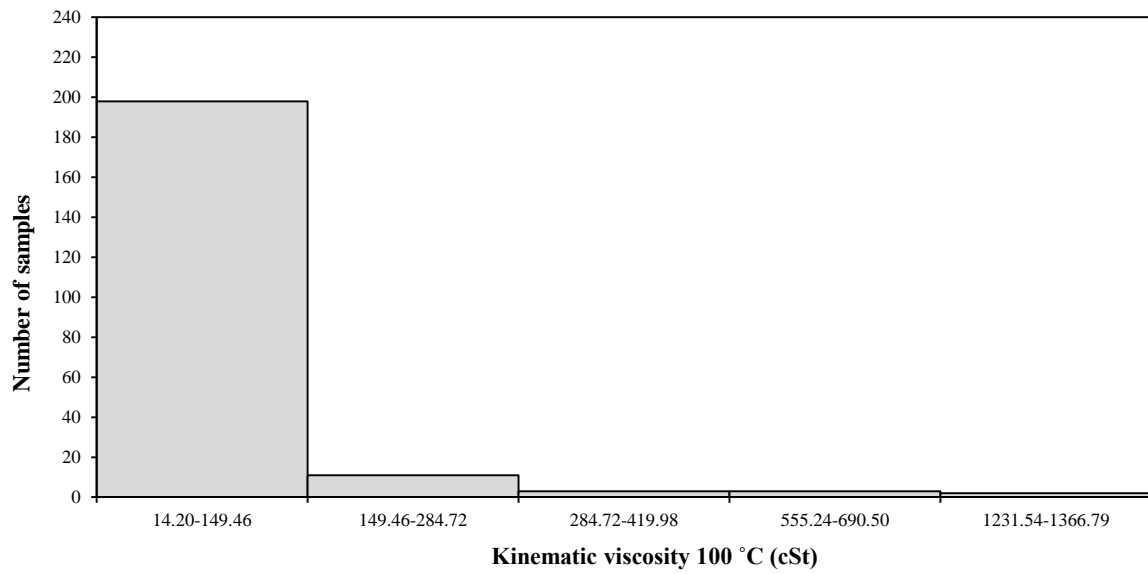
$$\nu_{1,2} = C_v \times t_{1,2} \quad (4.1)$$

The final result of the kinematic viscosity,  $\nu$ , was obtained as an average of the two calculated values for each sample ( $\nu_{1,2}$ ). To ensure that no mistakes were made, the kinematic viscosity of one sample was always measured twice using different viscometers, with different viscometer constants. With that, it was possible to calculate the repeatability, the difference between successive results obtained with the same apparatus under constant operating conditions and with the same sample, and to evaluate the precision of the test method.

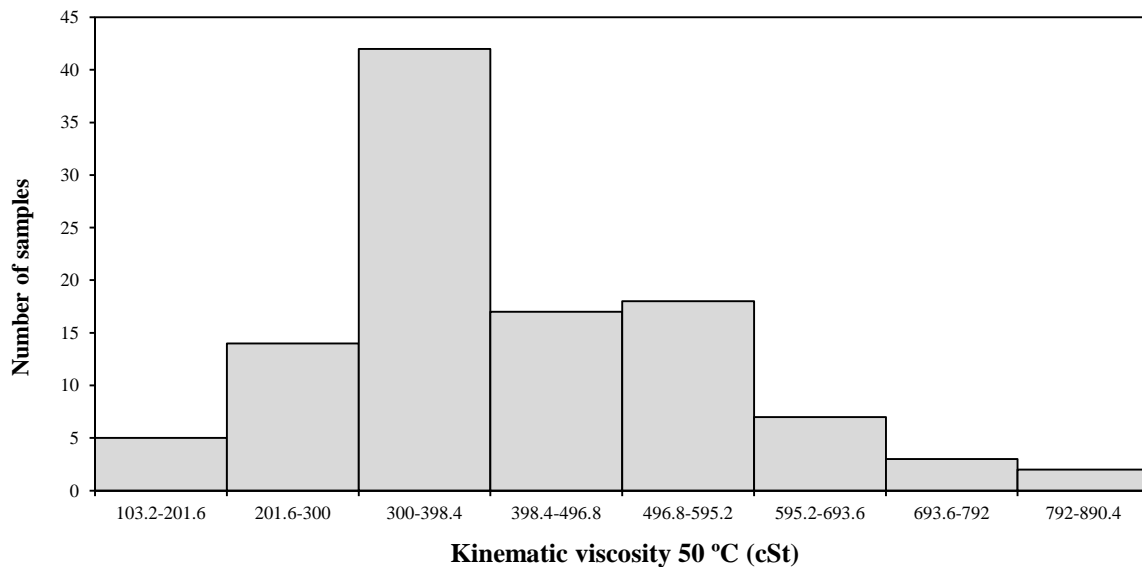
In Figure 4.4 and in Figure 4.5 the range of kinematic viscosity variations at 100 °C of all analysed HVGO and fuel oil samples is represented. While, in Figure 4.6 the range of the kinematic viscosity variations at 50 °C of all analysed fuel oil samples is demonstrated.



**Figure 4.4:** Histogram for 168 samples on the variation of HVGO kinematic viscosity at 100 °C.



**Figure 4.5:** Histogram for 217 samples on the variation of fuel oil kinematic viscosity at 100 °C.



**Figure 4.6:** Histogram for 108 samples on the variation of fuel oil kinematic viscosity at 50 °C.

#### 4.2.2. Density (ASTM D4052 and ASTM D5002)

Density is an important physical property that can be used to characterize petroleum products and petroleum fractions, in conjunction with other properties. There are different types of ASTM methods, to determine densities, used in the refinery for plant control and finished product quality. Examples of used ASTM densities were the ASTM D4052 (ASTM Standards D4052, 2011) for LVGO characterization, and the ASTM D5002 (ASTM Standards D5002, 2005) in the analysis of HVGO and fuel oil samples.

##### *ASTM D4052 – Standard test method for density, relative density and API gravity of liquids by digital density analyzer*

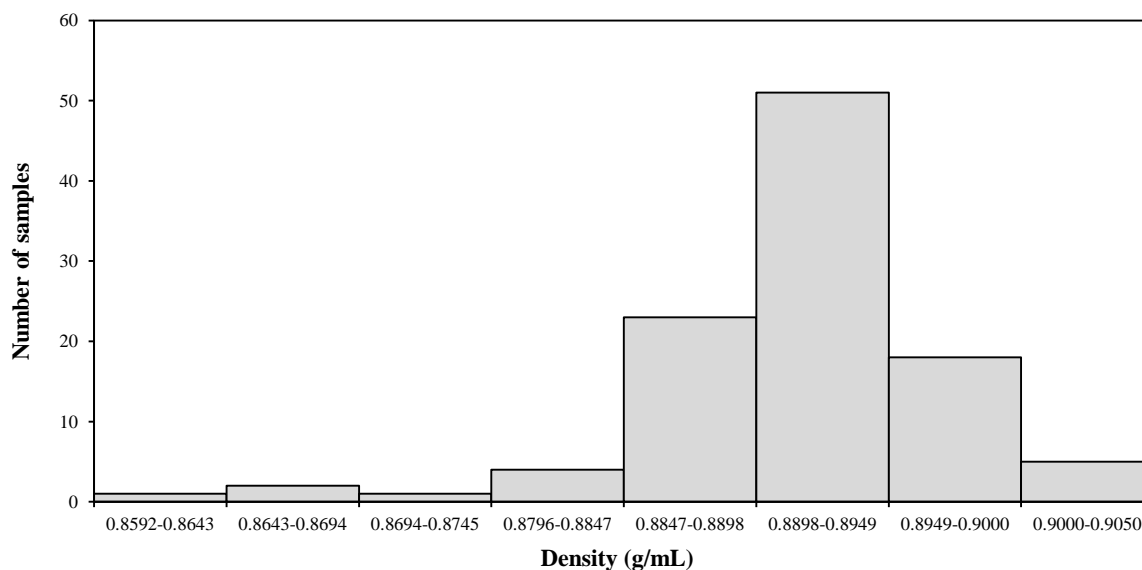
The ASTM D4052 method can be used for the determination of the density, relative density, and API gravity of petroleum distillates and viscous oils that can be handled as liquids at the tested temperature between 15 to 35 °C. However, this method is restricted to liquids with a viscosity value below 15000 cSt and a total vapor pressure typically below 100 kPa. Nevertheless, liquids with a total vapor pressure higher than 100 kPa can also be

analysed when ensured that the density determination is not affected by other perturbations such as bubbles in the oscillating sample tube. Gasoline and gasoline oxygenate blends, diesel, jet, base stocks, waxes and lubricating oils are examples of products that can be assessed by this method.

The density of LVGO was analysed by the standard test method using a digital density analyzer DMA 4500 (Figure 4.7). In this method, a small amount (about 1 to 2 mL) of liquid was measured in an oscillating sample tube and due to the existence of a system for electronic excitation and frequency counting, the total mass of a tube was measured by determining its natural frequency of vibration. The oscillating frequency depends on the dimension and elastic properties of the tube, and on the weight of the tube and of the sample. The results of the vibration frequency used in conjunction with calibration data provided the sample density. With one analysis the apparatus provided two determinations in the density of the sample and if the two determinations did not differ by more than 0.0002 g/mL the result was accepted. The range of density variations are presented in Figure 4.8.



**Figure 4.7:** Digital density analyzer, Anton Paar DMA 4500M (ASTM Standards D4052, 2011).



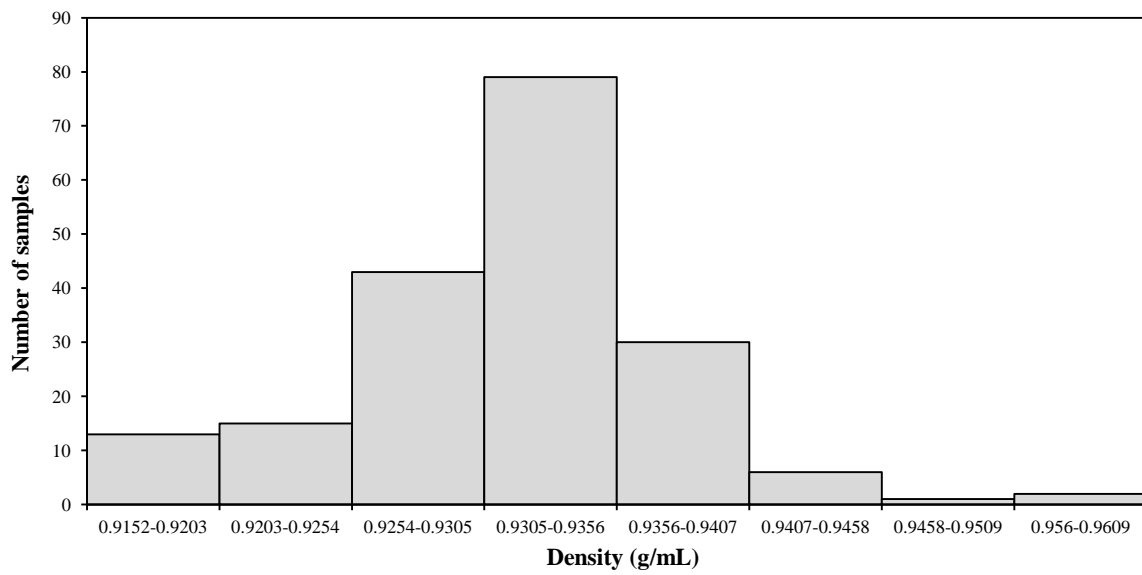
**Figure 4.8:** Histogram for 105 samples on the variation of LVGO density.

#### ***ASTM D5002 – Standard test method for density and relative density of crude oils by density analyzer***

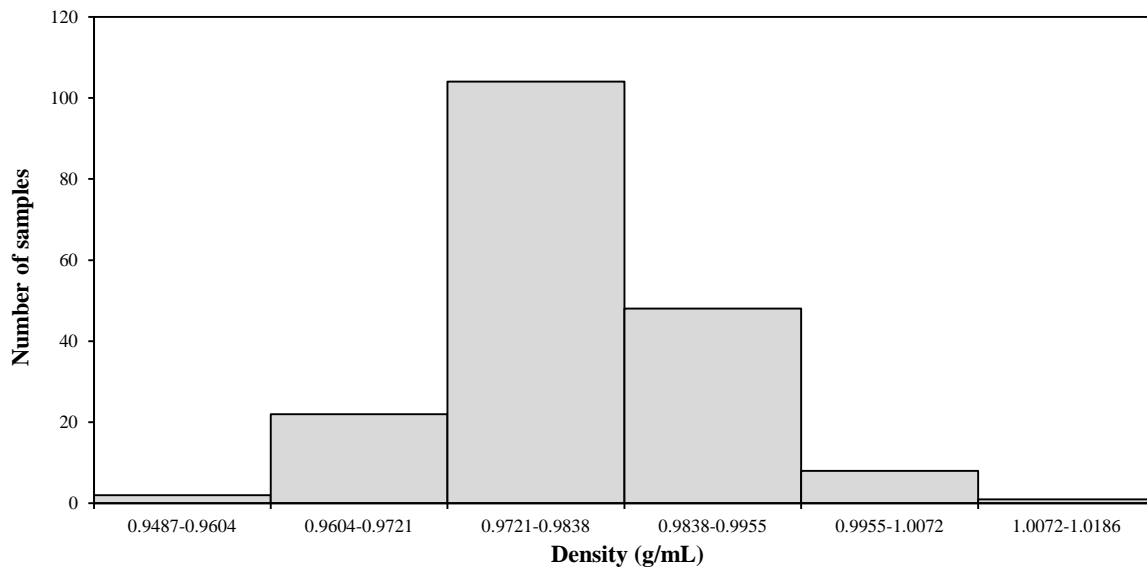
For HVGO and fuel oil samples it was more appropriate to use the ASTM D5002 standard test method. This method can be used to determine the density or relative density of crude oils that can be handled as liquids at tested temperatures between 15 and 35 °C. It can also be used in the analysis of crude oils with high vapor pressures and when necessary it can require measurements at higher temperatures to eliminate air bubbles in the sample. For these reasons samples as HVGO and fuel oil were analysed with this standard test method. We were referring to two dark samples that required some temperature increment to ensure that no unseen air or gas bubbles were in the sample. The basic principle of the operation in the equipment used (Figure 4.9) was similar to the ASTM D4052 standard test method. Beyond the required temperature, the main difference was related to the quantity of the sample that was required for the analysis, approximately 0.7 mL. The variations of the results obtained are presented in Figure 4.10 and Figure 4.11.



**Figure 4.9:** Digital density analyzer used, Mettler Toledo Density Meter De40 (ASTM Standards D5002, 2005).



**Figure 4.10:** Histogram for 189 samples on the variation of HVGO density.



**Figure 4.11:** Histogram for 185 samples on the variation of fuel oil density.

#### 4.2.3. Carbon Residue (ASTM D4530)

The standard test method used for the determination of carbon residue (ASTM Standards D4530, 2007) allowed evaluating the tendency of such materials (HVGO and fuel oil samples) to coke formation. The amount of carbon residue produced was dependent on the test conditions when evaporation and pyrolysis occurred.

The procedure used in the determination of this property can be resumed into two main parts. The first one involved the sample preparation while the second one was related to the processing of the samples. The first step in the preparation of the sample was to heat and stir the samples until a homogeneous liquid was obtained. This step was only necessary once the samples in analysis were HVGO and fuel oil, two samples that when cooled were difficult to handle. After that, an appropriate mass of the sample was transfer into a tare sample vial and then the vial was placed into a vial holder. The required sample mass depended on the samples description. In the case of HVGO samples, the recommended quantity of the sample was  $1.5 \pm 0.5$  g, while for black, viscous or solid samples such as fuel oil, the quantity recommended was  $0.15 \pm 0.05$  g. After the preparation and during the processing of the samples, the vial holder, with the vials, was placed into the oven at less

than 100 °C. The oven (see Figure 4.12) was slowly heated up to 500 °C under an inert (nitrogen) atmosphere in a controlled manner for a specific time. During this time, the volatiles formed with the cooking reactions were swept away by the nitrogen, leaving the carbon residue. After a specific period of heating and when the temperature in the oven was less than 250 °C, the vial holder was removed for further cooling. The cooling of the vials was made in a desiccator. When cooled, the vials were weighed.

The carbon content in the original sample was calculated as the ratio between the weight of the carbon left and a percentage of the weight of the original sample, as in the following equation:

$$\% \text{ carbon residue} = \frac{(A \times 100)}{W} \quad (4.2)$$

where  $A$  was the mass of carbon residue (g) and  $W$  the mass of the sample used (g).

In Figure 4.13 and Figure 4.14 the range of carbon residue (% m/m) variations of all HVGO and fuel oil samples analysed are represented. Information about the carbon residue of LVGO samples is not given since this property was not included in the characterization of the LVGO samples. As previously mentioned, the carbon residue gives indication of the amount of coke that will be formed during thermal processes as well as of the amount of high boiling constituents. As LVGO samples are rich in lower boiling fractions and contain insignificant quantities of coke precursors a lower carbon residue is expected.





Figure 4.12: Coking oven, Alcor MCRT-160 (ASTM Standards D4530, 2007).

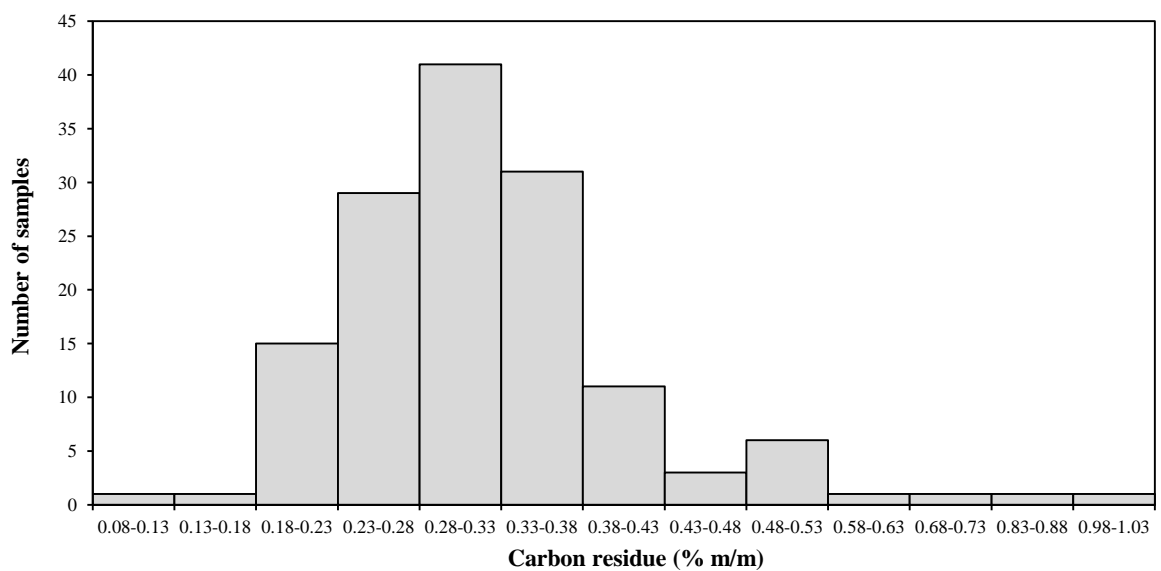
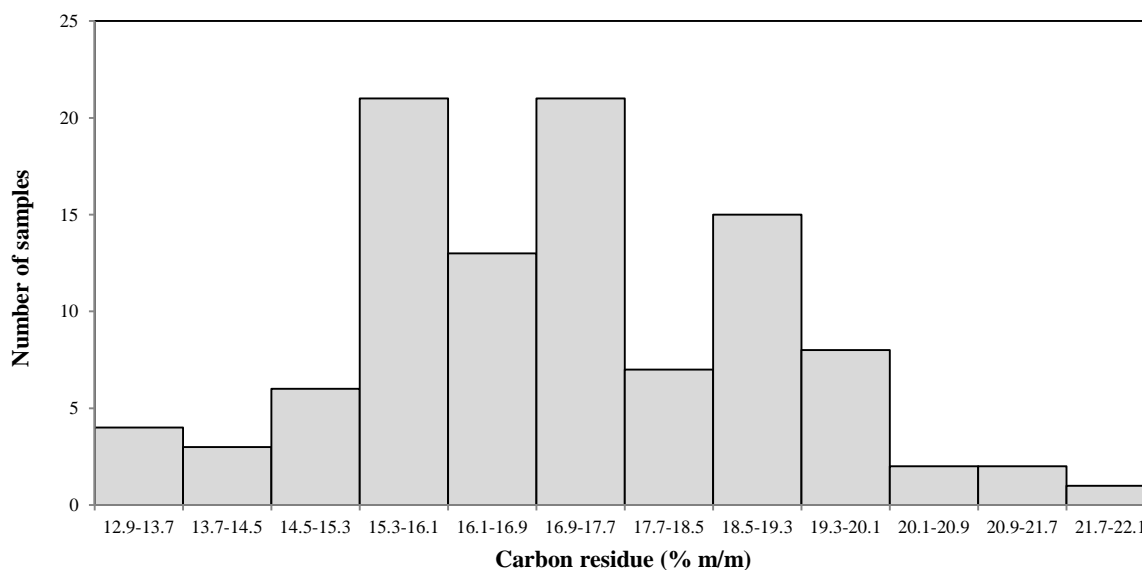


Figure 4.13: Histogram for 142 samples on the variation of HVGO carbon residue.



**Figure 4.14:** Histogram for 103 samples on the variation of fuel oil carbon residue.

#### 4.2.4. Sulfur Content (IP 336)

The IP 336 method (IP Standards 336, 2004) is used in the determination of sulfur content in products having sulfur contents in the range 0.03 (% m/m) to 5.00 (% m/m) such as naphthas, distillates, fuel oils, residues, lubricating base oils and unleaded gasolines. In this work the IP 336 method was used to determinate the sulfur content in fuel oils. LVGO and HVGO samples were not analysed by the IP 336 method once the sulfur content in these samples was expected to be very small. It is in the higher molecular weight fractions, such as the residue, that the greater concentration of sulfur compounds is located. Generally, the sulfur compounds found in crude oils have tendency to concentrate in residue. Farther, to determine the sulfur content of the fuel oil samples, the energy-dispersive X-ray fluorescence analyzer (Figure 4.15) was used.



**Figure 4.15:** Energy dispersive X-ray fluorescence analyzer, Oxford Instruments, Lab-X3500SCL (IP Standards 336, 2004).

In this procedure, a sample cell equipped with a thin transparent film was filled with the sample for the X-ray. The sample cell was placed in an analyzer detector in the beam emitted from a suitable low energy radioactive source, such as  $^{55}\text{Fe}$  source, and consequently irradiated by primary X-radiation. The basic operation principle of this analyzer is related to the X-ray fluorescence analytical method. This method explains that, due to the effect of the radiation, some electrons of the internal shells are ejected by the exciting radiation and with that they go to higher energy levels. Due to the movement of the electrons to other energy levels, electrons coming from higher levels occupy the “holes” create when the exited electrons turn away. This electronic rearrangement is responsible for the emission of the secondary X-radiation whose wavelengths are characteristic in the bombarded element. Each element can represent emissions of many different wavelengths depending on the internal electronic transitions. When electrons are ejected by the exciting radiation, the electrons from higher energy levels “fall down” to take the void place of  $K$ ,  $L$  and  $M$  left by the electrons (as shown in Figure 4.16). This means that an emission  $K_{\alpha}$  results from the movement of the electron arriving at the  $K$  shell and departing from the  $L$  shell. The index  $\alpha$  refers that the origin of the transition was at a level above. In  $\beta$  and  $\gamma$  index the origins of the transition were at two and three levels above, respectively. The sulfur content was analysed on the  $K_{\alpha}$  emission at  $5.573 \text{ \AA}$  (Wauquier, 1995). Figure 4.17 presents the range of sulfur variations of fuel oil samples.

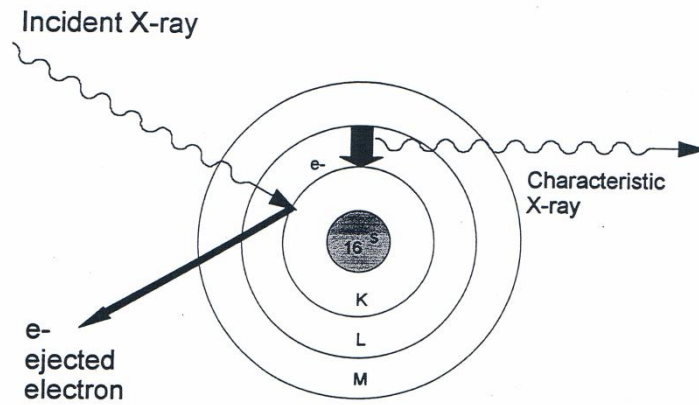


Figure 4.16: Fluorescence of sulfur atom.

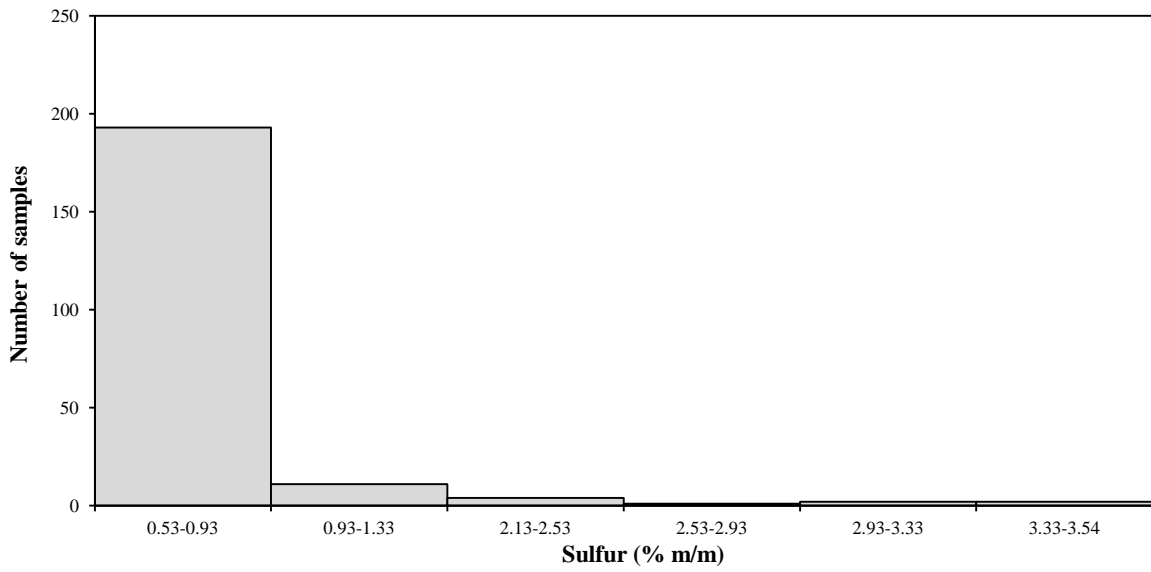


Figure 4.17: Histogram for 213 samples on the variation of fuel oil sulfur content.

#### **4.2.5. Flash Point (ASTM D93)**

The standard test method for flash point by Pensky-Martens closed cup tester (ASTM Standards D93, 2011) is the test method used in the determination of the flash point of petroleum products, at a temperature range from 40 to 360 °C, and of biodiesel at a temperature range of 60 to 190 °C. The flash point is an indispensable property that should be used for evaluating the flammability of a petroleum cut, since it is described as a measure in the tendency of the material to form a flammable mixture with air. The procedure in this method consisted in filling the brass test cup up to the inside mark with the analysed sample. After that the sample was heated and stirred at a specified rate in the automated apparatus – HFP 360 Flash Point Analyzers (Figure 4.18). Then an ignition source in the form of a small flame directly into the test cup at regular intervals was applied. When a flash was seen in the sample the flash point was found. The temperature at the moment of the flash was noted as the flash point of the sample. Figure 4.19 presents the results obtained with the application of the ASTM D93 on fuel oil samples. The flash point of LVGO and HVGO samples was not determined. Besides the important information obtained with the analysis of the flash point, this property was not used in LVGO and HVGO characterization once the determination of other properties were considered more significant to the required evaluation.



Figure 4.18: Pinsky-Martens closed flash tester, HFP 360 Flash Point Analyzers (ASTM Standards D93, 2011).

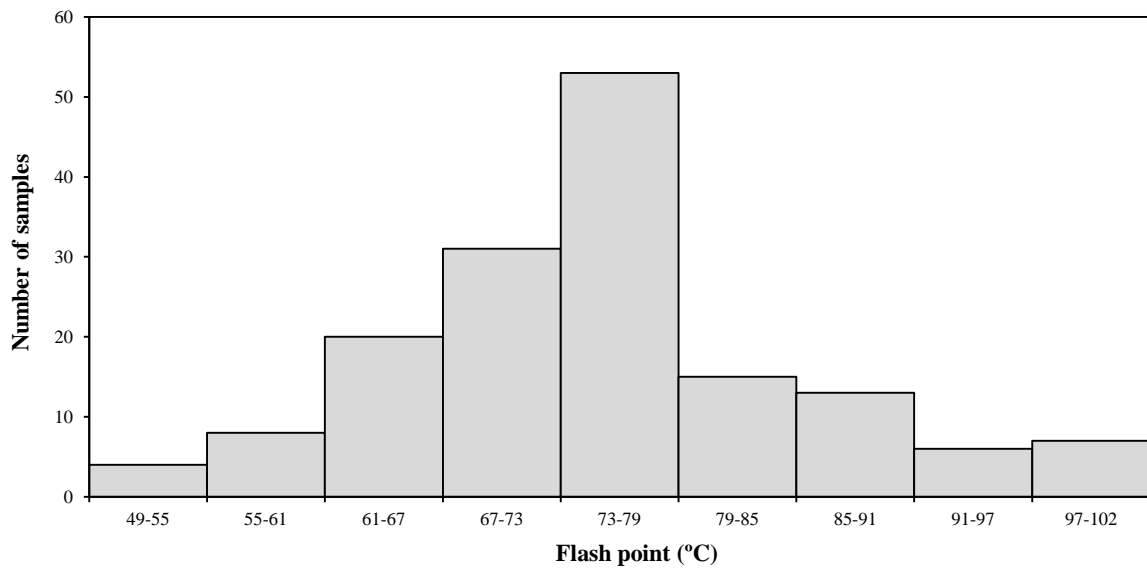
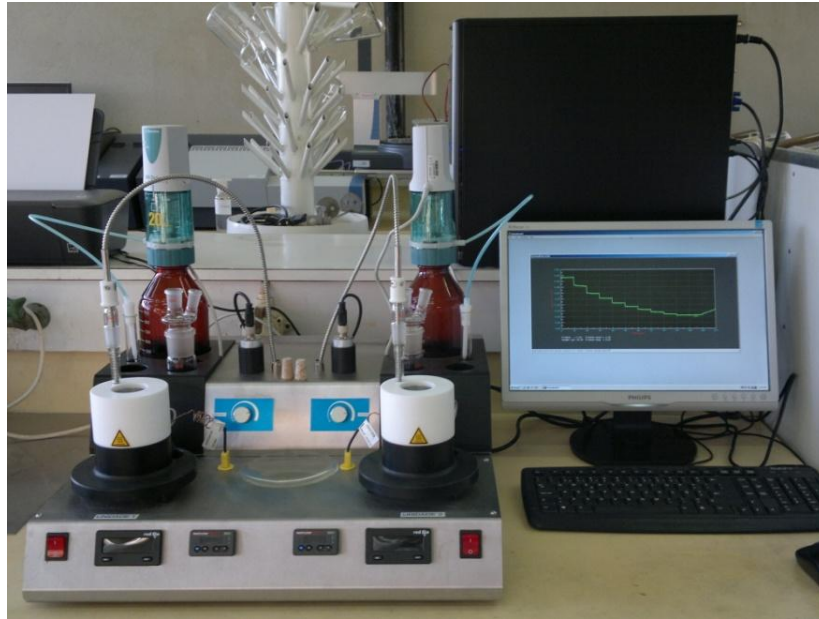


Figure 4.19: Histogram for 157 samples on the variation of fuel oil flash point.

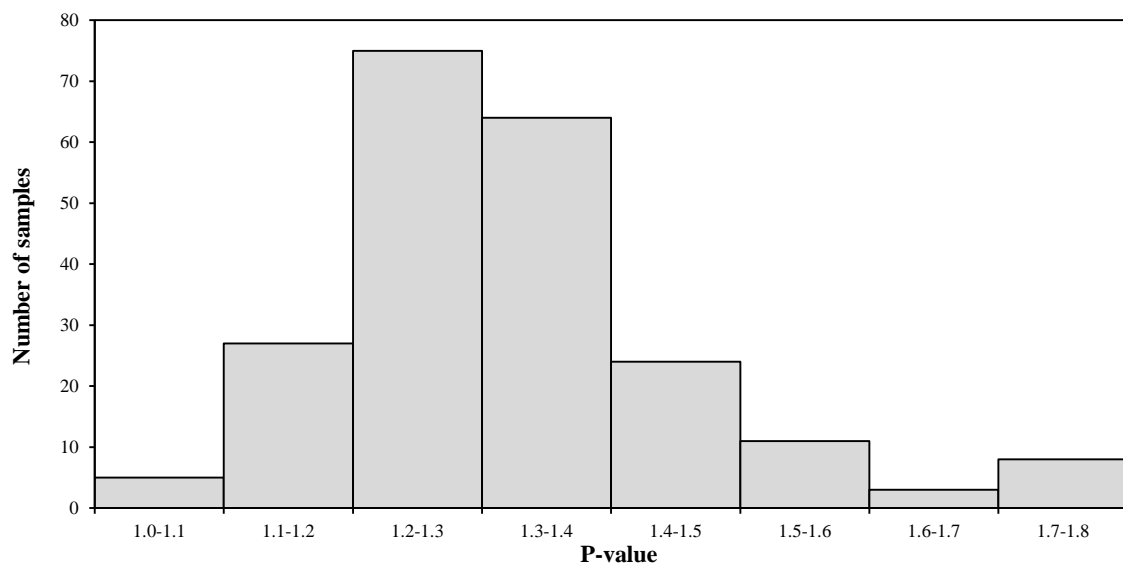
#### 4.2.6. P-Value (SMS 1600)

The P-value, also called peptization value test, allows obtaining information on stability and stability reserve required in refinery process control. This stability is directly related with the asphaltenes precipitation and the data obtained with this method will avoid undesirable asphaltenes precipitation and will improve production efficiencies. The SMS 1600 – *state of peptization of asphaltenes in heavy oil streams* – (Shell Method Series 1600, 2001) describes the tendency of asphaltenes precipitation in residues and residual fuel oils having a viscosity value less than  $2 \times 10^6$  cSt at 50 °C. In the realization of this method an automated procedure using a commercially available automated stability analyzer (Zematra ASA Automated Stability Analyzer; Figure 4.20) was used. An amount of the sample (approximately 12 g), heated and stirred at specified rates in the automated apparatus, was diluted with cetane (hexadecane). The quantity of cetane used depended on the quantity required to reach the point of instability, which meant, that the sample become unstable after the injection of a measured quantity of cetane. The point of instability (asphaltene flocculation point) was determined by measuring the reflection of light by the asphaltenes that had precipitated on the glass surface in the optrode device. The P-value (stability reserve) was automatically calculated (Figure 4.21) taking into account the number of mL of cetane with which one gram of the sample can be diluted until it doesn't flocculate the asphaltenes ( $X_{min}$ ):

$$P - value = 1 + X_{min} \quad (4.3)$$



**Figure 4.20:** Zematra ASA automated stability analyzer for residual refinery streams and fuel oils (Shell Method Series 1600, 2001).



**Figure 4.21:** Histogram for 213 samples on the variation of fuel oil P-value.



#### 4.2.7. Distillation at Atmospheric Pressure (ISO 3405)

To determine the distillation of LVGO products the ISO 3405 (ISO Standards 3405, 2011) – *determination of distillation characteristics at atmospheric pressure* - was used. This international standard method can be used to determine the distillation characteristics of light and middle distillates derived from petroleum and having initial boiling points above 0 °C and end points not above 400 °C.

An automated device, Herzog Atmospheric Distillation HAD 627, was used in the determination of LVGO distillation (Figure 4.22). This method used a standard round bottom distillation flask with a capacity of 125 mL attached to a water cooled condenser. The thermometer bulb was introduced into the top of the distillation flask using an appropriate snug-fitting device that held the temperature sensor in the middle of the neck in the distillation flask. A 100 mL of the sample was placed in the flask and heated by a small gas flame. The temperature of the initial boiling point was read when the first condensate drop entered the receiving cylinder at the end of the condensate bath. The test was continued at a constant rate and systematic observations of thermometer readings and volumes of condensate recovered were made at determined recovery levels of condensation. The final boiling point, the highest temperature observed, was also recorded (ISO Standards 3405, 2011).



**Figure 4.22:** Herzog atmospheric distillation HAD 627 apparatus (ISO Standards 3405, 2011).

#### 4.2.8. Distillation at Reduced Pressure (ASTM D1160)

The ASTM D1160 – *Standard test method for distillation of petroleum products at reduced pressure* – was used for determining the range of boiling points in HVGO products (ASTM Standards D1160, 2006). This method can also be used in the determination of the distillation characteristics of petroleum products that can be partially or completely vaporized at a maximum liquid temperature of 400 °C at controlled pressure between 1 and 50 mmHg. In this method, a 200 mL of the sample was weighed, into the distillation flask. The distillation apparatus, Herzog Vacuum Distillation HDV 632, (Figure 4.23) was evacuated until the pressure reached the prescribed level for distillation. After the desired pressure level was obtained, the heat was applied to the flask as quickly as possible taking into account that undue foaming of the sample cannot occurred. When vapor or refluxing liquid appeared at the neck of the flask, the heating rate was adjust to recover at 6-8 mL/min until the distillation was completed. Data from vapor temperature, time and pressure at different volume percentage fractions of the charge collected in the receiver were recorded.



**Figure 4.23:** Herzog vacuum distillation HDV 632 apparatus (ASTM Standards D1160, 2006).

All procedures previously described were used to determine the required physical-chemical properties. A summary of all properties analysed with indication of the number of samples studied is presented in Table 4.2. In Table 4.2 it is also indicated the test time required to analyse each sample. This test time comprises the time required to prepare the samples, the time needed to make the analysis using the standard methods and the time required to clean the material used. As visualized, the standard methods used in this work were time-consuming, some of them requiring many hours before the final result were obtained. Carbon residue was the standard method most time-consuming. The analysis of P-value of fuel oil samples could also take, at maximum 3 hours. However, depending on the sample composition this property could be determined in just 1 hour. To make a comparison between all procedures used to analyse each sample, Table 4.2 presents the time required to obtain a  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra. The procedure used to obtain a  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra will be explained in the following section. It was possible to conclude that, from all experimental procedures used, it was quicker to obtain a  $^1\text{H}$  NMR spectrum. It was verified that 25 minutes were enough to obtain a  $^1\text{H}$  NMR spectrum and, when using  $^1\text{H}$  NMR spectroscopy in combination with multivariate models, 25 minutes were enough to obtain the information about all the physical-chemical properties. On the other hand, many hours were needed to obtain a complete quality information of the samples in analysis when following the procedure of the traditional standard methods. In addition, the use of some of these standard methods required extensive handling and sample preparation.

Besides, the identification of the properties studied, the number of samples analysed and the time required to develop a procedure, in Table 4.2 the minimum and maximum value obtained for each property and for each sample is also shown.

**Table 4.2:** Number of samples, test time and minimum and maximum results for properties that were analysed.

	number of samples analysed			test time	results [minimum;maximum]			
	LVGO	HVGO	Fuel Oil		LVGO	HVGO	Fuel Oil	
<b>Properties</b>	Kinematic viscosity 50 °C (cSt) - ASTM D445		108	≅ 1 h 20 min			[103.2;890.4]	
	Kinematic viscosity 100 °C (cSt) - ASTM D445		168	217	≅ 1 h 20 min		[8.050;18.60] [14.20;1367]	
	Density (g/mL) - ASTM D4052	105			≅ 20 min	[0.8592;0.9050]		
	Density (g/mL) - ASTM D5002		189	185	≅ 20 min		[0.9152;0.9609] [0.9487;1.019]	
	Carbon residue (% m/m) - ASTM D4530		142	103	≅ 3 h		[0.08;1.03] [12.9;22.1]	
	P-value - SMS 1600			217	≅ 3 h		[1.00;1.8]	
	Sulfur content (% m/m) - IP 336			213	≅ 20 min		[0.53;3.54]	
	Distillation (°C) - ISO 3405	105			≅ 2 h	*		
	Distillation (°C) - ASTM D1160		116		≅ 2 h		*	
	Flash point (°C) - ASTM D93			157	≅ 40 min		[49;102]	
	<sup>1</sup> H NMR spectra	105	214	222	≅ 25 min	<sup>1</sup> H NMR spectrum		
	<sup>13</sup> C NMR spectra			63	≅ 8 h 20 min	<sup>13</sup> C NMR spectrum		

(\*) The results of the distillation depends on the distillation cut.

### 4.3. NMR Experiments

The NMR spectroscopy was used to chemically characterize all the samples previously analysed by means of the ASTM standard methods. The main idea was to carry out a quantitative analysis of all samples. To obtain high quality spectra, the NMR spectra were recorded with a Bruker Avance 300 spectrometer (Figure 4.24) operating at atmospheric temperature and 300.13 MHz for proton and 75.47 MHz for carbon.  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra were recorded in the work hereby reported.

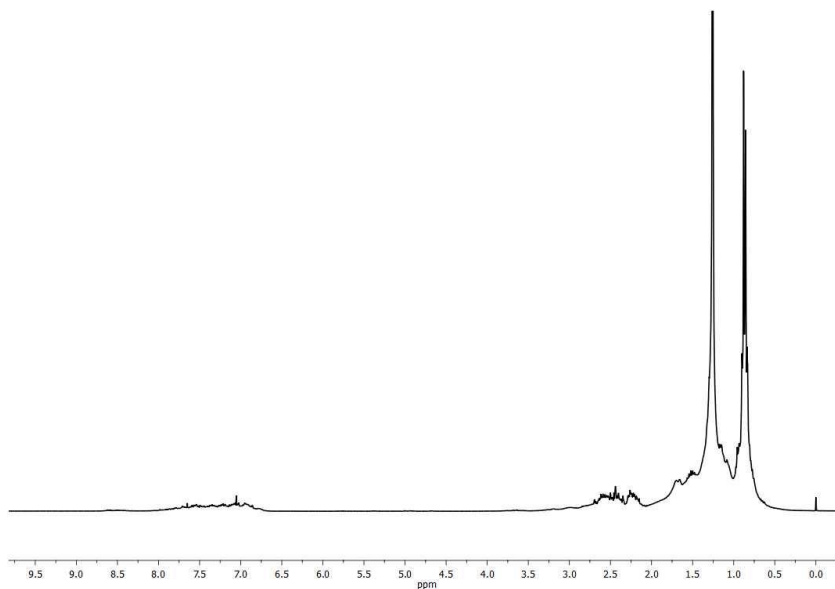


Figure 4.24: Bruker Avance 300 NMR spectrometer.

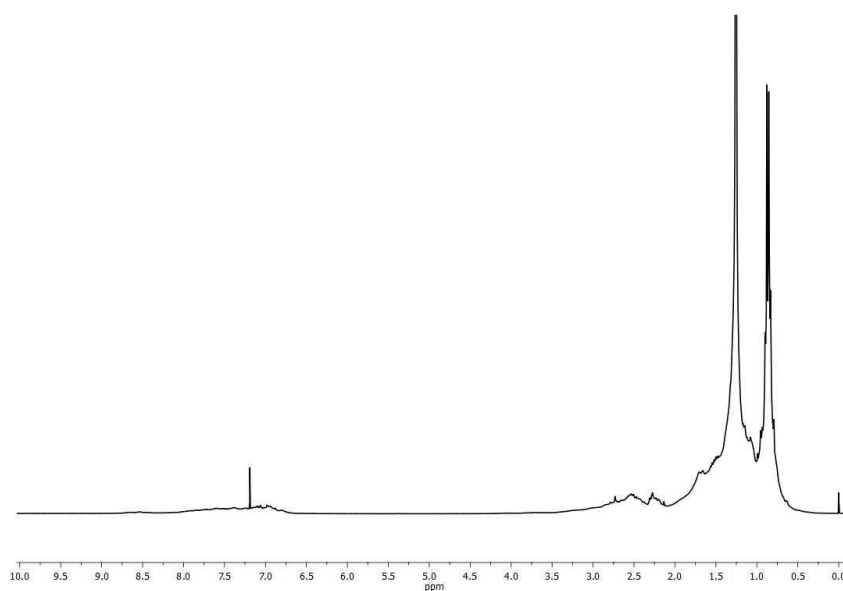
#### 4.3.1. $^1\text{H}$ NMR Experiments

The  $^1\text{H}$  NMR spectra were recorded using the following conditions: spectral width 6887 Hz, with a  $30^\circ$  pulse width ( $3.63 \mu\text{s}$ ), 2.4 s acquisition time, 1 s relaxation delay, 32768 K data points and 128 scans. Very good signal to noise (S/N) ratio was obtained under these experimental conditions. The NMR spectrum was acquired using a high quality tube with 5 mm diameter. The HVGO and fuel oil samples were prepared diluting 0.2000 g of the sample into 1 mL of deuteriochloroform ( $\text{CDCl}_3$ ) while for LVGO samples 0.35 mL of  $\text{CDCl}_3$  was used to dissolve 0.25 mL of the samples. In both preparations, an internal reference such as the tetramethylsilane [ $\text{Si}(\text{CH}_3)_4$ ], commonly abbreviated as TMS, was

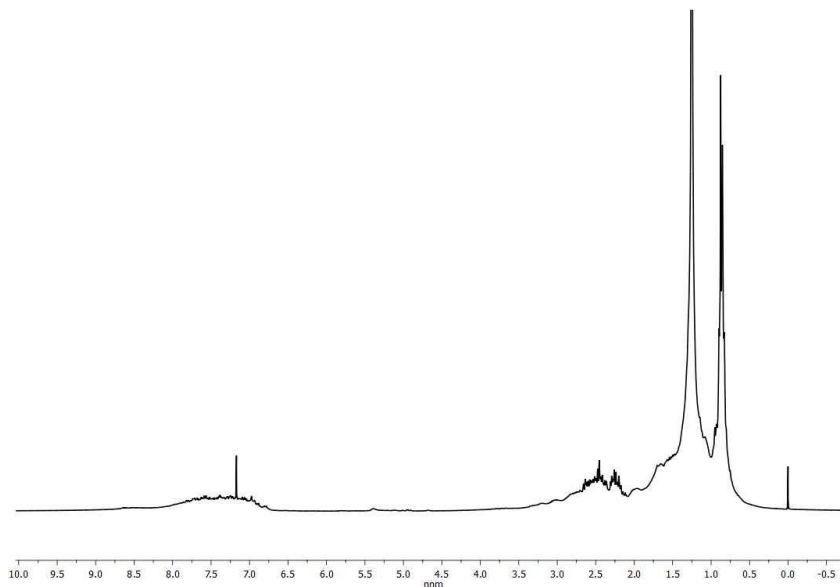
used. This internal reference defines the zero of the chemical shift scale and thus improves the precision of the quantitative and qualitative analysis. An example of LVGO, HVGO and fuel oil spectrum is presented in Figure 4.25, Figure 4.26 and Figure 4.27, respectively.



**Figure 4.25:** Example of a LVGO  $^1\text{H}$  NMR spectrum.



**Figure 4.26:** Example of a HVGO  $^1\text{H}$  NMR spectrum.



**Figure 4.27:** Example of a fuel oil  $^1\text{H}$  NMR spectrum.

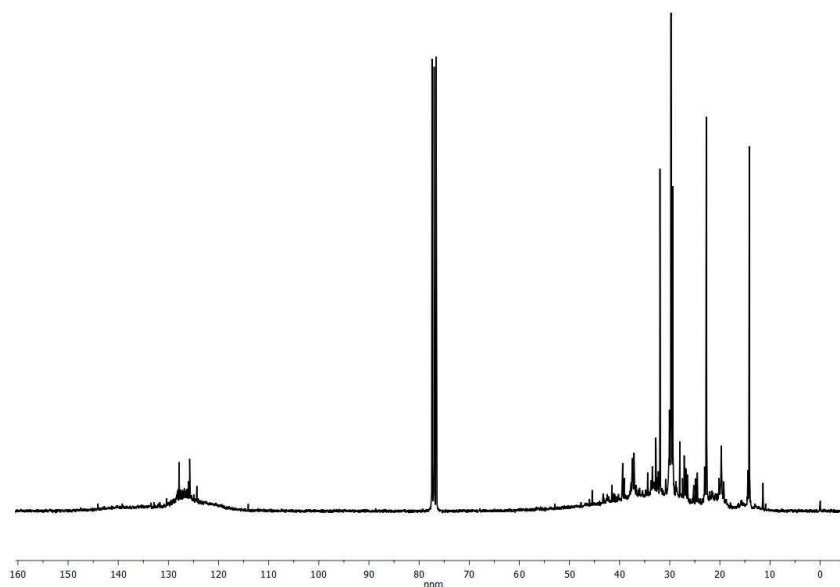
Apparently, it is not possible to highlight the main differences between the three spectra here presented. All spectra, LVGO, HVGO and fuel oil spectrum, present an aromatic and an aliphatic region. However, to indicate which of the samples present a high quantity of aromatic or aliphatic hydrogens a more detailed analysis of the spectra should be performed. For that, the integration of the spectrum is required. Such comparison is performed and discussed in Chapter 6.

#### 4.3.2. $^{13}\text{C}$ NMR Experiments

To obtain a quantitative spectrum, an inverse gate decoupling  $^{13}\text{C}$  NMR spectrum was used. The important parameters used in recording such  $^{13}\text{C}$  NMR spectrum were the following: i) spectral with 20000 Hz; ii) with a  $30^\circ$  pulse width (1.5  $\mu\text{s}$ ); iii) 8 h acquisition time; iv) 6 s relaxation delay; v) 32 K data points; and vi) 4224 scans. As in the  $^1\text{H}$  NMR experiments, very good signal to noise (S/N) ratio was obtained under the previously described experimental conditions. The  $^{13}\text{C}$  NMR solution was prepared using the same tube as the ones used in the  $^1\text{H}$  NMR experiments, a tube with 5 mm diameter. The samples analysed by  $^{13}\text{C}$  NMR spectroscopy were the fuel oil samples. The samples were prepared by dissolving 0.2000 g of the sample into 1 mL of deuteriochloroform ( $\text{CDCl}_3$ ).

The tetramethylsilane (TMS) as internal reference was also used. In contrast to  $^1\text{H}$  NMR spectroscopy, a small amount of paramagnetic relaxation reagent such as trisacetylacetonatochromium(III)  $[\text{Cr}(\text{acac})_3]$  was added into the NMR solution. The idea of using the  $\text{Cr}(\text{acac})_3$  was to reduce the long relaxation times of carbons and to avoid the nuclear Overhauser enhancement (NOE), related to the increase in signal intensity when C-H coupled protons were saturated by the decoupling field (Altgelt *et al.*, 1994). The use of  $\text{Cr}(\text{acac})_3$  contributed in obtaining a quantitative spectrum (as example, Figure 4.28), which the advantage was the peak areas, in different molecular positions, were proportional to their concentration, which didn't occur in a normal  $^{13}\text{C}$  NMR spectrum.

As will be presented in Chapter 6, the use of  $^{13}\text{C}$  NMR spectroscopy to provide the physical-chemical properties of the analysed samples will be not considered in this work. The main reason for not using the  $^{13}\text{C}$  NMR was the time required to obtain a quantitative result necessary to develop reliable models (see Table 4.2, page 126). Since  $^{13}\text{C}$  NMR spectroscopy will not be considered in this work, the analysis of the LVGO and HVGO samples were not made by  $^{13}\text{C}$  NMR spectroscopy.



**Figure 4.28:** Example of a fuel oil inverse gate decoupling  $^{13}\text{C}$  NMR spectrum.



As occur for the  $^1\text{H}$  NMR spectrum, to comment the  $^{13}\text{C}$  quantitative NMR spectrum (Figure 4.28) the integration of the spectrum is required. In Chapter 6 the integration of the  $^{13}\text{C}$  NMR spectrum is presented in an attempt to understand fuel oil composition.

### 4.3.3. Pre-Measuring Steps

As it is well known the quality of the NMR data depended on some experimental factors that need to be controlled and respected such as: i) sample preparation; ii) data acquisition; iii) pulse sequence; iv) optimization of instrumental parameters; and v) pre-processing data.

#### Sample preparation

As mentioned before, there were some precautions which had to be taken during the preparation of the sample. The quantity of each constituent that formed the final sample, the crude oil fraction, the solvent and the paramagnetic relaxation reagent, when used, need to be measured with some precision for obtaining adequate quantitative measurements. It was important that all samples, especially fuel oil samples, should be filtered before placed in the NMR sample tube to avoid the presence of solid impurities that will degrade the homogeneity of the magnetic field. The selection of the solvent used was also very important. The solvent used in all preparations, as referred, was the deuteriochloroform ( $\text{CDCl}_3$ ) once it had a high dissolving power (Christian *et al.*, 1986) and the spectrometer needed a deuterium ( $^2\text{H}$ ) signal to “lock” the magnetic field strength and prevent it from changing with time. TMS was chosen as an internal reference since it is an inert, a low-boiling liquid, and gives rise to a single line in the spectrum. Such facts occur due to all the protons in the symmetrical molecule having an identical environmental, and all protons in TMS resonate at a higher field than practically any other proton (Christian *et al.*, 1986; Claridge, 1999).

### Computer controlled NMR spectrometers

As seen in Section 3.2.3, the NMR spectrometer was controlled by a computer which was used to collect, digitize and perform the Fourier transformation of the signals. There were some experimental variables that were digitally defined at the beginning of the experiment. These variables were the size of the data table ( $N$ ), the acquisition time ( $AT$ ), in seconds, and the spectral width ( $SW$ ), in Hertz, of the transformed spectrum. The data table can be expressed as followed:

$$N = 2AT \times SW \quad (4.4)$$

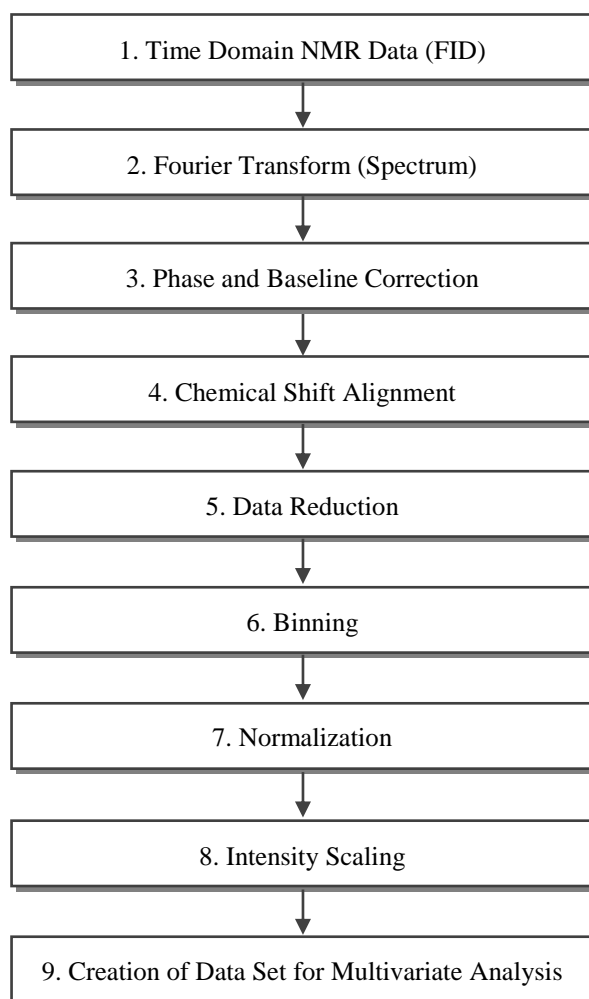
with  $N$  related to the number of addresses to be filled and responsible with the limits for the eventual resolution attainable (Christian *et al.*, 1986). Additionally, there were other variables with high importance such as the pulse width and the delay time. The pulse width was defined as the time, in microseconds, during which the radio frequency pulse was applied. The delay time was referred to the time, in seconds, between pulses sequences and can be used to allow nuclei to reach magnetic equilibrium or “relax”.

### Instrumental variables

There were many instrumental variables that were controlled such as the amplification, noise filtering, phasing of signals, width of sweep, adjustments of field homogeneity and the amplitude of the radio frequency radiation. This last variable, the amplitude of the radio frequency radiation, when too high can be responsible for increased sensitivity. For this reason the amplitude of the radio frequency radiation was set at a value that only saturates some of the resonances, in a real sample, over the finite range of relaxation times (Christian *et al.*, 1986).

#### 4.3.4. Data Pre-Processing

In this section some of the pre-processing and data manipulations used to improve the performance of the subsequent multivariate data methods are reviewed. This pre-processing of the NMR data is always necessary if quantitative NMR spectroscopy is desirable. A set of necessary processing steps, as shown in Figure 4.29, must be implemented to minimize or eliminate artifacts (resulting from experimental variations) and to enhance resolution or signal to noise ratio (Becker, 2000). To apply these pre-processing steps it was important to obtain the spectrum that was acquired with the application of the Fourier transform to the FID signal. After that, the spectrum can be manipulated or transformed to produce the data set for multivariate analysis. These different types of data manipulations are described:

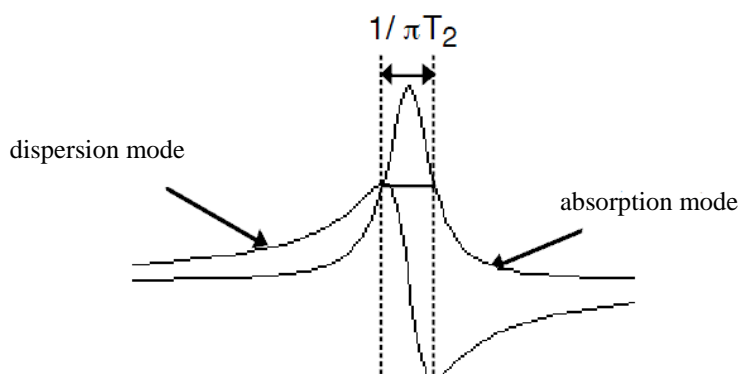


**Figure 4.29:** Example of a flow chart of the different steps that can be utilized during the analysis of NMR spectral data.

#### 4.3.4.1. Phase and Baseline Correction

The main difficulties encountered in the analysis of NMR data were related with the baseline distortions and variations in the spectral phasing between spectra. These variations could be the result from chemically induced changes and/or instrumental and temperature instability, field inhomogeneities and sample susceptibility effects (Alam *et al.*, 2004). If in each single spectrum no correction was made, difficulties in accurately determining peak areas (quantification) and problems in classification when using some multivariate methods could be found.

As previously stated, the phase correction of the spectrum was indispensable to optimize the appearance of the overall spectrum and to obtain the desired peak shape (Jacobsen, 2007). After Fourier transform, the signals of the spectrum could be decomposed into absorptive and dispersive signals as indicated in Figure 4.30. The application of the phase correction was needed to yield purely absorptive signals (absorption mode). The NMR spectra were preferred in the absorption mode as it comprises a higher resolution and a higher proportionality between the integral of peaks and the number of nuclei which cause the signal, when compared with the dispersion mode.



**Figure 4.30:** Peak signal in pure absorption or dispersion mode (Zerbe, 2010).

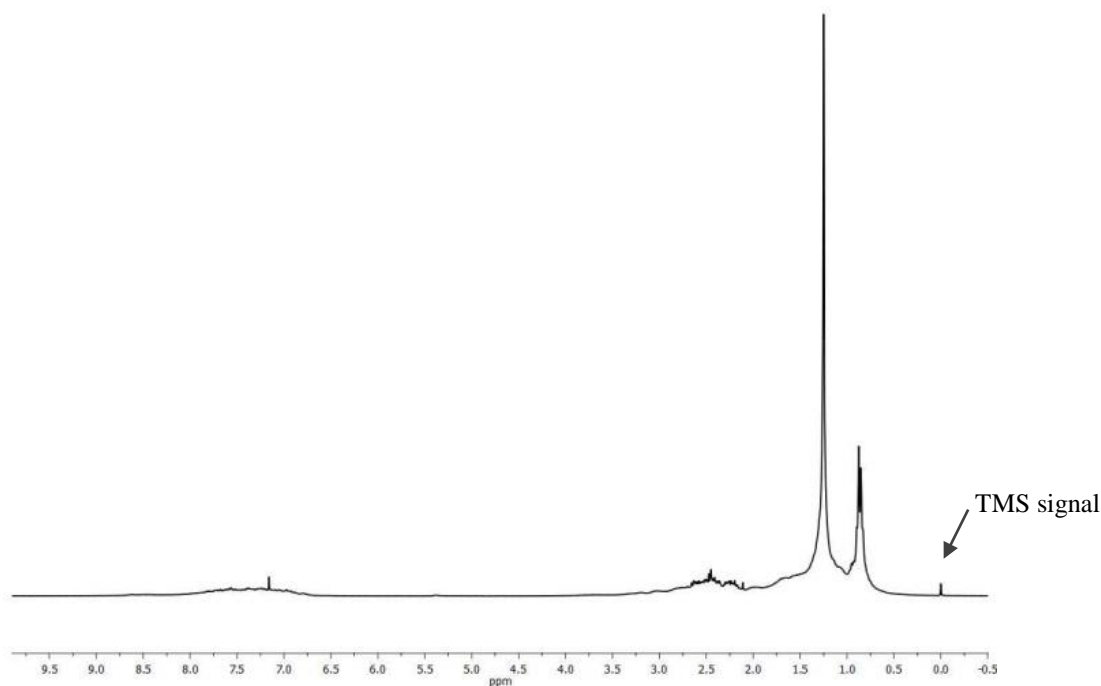
The baseline correction was also very important to obtain all signals connected by a flat line and with that, obtained a good quality spectrum. Besides that, a good baseline gave good results in quantitative measurements where integral calculations were required

(McKenzie *et al.*, 2011). The baseline can be defined as the average of the noise part of the spectrum that should ideally be straight and a horizontal line representing zero intensity.

Phasing and baseline distortions affect the quality of the spectra, reason why it is important to adjust and correct the spectra, automatically or manually. In the past, manually correction was used but soon standard commercial and public domain NMR software programs including methods to automatically correct these variations were developed. In this study, the phase and baseline were automatically adjusted and corrected by using the MestreNova 5.0.3 (Mnova) software (Mestrelab Research - Chemistry Software Solutions, 2012).

#### 4.3.4.2. Chemical Shift Alignment

All spectra obtained were aligned to the TMS signal at zero ppm (Figure 4.31). This procedure, the alignment of the spectrum to a reference peak, was important to acquire a useful chemical shift scale and to compare all spectra considered.



**Figure 4.31:**  $^1\text{H}$  NMR spectrum of a fuel oil with reference of the TMS signal.

To ensure that all spectra analysed had the TMS signal at zero ppm, a simple procedure implemented in Matlab 2011a (Mathworks Inc., Natick, MA, USA) was used to automatically adjust the TMS signal at 0.0 ppm. This type of chemical shift alignment was always used in this work due to the good results obtained. However other procedures were also tried and implemented, such as:

- alignment of the spectra to the solvent peak as a chemical shift reference;
- alignment of the spectra to another important and non-sensitive signal such as CH<sub>3</sub> signals;
- alignment of the spectra according to a reference spectrum;
- alignment of the aromatic region to the solvent peak (CDCl<sub>3</sub>) and the aliphatic region to the TMS signal.

These procedures were not used during the pre-processing of the data once the results obtained when applying the multivariate data methods were not as good as when compared with the alignment of the spectrum to the TMS signal.

In the scientific literature there are some studies referring the use of chemical shift alignment in terms of peak position (Stoyanova *et al.*, 2004; Forshed *et al.*, 2005; Guo *et al.*, 2008; Savorani *et al.*, 2010; McKenzie *et al.*, 2011). Example of some methods are the partial linear fit (PLF) (Forshed *et al.*, 2005), dynamic time warping (DTW) (Savorani *et al.*, 2010; McKenzie *et al.*, 2011), the correlation optimized warping (COW) (Savorani *et al.*, 2010; McKenzie *et al.*, 2011), the multiplicative signal correction (MSC) (Kramer *et al.*, 2008) and the icoshift algorithm (Savorani *et al.*, 2010). In this study, the multiplicative signal correction (MSC) algorithm was applied. With the MSC each spectrum was regressed against a “standard” spectrum and compared one by one with the line segments of the “standard” spectrum. Each spectrum was then shifted and aligned with the “standard” spectrum (Kramer *et al.*, 2008). However, this type of algorithm was not used in this study due to the unsatisfactory obtained results after the multivariate data methods, when comparing with the alignment of the spectrum to the TMS signal. This lead to conclude that the use of peak alignment algorithm in these types of samples, which forces

the variation of the peak position, can slightly reduce the predicted ability of the model developed and should only be employed when strictly necessary.

#### **4.3.4.3. Data Reduction**

Data reduction is used to advantageously reduce the unused or undesired spectral regions. It can be used to remove just a signal or a spectral area with some noise or a non-viable region. In this research, data reduction was used: i) to remove the TMS signal that didn't bring chemical information; ii) to remove the signal of the solvent used ( $\text{CDCl}_3$ ); and iii) to remove the region between 4.5 and 6.3 ppm since it was a region where no important information occurred and where spectral signatures of additives or other chemical species appeared.

#### **4.3.4.4. Binning**

Binning process has been considerably used for NMR data especially in the development of multivariate classification models. Binning involves the calculation of peak areas within specified segments of a spectrum (binning or bucketing) (Craig *et al.*, 2006; McKenzie *et al.*, 2011) becoming the resulting "spectrum" the input variable instead of the real NMR spectrum. It is evident that there is a reduction in the number of variables in the resulting "spectrum" and a reduction in the spectral resolution. This procedure is a rapid and consistent method that simplifies the data by reducing its size and smoothes out alignment errors between different samples. It also contributes to the elimination of the impact of small chemical shift variations (especially due to concentration) between different samples on the subsequent multivariate analysis. The most frequently bin size used in the NMR spectra is a bin width in the range of 0.02 to 0.04 ppm (Alam *et al.*, 2004; Flumignan *et al.*, 2012). In this work, the most frequent used bin was 0.02 ppm due to better results that were obtained after applying multivariate data methods.

#### 4.3.4.5. Normalization

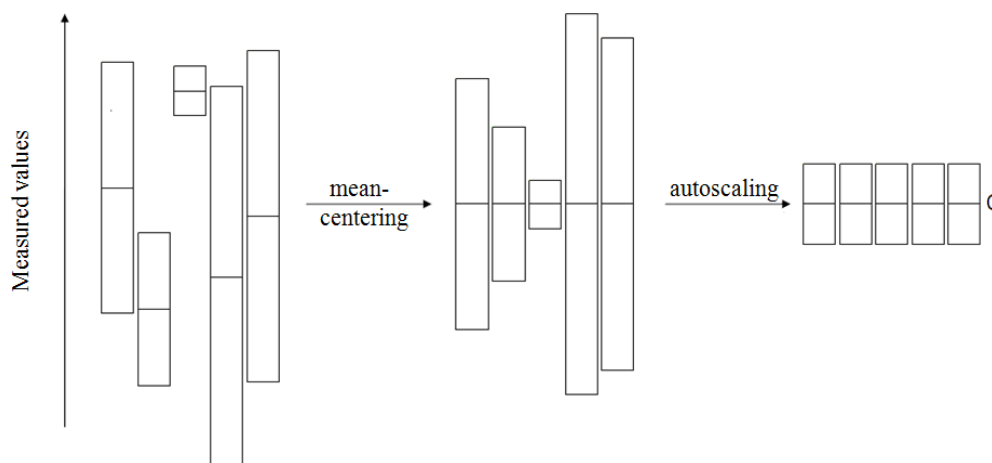
Normalization methods are data pre-processing approaches used when comparisons between spectra are required. Normally, in optimal conditions, the NMR peaks integrals are directly proportional to concentrations, in quantitative analysis. However, there are some experimental and instrumental variables that can affect the absolute NMR peak intensities. For this reason the normalization, to remove or minimize the effects of such variables, is used. In the present work, the normalization of the data used involved setting each spectrum to have unit total intensity by expressing each spectrum segment as a fraction of the total spectrum integral. Other alternative normalization approach, such as, normalization by the maximum peak of the spectrum and normalization by a given area (0.25 to 0.27 ppm) were also applied. The best results were obtained when using the normalization by the total spectrum integral.

#### 4.3.4.6. Intensity Scaling

Scaling is a part of multivariate analysis which can significantly improve the rate of identification and separation of groups during classification, normalizing the relative importance of each variable (Alam *et al.*, 2004). Scaling should be used when there is no information about variable importance and, as previously stated, to regulate the importance of each variable in the model due to the existence of largest and smallest spectral features. When no scaling is applied the largest spectral features, which are not necessary more important than those present in a lower concentration, can obscure systematic variation of interest in the low intensity regions (Alam *et al.*, 2004). In the literature it is possible to find some scaling methods including mean-centering (Cloarec *et al.*, 2004; Craig *et al.*, 2006), variable stability (VAST) scaling (Keun *et al.*, 2003), scaling to unit variance (autoscaling) (Craig *et al.*, 2006) and Pareto scaling (Keun *et al.*, 2003). In this study, mean-centering scaling was used to emphasize weaker spectral component and the autoscaling was also used to obligate the variables to have equal probability of influencing the model (Figure 4.32). With the mean-centering the average value of each column was calculated and then each column could be given a mean of zero by subtracting the column mean from each value in the column. While, applying the autoscaling meant scaling each



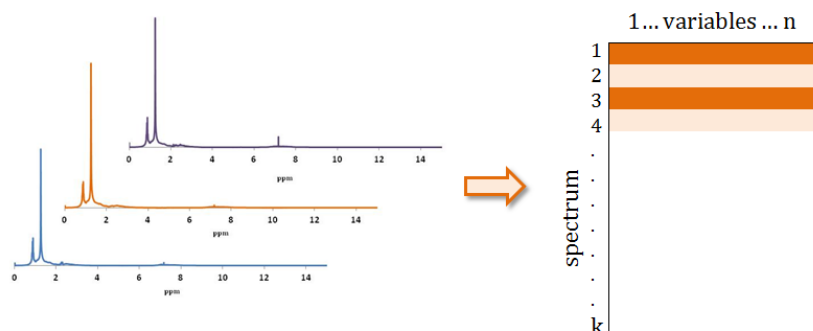
column to have unit variance by dividing each variable in the column by the standard deviation of the column and, therefore, giving equal weight to each variable.



**Figure 4.32:** Representation of the scaling methods used: after mean-centering and autoscaling all variables will have mean zero and unit variance. Adapted from Eriksson *et al.* (2006).

#### 4.3.4.7. Creation of Data Set for Multivariate Analysis

After applying all different types of data manipulations to each spectrum, a NMR data matrix composed by the spectra was obtained. Each different spectrum was composed of  $n$  areas for the frequencies within the observable spectral window as shown in Figure 4.33. This NMR data matrix could now be submitted to a multivariate analysis.



**Figure 4.33:** Creation of data set for multivariate analysis after applying data pre-processing to each spectrum.

#### **4.4. Final Remarks**

Analysing the procedures and sample preparation in the methods used during this research it was verified that these methods were elaborate, time-consuming, presenting some of them tedious preparations and requiring large amount of samples when compared with the NMR spectroscopy. As demonstrated, the impact of the pre-processing procedure on the analysis of NMR spectral data can be very significant in the subsequent multivariate analysis. Phase and baseline correction, chemical shift alignment, intensity scaling, normalization, binning and data reduction were the main pre-processing steps used. These different types of data manipulations were demonstrated to be very important and necessary to exploit the information content on the NMR data, increase the prospect of detecting subtle changes in the NMR data and enhance the interpretability of the resulting multivariate models.

# 5

## **PREDICTION USING MULTIVARIATE METHODS**

The difficulty and complexity associated with the interpretation of NMR spectra, from crude oil and derivatives require the development of powerful multivariate models to explore the information of such complex spectra. The use of multivariate models in the study of crude oil spectra is not new, but there are only few researchers using such models in the analysis of crude oils. In addition, studies referring the use of statistical models in the analysis of fuel oil, heavy vacuum gas oil and light vacuum gas oil samples are very scarce.

## **5. PREDICTION USING MULTIVARIATE METHODS**

5.1. State of the Art.....	143
5.2. Methods Applied in this Thesis .....	146
5.2.1. Principal Components Analysis - PCA.....	147
5.2.2. Principal Component Regression - PCR.....	150
5.2.3. Partial Least Squares – PLS.....	151
5.2.4. Artificial Neural Networks – ANN.....	157
5.3. Models Pre-Treatment .....	164
5.4. Performance Indexes .....	168
5.5. Final Remarks.....	173

## 5.1. State of the Art

The analysis of crude oil and derivatives using NMR spectroscopy requires the application of sophisticated chemometric techniques to extract the meaningful information from the complex and highly informative spectra. Normally the information about these types of samples, whose molecular structure of most compounds were very complex, could not be easily assessed due to the many overlapping peaks in the NMR spectra. The development of chemometric techniques overcome this NMR disadvantage and the possibility of applying multivariate methods to establish correlations between the spectral data and sample properties which were determined by the conventional techniques.

The use of chemometric techniques to compute models for extracting chemical information in a NMR crude oil spectrum has quickly become an important tool for the NMR spectroscopy, allowing both qualitative and quantitative information to be obtained from these complex spectral signals. Chemometric analysis are mathematical and statistical methods responsible for decomposing the complex multivariate data into simple and relevant information, focusing on the important variances within the data (Miller *et al.*, 2005; Kiralj *et al.*, 2006). Many developments in the chemometric methods have emerged and the combination of multivariate models and the NMR spectra have been used in the analysis of crude oil (Kvalheim *et al.*, 1985; Molina *et al.*, 2007; de Peinder *et al.*, 2009; Silva *et al.*, 2011; Masili *et al.*, 2012). Although very few, some studies have been found in literature referring the use of statistical models in the analysis of heavy crude oil fractions such as fuel oil (Nielsen *et al.*, 2008). However, to our knowledge, published works using statistical models in the analysis of LVGO and HVGO were not found.

Principal component analysis (PCA), principal component regression (PCR), partial least squares (PLS) and artificial neural networks (ANN) are examples of important techniques of multivariate data analysis that can be applied to spectral data to extract significant information from the complex and highly informative spectra. These techniques allow the construction of models that can predict macroscopic properties of mixtures of crude oils from NMR measurements (Aske *et al.*, 2002).

Very briefly, PCA is a method of data analysis applied in contexts where there are several variables collected in each observation. The idea is to transform all these correlated

variables into a reduced set of variables called the principal components. The main objectives of PCA are to reduce the size of the data and analyse the structure relationship between the variables and observations (Alam *et al.*, 2004; Miller *et al.*, 2005). In simple terms, PCR is a regression model that combines principal component analysis and linear regression. In PCR, some of the principal components are used instead of all original variables, reducing in this way the number of predictor variables (Alam *et al.*, 2004). PLS regression generates and combines characteristics from principal component regression and multiple regression. PLS seeks to estimate the subspace that best explains the variability in the  $Y$ -axis, also describing the variability in the  $X$ -axis. The interpretation of PLS model allows the analysis of the relative importance of variables  $X$  in predicting  $Y$  (McKenzie *et al.*, 2011). ANN regression is a mathematical model that contributes to simulate the structure and/or the functional aspects of the sample. They are usually associated with nodes or processing units and responsible to establish relationships between other units, receiving inputs and sending outputs, finding some patterns in complex data sets (Fraser *et al.*, 1997).

Molina *et al.* (2007; 2010) developed a new PLS based method contributing to the yield analysis and to the analysis of the principal physical-chemical properties of a crude oil mixture. Taking into account some studied physical-chemical properties, like the gravity values ( $^{\circ}\text{API}$ ),  $K_{\text{UOP}}$  factor, wax content, and correlation index (IC) it was confirmed the composition of the crudes studied (paraffinic or aromatic) (Molina V *et al.*, 2010). For example, the crude classified as the most paraffinic was characterized as presenting the highest value of  $^{\circ}\text{API}$  and  $K_{\text{UOP}}$  factor, while the most aromatic one was characterized for having low values of  $^{\circ}\text{API}$  and  $K_{\text{UOP}}$  factor. It was also observed that when the paraffinic hydrocarbons content decreases and the aromatic resins and asphaltenes contents increases, other properties' values increase such as the sulfur, nitrogen, vanadium and nickel content, the percentage of insolubles in  $n$ -pentane ( $n\text{C}_5$ ) and  $n$ -heptane ( $n\text{C}_7$ ) and the percentage of carbon residue (CR). The most naphthenic studied crude has presented the lowest pour point and the highest acid number. The authors justified the achieved correlation between  $^1\text{H}$  NMR spectrum and the content of vanadium and nickel by the PLS method probably due to the presence of paramagnetic metallic centers (for example vanadyl porphyrins). These paramagnetic metallic centers in the molecules affect the chemical shifts of the entire  $^1\text{H}$  NMR spectrum, the area of the spectrum of more distant protons and also the

protons belonging to the closest surrounding molecules. Consequently, their perturbation range is extended further than the molecules containing metallic centers, influencing different types of protons. Using the PLS regression method and the  $^1\text{H}$  NMR spectra of different crudes it was possible to meet the physical-chemical properties and the refining product yields with a high precision.

Nielsen *et al.* (2008) used multivariate models in the study of different heavy crude oil fractions, aiming to identify physical-chemical properties. The physical-chemical parameters analysed were the calculated carbon aromaticity index (CCAI), density at 15 °C, gross (GCV) and net calorific values (NCV), viscosity at 50 °C, ash content (ASH), total sediment accelerated (TSA), carbon residue (CR) plus aluminum, silicon, zinc and water contents. The obtained results, especially of density, showed that high aromatic content is associated with a high density, while a high fraction of aliphatic compounds normally has a lower density. A higher aromatic content means a higher value of calculated carbon aromaticity index (CCAI) of the sample. Relatively to the GCV and NCV, which are very important parameters in the oil quality analysis, it was verified that they have a positive contribution for the signals at 1.5, 0.7 and 0.5 ppm and a negative contribution for the signal at 2.5 ppm in the  $^1\text{H}$  NMR spectra. The MCR parameter, related with the oil stability, can be determined in the  $^1\text{H}$  NMR spectrum by the signals at 4.7-6.2 ppm, having therefore a slighter contribution for the signals at 3.2-3.7 ppm and a negative effect for that at 1 ppm. The sulfur content can be determined by taking into account the region between 2.1 and 3.0 ppm. The water content is predicted near 4.5 ppm for undiluted samples and  $^1\text{H}$  NMR measurements at 50 °C. The combination of the PLS and PCA techniques with the NMR spectroscopy have demonstrated, once again, to be very important in the analysis of physical-chemical properties on the heavy crude oil due to the facility, fast and reliable way to obtain this type of information.

Peinder *et al.* (2009) analysed the potential of PLS to predict some properties of crude oils, such as yield long-on-crude (YLC), density (D), viscosity (V), sulfur content (S), pour point (PP), asphaltenes (Asph) and carbon residue (CR). The results obtained with the multivariate analysis and the traditional methods, like the ASTM and IP, demonstrated the importance of the PLS modelling that gave valuable and fast results. In this research,  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra and also other spectroscopic technique, the IR technique, were used,

primary to the analysis of each single technique and then associated, to determine the properties previously described. It was verified that a single spectrum from different techniques did not lead to any improvement in the results concerning the crude oil properties. The combination of the two techniques did not present additional information once the sample in the analysis was the same crude oil. However, each single spectrum was modeled by PCA model and then compared with the results of the PLS models with the single spectrum of each technique without the PCA method. With that it was concluded that modelling PCA scores instead of all spectrum only contribute to a better results if the  $^{13}\text{C}$  NMR spectroscopy was used.

In 1985, Kvalheim *et al.* (1985) used PCA to help in the NMR interpretation of the crude oil naphtha fraction. The main idea was to analyse samples of the same geographical area aiming to identify correlations between samples from the same source. Dependence between the samples, especially in the composition, was found, for example in long chain versus short chain alkanes, branching, cyclization, and aromatization. The main differences were obtained in the quantities of the different structures.

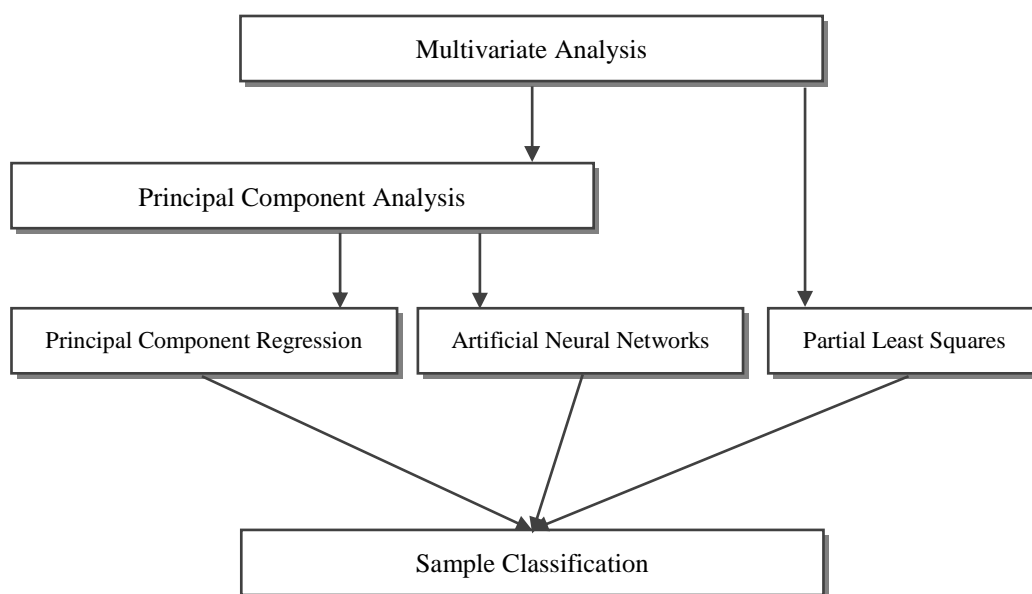
Masili *et al.* (2012) used  $^1\text{H}$  NMR spectroscopy and PLS regression models to predict the properties of 64 crude oil samples. The idea of such investigation was to develop a model that contributed in a fast and low cost effective way to the characterization of the different crude oils. Properties such as density, sulfur content, total acidity number, true boiling point distillation yields and  $K_{\text{UOP}}$  factor were determined using the ASTM standard methods and were the properties predicted by the developed models. Successful results were obtained when using the PLS models to predict such properties.

## **5.2. Methods Applied in this Thesis**

As the main objective of this research was the characterization of heavy residues and oil fractions such as fuel oil, HVGO and LVGO, using NMR spectroscopy, it was important to implement statistical methods to correlate the NMR determinations with the physical-chemical properties used in the characterization of such products. After creating the data set for multivariate analysis, as referred in Chapter 4, the statistical methods for the analysis of NMR spectral data could be applied. The statistical methods used, as shown in



Figure 5.1, were: i) Principal Component Regression (PCR) [based on principal components analysis (PCA)]; ii) Partial Least Squares (PLS); and iii) Artificial Neural Networks (ANN) [based on principal components analysis (PCA)].



**Figure 5.1:** A flow chart of the different multivariate statistical methods used in this work.

### 5.2.1. Principal Components Analysis - PCA

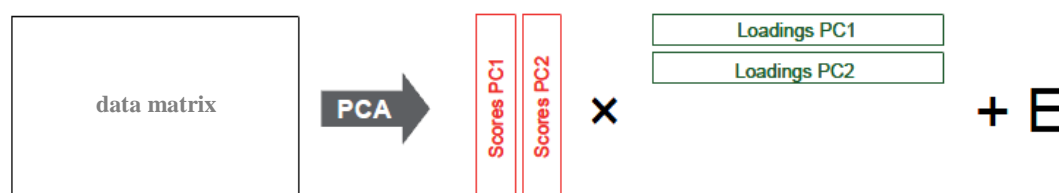
Principal components analysis, also known as eigenanalysis or principal factor analysis (PFA), is a multivariate statistical method of data analysis applied in contexts where there are several variables collected in each observation (Miller *et al.*, 2005). The idea is to transform all these predictive variables into a reduced set of new variables called the principal components  $Z_1, Z_2, \dots, Z_n$  (PCs). The principal components are linear combinations of the original variables describing each specimen  $X_1, X_2, \dots, X_n$ . The first principal component,  $Z_1$  ( $PC_1$ ), describes the first greatest variance and the second principal component,  $Z_2$  ( $PC_2$ ), describes the second greatest, and so on, with the later components describing the smallest variance. These principal components are orthogonal and uncorrelated with each other (Martens *et al.*, 1989). They are calculated by linear combinations of the original variables by means of the following equations:

$$Z_1 = a_{11}X_1 + a_{12}X_2 + a_{13}X_3 + \cdots + a_{1n}X_n \quad (5.1)$$

$$Z_n = a_{n1}X_1 + a_{n2}X_2 + a_{n3}X_3 + \cdots + a_{nn}X_n \quad (5.2)$$

with  $a_{11}$ ,  $a_{12}$ ,  $a_{nn}$  representing the coefficients chosen with the aim to obtain new variables that are not correlated with each other, unlike the original variables, and their variance as large as possible. The principal components contain, in a much smaller number of variables, more than 90% of the information (cumulative percentage variance) that is available in the original data set, meaning that these variables include the greatest variability of the data and exclude the ones that contain modest information.

Geometrically, the PCA is a rotation of the axes where the first principal component is in the direction of maximum variation, the second greatest principal component in the direction of the next greatest variation, and so on. With this, PCA reduces the data matrix from  $(n \times d)$  where  $n$  is the number of samples and  $d$  the number of data points per spectrum to a  $(n \times p)$  matrix with  $p$  representing the number of PCs determined. The sample coordinates on the new PC axes are called scores while the relationship between the new and the original axes are referred as loadings. Consequently, the decomposition of the data matrix occurs as indicated in Figure 5.2. With the scores it is possible to have information relative to the relation between samples variation (concentrations of multivariate), while with loadings it is possible to have an idea about which region of the spectrum represent the highest variation, which means, a description about the PCs and its relation with the original variables (chemical shifts) (Alam *et al.*, 2004; Sousa *et al.*, 2007; Winning *et al.*, 2008; McKenzie *et al.*, 2011).



**Figure 5.2:** Decomposition of the data matrix into scores, loadings and the residual (E).

It is important to have in attention that the principal components are not scale invariant, i.e. they are dependent on the measurement scale in which the original variables are expressed (Miller *et al.*, 2005). Consequently, it is important to analyse both the units used as well as the range of values that the original variables assume and standardized the data before applying PCA, using the scaling methods as described in the Chapter 4. Once again, standardizing the original variables ( $X_j$ ), by using the scaling methods, corresponds in calculating the variable ( $Z_j$ ) values using the following expression:

$$Z_j = \frac{X_j - \mu_j}{\sqrt{Var(X_j)}} \quad (5.3)$$

which, using the mean of each variable ( $\mu_j$ ) and the variance ( $Var$ ), transform all original variables to have a zero mean and unit standard deviation. After applying this method, e.g. after the standardization of the original data, the covariance matrix is determined. Following, the eigenvalues ( $\lambda$ ; Equation 5.4) are determined by using the subsequent expression:

$$|Cov - \lambda I| = 0 \quad (5.4)$$

where  $I$  represents the identity matrix and  $Cov$  the covariance matrix. When the standardization is used, the covariance matrix of the standardized data is equivalent to the correlation matrix of the original data. Using the Equation 5.5, the weights of the variables in the PCs are calculated (Çamdevýren *et al.*, 2005; Sousa *et al.*, 2007).

$$(Cov - \lambda I)W = 0 \quad (5.5)$$

where  $\mathbf{W}$  is the matrix of the weights (or eigenvectors). The principal components, in mathematical terms, are the eigenvectors of the matrix and to each principal component (eigenvector) corresponds an eigenvalue which indicates the amount of variance in the data set explained by the principal component (Miller *et al.*, 2005). Using the original data set standardized and the weights it is possible to obtain the principal components:

$$\text{PC} = Z_j \times \mathbf{W} \quad (5.6)$$

The number of principal components obtained is equal to the number of original variables used. However, the number of principal components used into further analysis is smaller than the number of the original variables. The selection of the principal components used in further analysis depends on several criteria, which will be explained in the following section.

Concluding, PCA is used to reduce the amount of data for further visualization and analysis. It helps in the interpretation of the process and in the identification of groups of variables that are interrelated. This interrelation is not directly observed and is also useful to analyse which data is considered an outlier, i.e. samples with significant different characteristics when comparing with other samples (Alam *et al.*, 2004; Miller *et al.*, 2005; Rezzi *et al.*, 2005; Eriksson *et al.*, 2006; Winning *et al.*, 2008).

### **5.2.2. Principal Component Regression - PCR**

Principal component regression is a regression model that combines multiple linear regression (MLR) and principal component analysis (PCA) (Martens *et al.*, 1989; Alam *et al.*, 2004; Miller *et al.*, 2005; Pires *et al.*, 2008a), i.e. the spectral data is compressed by PCA and then a multiple linear regression is applied using as predictive variables the PCA scores. The variables  $X$ , from a linear combination, are substituted by a matrix of their PCs which are correlated with the output variable  $y$ . When comparing with PCA, PCR is another method that can be used to reduce the number of predictor variables by using some

of the principal components instead of using all original variables. There are several criteria that can be used to determine the number of principal components used in the regression analysis, such as: i) retain a sufficient number of principal components taking into account the cumulative percentage variance (Alam *et al.*, 2004), with a number of components that explained more than 90% of the original data set; and ii) retain the components whose eigenvalues are greater or equal to unit (Kaiser criterion) (Çamdevýren *et al.*, 2005) once, like this, the variance of each principal component is always greater or equal to the variance of the original variable. In this work, these methods typically agree on the number of components to retain, which is indifferent using either. The criteria mostly used is the one which selects PCs with eigenvalues greater than 1, and when this criterion does not achieve 90% of the original data variance the other criterion must be used. In this way, it is ensure that the select PCs represented at least 90% of the original data variance. The most important aspect to have in attention when choosing the components is the amount of variation they explain of  $X$  (Martens *et al.*, 1989; Miller *et al.*, 2005).

### **5.2.3. Partial Least Squares – PLS**

Partial least squares regression (PLSR) generates and combines characteristics from principal component regression (PCR) and multiple linear regression (MLR) to estimate the subspace that best explains the variability in the relation between the response variable ( $Y$ ) and predictor variables ( $X$ ). When using PCR, the principal components chosen explains the greatest amount of variance in  $X$ , independently of the strength of the relationships between the predictors and the response variables. However, PLS finds a set of uncorrelated linear combinations of the predictor variables which, as happens in PCR, have the highest capacity to explain the overall variability in the predictor variables but, moreover, still are highly correlated with the response variables (Miller *et al.*, 2005). Normally, PLS is preferred relatively to other traditional multivariate methods when there are smaller numbers of observations compared to the numbers of predictor variables, which is the reason that it is considered to be one of the least restrictive methods compared with other multivariate extensions of the MLR models (Wold *et al.*, 2001; Molina *et al.*, 2007). Besides this, it is also chosen in situations where only partial knowledge of the data

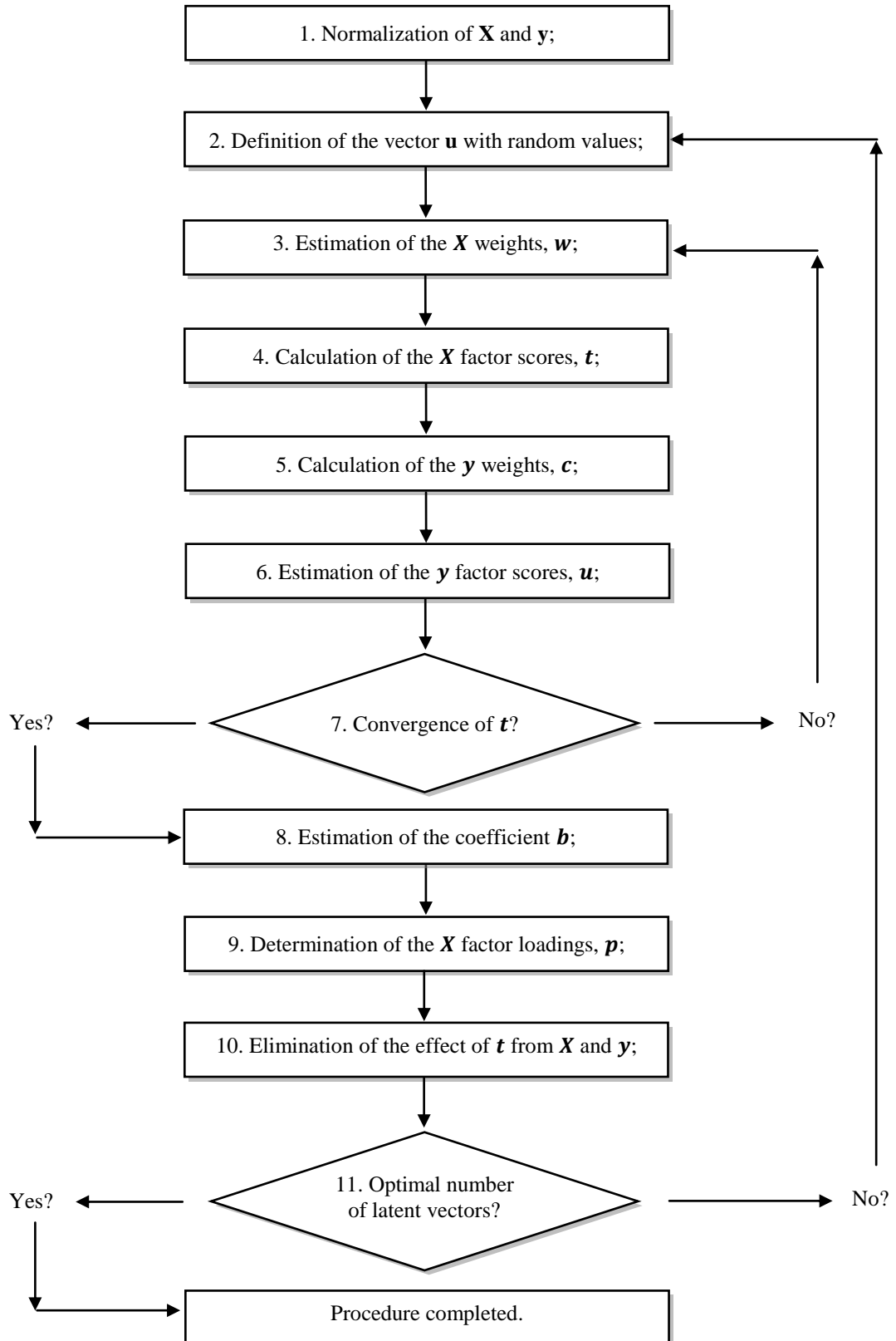
and multicollinearity among the explanatory variables exists (Eriksson *et al.*, 2006). As already stated, PLS has been designed to find the relation between the explanatory variables ( $\mathbf{X}$ ) and the output variables ( $\mathbf{y}$ ) and searches a set of orthogonal components, the latent variables, which performs a simultaneous decomposition of  $\mathbf{X}$  and  $\mathbf{y}$  with the constraint that these components explain as much as possible the covariance between  $\mathbf{X}$  and  $\mathbf{y}$  (Wise *et al.*, 1996; Abdi, 2003; McKenzie *et al.*, 2011). The number of latent variables depends on the correlation between these variables and the dependent variable, but the main idea is to have the highest correlation and, if possible, with the smallest number of latent variables. If the number of latent variables is too small then there is not enough information to create the model and it will be a poor prediction. However, if the number of latent variables is higher, then there is too much variation and the model will not be robust enough to small variations. There are different mathematical algorithms that can be applied when using PLS as the PLS with cross validation (Keun *et al.*, 2003; Molina *et al.*, 2007; Satya *et al.*, 2007; Flumignan *et al.*, 2012; Zerzucha *et al.*, 2012) and with external validation. With these algorithms the idea is to determine the correct number of latent variables which contributes to good predictions and small errors. Normally the cross validation is used when the number of samples is modest while the external validation is more often used when a higher number of samples is presented. The best model, in the two algorithms, is the one that presents a smaller error. In the cross validation the RMSECV (Root Mean Squared Error of Cross Validation) given by Equation 5.7 is calculated and in the external validation the RMSEP (Root Mean Squared Error of Prediction) is calculated using Equation 5.8.

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^n (y_{(i)} - \hat{y}_{(i)})^2}{n}} \quad (5.7)$$

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n_{val}} (\hat{y}_{(i)} - y_{(i)})^2}{n_{val}}} \quad (5.8)$$

For the external validation an independent validation data set is required and so  $n_{val}$  refers to the validation data set while, for the cross validation a data set encompassing the training and validation data sets, consequently,  $n$  refers to the number of samples of the training and validation data sets. This splitting of the data will be explained in Section 5.3.

The complete procedure used when applying the PLS model can be defined as follows (Wise *et al.*, 1996; Wold *et al.*, 2001; Abdi, 2003; Pires *et al.*, 2008b):



**Figure 5.3:** A flow chart of the complete procedure used when applying the PLS model.



1. Normalization of  $\mathbf{X}$  and  $\mathbf{y}$ ;

$$\mathbf{X}_0 = \frac{\mathbf{X}}{\|\mathbf{X}\|} \quad (5.9)$$

$$\mathbf{y}_0 = \frac{\mathbf{y}}{\|\mathbf{y}\|} \quad (5.10)$$

2. Estimation of a vector  $\mathbf{u}$  using random values (normally one of the  $\mathbf{y}$  columns);

3. Estimation of the  $\mathbf{X}$  weights,  $\mathbf{w}$ ;

$$\mathbf{w} = \frac{\mathbf{X}_k^T \mathbf{u}}{\|\mathbf{X}_k^T \mathbf{u}\|} \quad (5.11)$$

4. Calculation of the  $\mathbf{X}$  factor scores,  $\mathbf{t}$ ;

$$\mathbf{t} = \frac{\mathbf{X}_k \mathbf{w}}{\|\mathbf{X}_k \mathbf{w}\|} \quad (5.12)$$

5. Calculation of the  $\mathbf{y}$  weights,  $\mathbf{c}$ ;

$$\mathbf{c} = \frac{\mathbf{y}_k^T \mathbf{t}}{\|\mathbf{y}_k^T \mathbf{t}\|} \quad (5.13)$$

6. Estimation of the  $\mathbf{y}$  factor scores,  $\mathbf{u}$ ;

$$\mathbf{u} = \mathbf{y}_k \mathbf{c} \quad (5.14)$$

7. Repetition of the steps 3 to 6 until the convergence of  $\mathbf{t}$ ;

8. Estimation of the coefficient  $\mathbf{b}$  used in the prediction of  $\mathbf{y}$  from  $\mathbf{t}$ ;

$$\mathbf{b} = \mathbf{t}^T \mathbf{u} \quad (5.15)$$

9. Determination of the  $\mathbf{X}$  factor loadings,  $\mathbf{p}$ ;

$$\mathbf{p} = \mathbf{X}_k^T \mathbf{t} \quad (5.16)$$

10. Exclusion of the effect of  $\mathbf{t}$  from  $\mathbf{X}$  and  $\mathbf{y}$ ;

$$\mathbf{X}_{k+1} = \mathbf{X}_k - \mathbf{t}\mathbf{p}^T \quad (5.17)$$

$$\mathbf{y}_{k+1} = \mathbf{y} - \mathbf{b}\mathbf{t}\mathbf{c}^T \quad (5.18)$$

11. Repetition of the steps 2 to 10 until the determination of a select number of latent variables.

With this procedure, the vectors estimated,  $\mathbf{u}$ ,  $\mathbf{w}$ ,  $\mathbf{t}$ ,  $\mathbf{c}$  and  $\mathbf{p}$ , were stored in the columns of the correspondent matrices  $\mathbf{U}$ ,  $\mathbf{W}$ ,  $\mathbf{T}$ ,  $\mathbf{C}$  and  $\mathbf{P}$ , while the scalar  $b$  was stored in the diagonal matrix  $\mathbf{B}$ . The entire procedure was repeated until all latent variables were determined, i.e., until  $\mathbf{X}_k$  becomes the zero matrix.

The dependent variables were predicted using the following multivariate regression expression (Equation 5.19):

$$\hat{\mathbf{y}} = \mathbf{TBC}^T = \mathbf{XB}_{PLS} \quad (5.19)$$

This regression become equivalent to a principal component regression (PCR) if all the latent variables were used (Abdi, 2003).

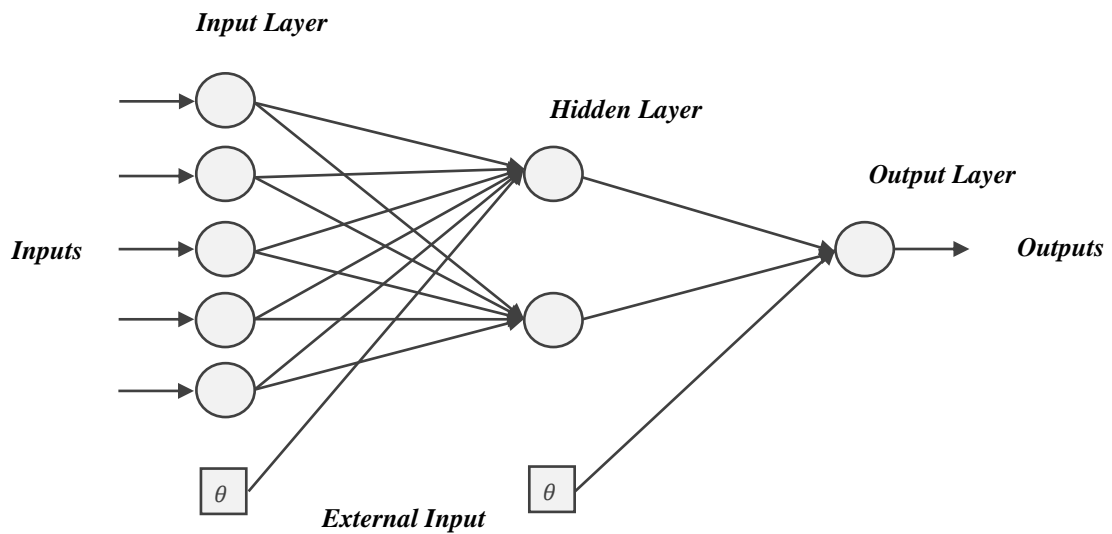
#### 5.2.4. Artificial Neural Networks – ANN

The artificial neural networks (ANN) are powerful mathematical models that can be used in the process control, in developing predictive models and in matching patterns (Fraser *et al.*, 1997). They are considered as a tool to solve chemical problems having been already applied to spectroscopic analysis in the interpretation of mass spectral data (Curry *et al.*, 1990), infrared spectral analysis (Roggo *et al.*, 2007) and nuclear  $^1\text{H}$  and  $^{13}\text{C}$  spectral analysis (McKenzie *et al.*, 2011). However, when comparing with the use of other techniques such as PCA and PLS, ANN are not widely used in the interpretation of NMR spectroscopy data. To our knowledge, it has not yet been found in scientific literature the use of ANN in the analysis of NMR spectral data of heavy petroleum fractions such as LVGO, HVGO and fuel oil. Related with petroleum refining industry, the ANN have been used in the prediction of distillate yields in a visbreaking process (Foddi *et al.*, 2009), in the prediction of diesel fuel cold properties (Marinović *et al.*, 2012) and diesel fuel lubricity (Korres *et al.*, 2002).

In a simple form, the ANN are learning models inspired in the biological neural processing system (Fraser *et al.*, 1997; McKenzie *et al.*, 2011) that attempt to mimic the operation of

neurons in the brain trying to imitate the human brain's ability to learn. Therefore, ANN can be considered as a processing unit that stores knowledge based on learning such as the human brain. The human brain is able to learn and make decisions based on human learning that results from human interaction with their surroundings.

The fundamental elements that characterize the ANN models are: i) the neurons or also known as nodes; ii) a pattern of connectivity among neurons; iii) an activation function for each neuron; and iv) a learning algorithm (Ahmadloo *et al.*, 2010; Wang *et al.*, 2010). All processing nodes that forms the ANN are interconnected by synapses in a net like structure, operate in parallel, communicating with each other by means of connecting weights (Wang *et al.*, 2010) and are distributed by several layers (input layer, output layer and hidden layers) (Miller *et al.*, 2005; Liu *et al.*, 2007; Roggo *et al.*, 2007), as shown in Figure 5.4. The structure of the network that is defined by the nodes and the way that they are linked can be divided into two of the most common types: i) feedforward networks; and ii) recurrent networks (Fraser *et al.*, 1997; Maier *et al.*, 2010; Wang *et al.*, 2010). The feedforward networks are characterized for only present connections starting in inputs nodes and ending in outputs nodes and for not having connections starting in outputs of nodes and ending in inputs of nodes in the same layer or previous layer. In these cases, the information flows only feedforward from input to output nodes and never feedback. On the other hand, when the networks present the two types of connections, the feedforward and the feedback, the neural network architecture is designated by recurrent network. From these two, the feedforward is the most applied in the real world applications and normally is characterized from having an output layer, not many hidden layers (one or two hidden layers will be preferred) and an input layer.



**Figure 5.4:** Example of a feedforward artificial neural network.

The learning process of the ANN is progressive and depends on the variability of the input data, and on the linkages that is established between nodes to obtain the output information. In the linkages established between nodes, also called synapses, there are values assigned that are defined as the weights. These weights are very important and considered adjustable parameters that are responsible by the neural structure and have the ability of learn and keep the information given on memory. It is important that the information received by the artificial neural network is descriptive about the physical phenomenon to allow that the ANN extracts the best relationship between the inputs and the desired outputs. There are three types of learning algorithms used in artificial neural networks: i) unsupervised learning; ii) reinforcement learning; and iii) supervised learning. In the first one, during the training of the artificial neural network, the response desired is not given. In the reinforcement learning and during the training of the neural network only a reward response is given. In the last one, the supervised learning, during the training the desired response is given to the ANN (Fraser *et al.*, 1997; Wang *et al.*, 2010). From all these learning algorithms, the most commonly used is supervised learning. As stated earlier, in the supervised learning a set of input-output pairs are given to the ANN. The data received by the input is then propagated to output passing through the interconnections between neurons. The neurons are responsible for processing the

information received by the inputs and calculating the output. There are some activated functions associated to the artificial neuron that are used to obtain the output. The most common functions are: i) the pure linear; ii) the sigmoid; and iii) the hyperbolic tangent (Marques, 1999; Hernández-Caraballo *et al.*, 2003; Liu *et al.*, 2007). The general working principle of the artificial neuron and the way to obtain the output of node  $i$  ( $Y_i$ ) can be demonstrated by means of the following expression:

$$Y_i = f \left( \sum_{j=1}^n w_{ij}x_j + \theta_i \right) \quad (5.20)$$

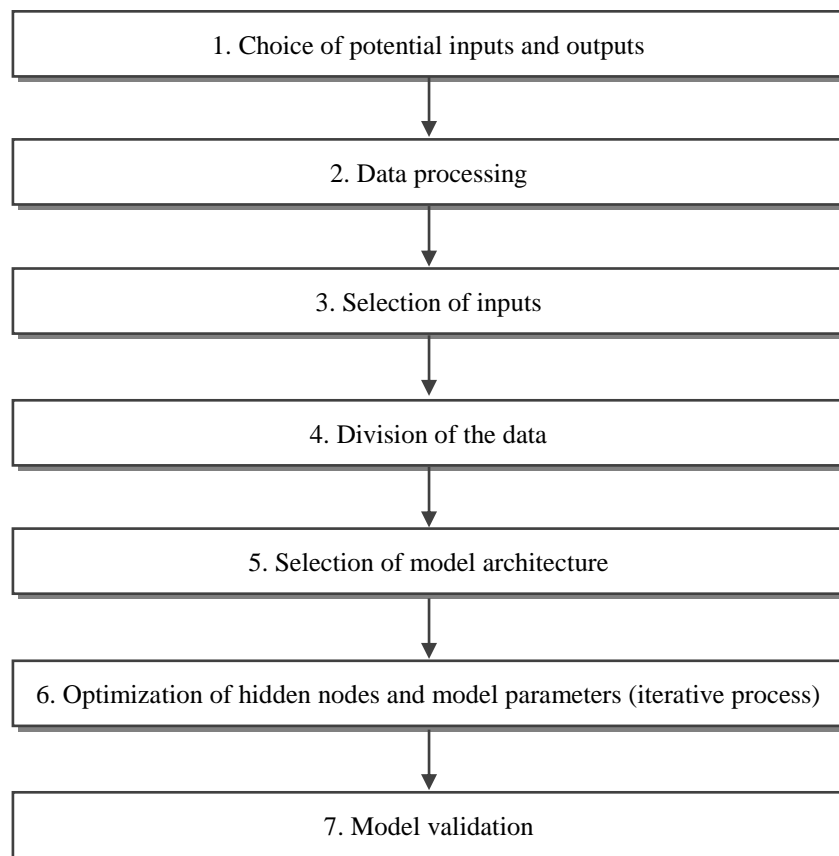
where  $f()$  is the activation function,  $x_j$  the input value from the previous node  $j$ ,  $w_{ij}$  the weight value that connects node  $i$  and node  $j$ ,  $\theta_i$  the bias value of node  $i$  and  $n$  the total number of previous nodes connecting with node  $i$ . The network continues to correlate the inputs and outputs to find the best weights and bias that minimize the output error, which is acquired when comparing the predicted output of the ANN with the actual output value. This process is called the training procedure and it is repetitively applied in an iterative process until the bias values and weights are adjusted to reduce the global error (Kapur *et al.*, 2004; Liu *et al.*, 2007). To train the ANN a learning rule must be implemented and the most popular in the ANN is the back-propagation algorithm (Liu *et al.*, 2007). The back-propagation algorithms has different variants being the Levenberg-Marquardt algorithm one of the most widely used (Chaloulakou *et al.*, 2003; Motlaghi *et al.*, 2008). As some authors (Chaloulakou *et al.*, 2003) announced, the Levenberg-Marquardt algorithm compared with the original back-propagation algorithm has showed significantly faster convergence and has been proved to be accurate enough in most cases. The main idea of using these algorithms is to minimize the error between the predicted and the observed output values and so the error of the output neurons is back propagated through the network to adjust the weights and bias values during the training period (Gershenson, 2003; Motlaghi *et al.*, 2008; Mashrei *et al.*, 2010).

During the training procedure it is possible to find some overfitting problems (Kapur *et al.*, 2004). Overfitting indicates that the artificial neural network models memorize the training examples and have not learnt the ability to generalize, consequently the application of the neural network models to new situations can lead to higher errors. The same doesn't occur with the values used to train the network, since the error is reduced due to the high number of iterations applied to obtain the best network. To overcome the overfitting problem and to improve the generalization of the network the available data is divided into three data sets, the training set, the validation set and the test set (Chaloulakou *et al.*, 2003; Maier *et al.*, 2010). The training data set is the only one used to develop the ANN models and to determine the network weights and bias values by solving a non-linear optimization problem. The validation data set is important to examine the accuracy of the ANN during the training set and plays a role of preventing overfitting on the training set. At the end of each training iteration the mean squared error (MSE) is determined (see Equation 5.21) by cross-validation for the training and validation data sets simultaneously. Normally, the errors on both data sets decreases as the proceeding of the training processes, although when the ANN begins to overfit, the error on the validation data set starts to increase while the error of the training data set progress decreasing (Mi *et al.*, 2005). The continuous increasing of the error on the validation data set, for a specified number of iterations, stops the training processes. The best network parameters corresponds to an ANN which presents a minimum global value of MSE for the validation and training data sets and generally a high number of attempts is performed. This ANN presenting the minimum global value of MSE for the validation and training data sets will have a better generalization. The MSE, used as the performance function to interpret the difference between the predicted output value of the ANN and the observed output value, is determined by means of the following equation:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (5.21)$$

where  $Y_i$  represents the observed values,  $\hat{Y}_i$  the modeled ones and  $n$  the number of samples from the data set. The test data set, never used for the training of the ANN, is used as a final test to analyse the trained ANN performance using another different set of samples (Kapur *et al.*, 2004; Ahmadloo *et al.*, 2010).

The main steps used in the development of ANN prediction models are resumed in Figure 5.5 (Maier *et al.*, 2010):



**Figure 5.5:** Steps used in the development of an ANN prediction models.

**1. Choice of potential inputs and outputs** – using a *priori* knowledge and/or availability of data, the modeller should be able to select a set of potential model input variables from the available data. It is important to select the right pairs of input-output data that is going to be used in the develop of the best possible model (Maier *et al.*, 2010);



**2. Data processing** – consists in the preparation of the data to subsequent steps of the model development process. Standardization and scaling are examples of data processing that can be applied. In this study, as already mentioned in Chapter 4, all input variables (NMR spectra) have undergone a pre-treatment;

**3. Selection of inputs** – the selection of the appropriate set of inputs is determinative in the ANN process. The idea is not having too many or too few inputs. If too many inputs are used, it is possible that some of them are related with each other and so redundant information or even no additional information is given to the model, which can contribute to overfitting and to identical model performance. While when using too few inputs it is possible that some of the inputs that have significant information are excluded and so the available data is not enough in the development of the resulting model. In this study, the principal components obtained after the application of PCA were used as inputs of the network. The use of principal components will contribute to simplify the network architecture due to the decrease of input variables, trying also to mitigate overfitting problems in the ANN model (Warne *et al.*, 2004; Sousa *et al.*, 2007; Maier *et al.*, 2010);

**4. Division of the data** – as stated earlier, the data is divided into training, validation and test data sets. These different data sets must be representative of all information available;

**5. Selection of model architecture** – as stated before, the network architecture gives important information about the overall network structure and indicates how the information flows in ANN models. It depends on the number of layers, the number of nodes and how they are linked with each other. The number of nodes in the input and output layers are determined by the nature of the problem while the optimum number of nodes in the hidden layer can be determined by trial and error. The most commonly ANN architectures used are the feedforward and recurrent networks (Maier *et al.*, 2010);

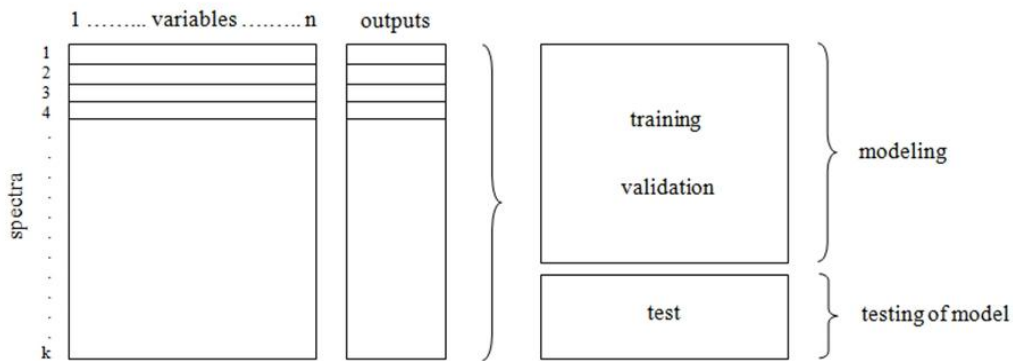
**6. Optimization of hidden nodes and model parameters** – for such optimization it is important to have in account: i) the structure of the network as the number of hidden nodes and the activation functions; ii) the model network parameters, as weights and bias values; and iii) the performance evaluation of the model. Following an iterative process of training and evaluation steps, the best model structure, which represents the desired input/output relationship, is obtained. With this iterative process the goal is to search a combination of model parameters (weights and bias values) that minimizes the output error between predicted and observed output values;

**7. Validation of the model** – after identifying the best network configuration, to validate the model it is necessary to verify its performance when using an independent data set, the test data set. In this study, the best network configuration corresponds to an ANN which presents a minimum global value of MSE for the validation and training data sets, after the network being trained during one thousands of attempts with different initial values for weights and bias. The best network obtained has a better generalization and its performance is evaluated calculating some statistical performance indexes (Maier *et al.*, 2010). The statistical parameters used in this work are presented in Section 5.4.

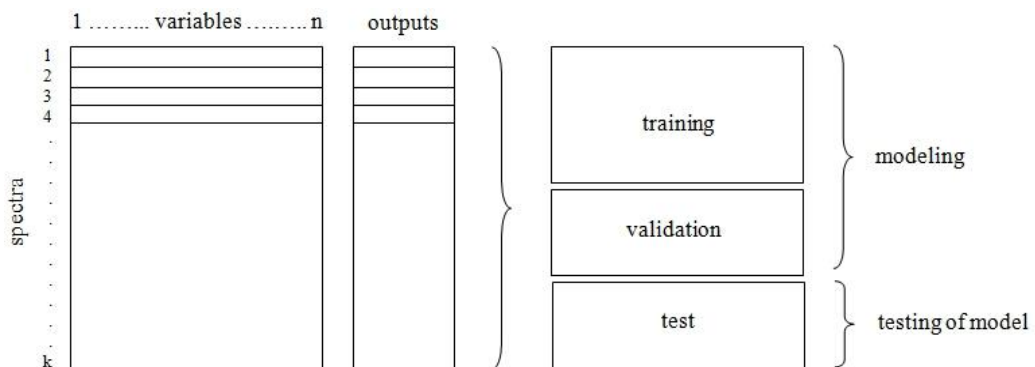
### **5.3. Models Pre-Treatment**

This study aims to evaluate the performance of the statistical models described above (PCR, PLS and ANN) for predicting the crude oil fractions properties from NMR. The data obtained with the NMR spectroscopy was used as the predictors' variables (**X**) of the models, while the physical-chemical properties determined using the ASTM and IP standards methods were used as outputs of the models (**y**). For the development of the models, the data was divided into different groups: training, validation and test data sets, in the ratio of 60%, 20% and 20%, respectively. All different data sets were representatives of the data available. For the PCR model and PLS model with cross validation the samples forming the training and validation data sets were grouped (Figure 5.6) while for the ANN model, as visualized in Figure 5.7, a separate validation data set was required. The training data set, the largest one, was used to set up the model and to determine the models

parameters, while the test data set was inevitable for evaluating the models performance when applied to a new data set. The validation data set was used to find the optimal dimensionality of the multivariate model, avoiding either overfitting or underfitting. The splitting of the data is a very important procedure for models that are aimed for applications in the routine processes control since it is necessary to perform a model that will be applicable to the overall data set and to new samples. If the splitting of the data was not made, e.g. if using all samples at the training data set, the developed model would become very dependent on the data and when applied to a new data set it probably would not respond with a good accuracy and precision leading to large prediction errors.



**Figure 5.6:** Schematic representation of variable selection for the PCR model and PLS model with cross validation.



**Figure 5.7:** Schematic representation of variable selection for the and ANN models.

The split of the original data has been performed using the percentile distribution. Using the percentile distribution the objective was to divide the data into three sets covering approximately the same region and having similar statistical properties. The percentile was determined applying the percentile function in the range of the corresponding data set for values of percentile between 0.01 to 0.99. Taking into consideration the percentile distribution determined, the original data was randomly selected and allocated to the different data sets. Hence, it was guaranteed the representative of the original data in three data sets. The results obtained using the percentile distribution to create the different data sets were similar with the results obtained by applying a different algorithm based on works performed by Snee and Fernández Pierna *et al.* (Snee, 1977; Fernández Pierna *et al.*, 2009) for selecting the different data sets. Their algorithm was based on the Euclidean distance between all possible pairs of inputs points. Hence, in this work the Euclidean distance between two input vector (input data obtained by NMR spectroscopy of the sample  $k$  and  $l$ , for example) for a given physical-chemical property ( $i$ ) was calculated by using the following expression:

$$d_{kl} = \sqrt{\sum_{j=1}^J (x_{kj} - x_{lj})^2} \quad (5.22)$$

with  $j$  representing all the samples analysed for a given physical-chemical property and  $x_{ij}$  indicating the  $j$ th sample analysed of the input data obtained by NMR spectroscopy for the physical-chemical property determined ( $i$ ). The procedure consisted in determining the distance between all input vectors and then selecting the two vectors which were furthest away from each other. These vectors were assigned for the training data set. Subsequently, after removing the two vectors for the training data set, from the remaining vectors, the other two vectors which were farthest away from each other were selected for the grouped validation and test data set. The procedure continued to be applied until all vectors from the original data were added alternately to each set, the training data set and the validation and test data sets. Afterward, the same procedure was performed to the grouped validation

and test data set in order to generate the two new data sets, the validation data set and test data set. In this stage of the analysis, the training data set was composed by 50% of the original data while the validation and test data sets were formed by 25% of the original data set, respectively. Since the results obtained with this algorithm were very similar to the ones achieved by the percentile distribution it was considered that both approaches could be applied. For this reason, the percentile distribution was the selected method to be used to create the different data sets.

When using the PCR method, the principal components obtained in the PCA, were used as inputs of the model and the physical-chemical properties analysed were used as outputs. To determine the number of principal components to be used in this regression analysis, the criteria indicated to retain the components whose eigenvalues were greater or equal to the unit, was used (Kaiser criterion), although, from all principal components only the ones with statistically significance were required. With this, it was guaranteed that the components which had important contributions from the original variables were included in the model. In this algorithm the training data set include 80% of the original data while the test data set comprised 20% of the original data. Meaning that in the PCR model the training data set (80%) encompassed the samples forming the training data set (60%) and the ones from the validation data set (20%), after the separation in the three different groups by the percentile distribution.

When using PLS, the cross-validation and the external validation method were test to determine the number of latent variables. However, once the results obtained using cross-validation were always better than the ones obtained with external validation, the last method was no longer used. The number of latent variables, estimated using the cross-validation algorithm, was selected by determining the minimum of the mean squared error (RMSECV), given by Equation 5.7, considering a wide range of latent variables. With this algorithm an iterative procedure was used where one vector was put aside at each step and the response values for the vector that had been left out was determined and the resulting error of prediction estimated. Thereby, the mean squared error was calculated from  $n$  multiple partial regressions, where in each regression the number of input vectors used to determined the regression parameters was  $n-1$ , while the other vector was used in the prediction of the error. This procedure was carried out until all the input vectors were used

only once in the estimation of the error. The value  $n$  was assigned to the number of input vectors of the training and validation data sets. The input variables of the PLS model were the NMR spectra and the properties of the samples the output variables.

For the ANN method, the principal components were used as input variables while the predicted properties were used as outputs. The networks used were of the feedforward artificial neural network type. These networks were characterized for having three types of layers (the input, the hidden and the output layer) and were tested using two, three and four nodes in the hidden layer. The best results were obtained when three nodes in the hidden layer were used. More than four nodes in the hidden layer were not considered as it can arrive to networks with a higher number of parameters compared with the dimension of data available. The transfer function used to obtain the output value was the hyperbolic tangent at the hidden layer node and the linear in the output layer node. The best network obtained corresponded to the ANN that presented a minimum global value of MSE (Equation 5.21) for the training and validation data sets, after the network had been trained during one thousands of attempts with different initial values for weights and bias. Hence, the training data set was used to determine the network weights and bias, while the validation data set was applied to evaluate the performance of the network during the training data set. The test data set was used to evaluate the performance of the obtained ANN in different situations than those used in the training of the ANN.

All models used were developed in Visual Basic<sup>®</sup> for Applications develop for Microsoft Excel<sup>®</sup> and in Matlab<sup>®</sup> 2011a (Mathworks Inc., Natick, MA, USA) using available models of principal component analysis, principal component regression, partial least squares and artificial neural networks.

#### **5.4. Performance Indexes**

To evaluate the model's performance in both training and validation data sets and test data set there were some statistical parameters (see Table 5.1) that were calculated, such as the: i) coefficient of determination ( $R^2$ ; Equation 5.23) or the Nash-Sutcliffe efficiency (NSE; Equation 5.24); ii) mean absolute error (MAE; Equation 5.25); iii) the root mean squared error (RMSE; Equation 5.26); iv) mean bias error (MBE; Equation 5.27); v) percent bias

(PBIAS; Equation 5.28); and vi) RMSE-observations standard deviation ratio (RSR; Equation 5.29). These statistical parameters analysed can be divided in different major categories: i) standard regression; ii) dimensionless; and iii) error index. The standard regression statistics, which included the coefficient of determination, determine the strength of the linear relationship between the predicted and observed data (Moriiasi *et al.*, 2007). The Nash-Sutcliffe efficiency (NSE) is an example of dimensionless techniques, which provide a relative model evaluation between the residual variance (“noise”) compared to the measured data variance (“information”). Moreover, it indicates how well the plot of observed versus measured data fits the 1:1 line (Legates *et al.*, 1999; Hall, 2001; Moriiasi *et al.*, 2007). The NSE statistical parameter, introduced by Nash and Sutcliffe (1970), is analogous with the coefficient of determination ( $R^2$ ) and, in this case, where both determined by using the same expression, as visualized in Table 5.1. Although these statistical parameters were determined in the same way, it was considered important to determine both for model evaluation. The tendency is to use the coefficient of determination as a first way to evaluate how much of the observed dispersion is explained by the prediction. However, it should not be used alone when model evaluation is required and therefore, other statistical parameters are necessary to evaluate the relationship between measured and predicted properties. Since the general performance ratings presented by Moriiasi *et al.* (2007) were used in this work, the NSE statistical parameter was also determined. Besides the standard regression and dimensionless categories, the error indices measure the deviation between the predicted and experimental data. Examples of error indices are the mean absolute error (MAE), mean bias error (MBE), root mean squared error (RMSE), percent bias (PBIAS), and the RMSE-observations standard deviation ratio (RSR).

To evaluate each model performance, all the statistical parameters indicated in Table 5.1 were determined. However, since the general performance ratings presented by Moriiasi *et al.* (2007) were based on the RSR, NSE and PBIAS results, only these statistical parameters, in addition to the reported performance ratings presented in Table 5.2, were used to classify the developed models. In this way, taking in consideration the RSR, NSE and PBIAS results, the developed models can be evaluated, compared and classified as very good, good, satisfactory and unsatisfactory. These general performance ratings recommended by Moriiasi *et al.* (2007) (Table 5.2) were defined to compare simulated

output with measure data and consequently used to establish a platform for model evaluation. The idea of Moriasi *et al.* (2007) research was to demonstrate that the combination of graphical techniques and dimensionless and error index statistics should be used for model evaluation. Consequently, the quantitative statistics NSE, PBIAS and RSR were recommended as well as its performance ratings.



**Table 5.1:** The statistical parameters analysed.

Statistical Parameters	Expression	Definition
Coefficient of Determination (R <sup>2</sup> )	$R^2 = \frac{\sum_{i=1}^n (Y_i - \bar{Y}_i)^2 - \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y}_i)^2}$	<p>With the coefficient of determination it is possible to have an idea about the variability of the results reproduced by the development model when compared with the observed data. High values of R<sup>2</sup> indicate small errors in variances and consequently correspond to better models. (5.23)</p>
Nash-Sutcliffe Efficiency (NSE)	$NSE = 1 - \left[ \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y}_i)^2} \right]$	<p>The Nash-Sutcliffe efficiency, analogous to the R<sup>2</sup>, is a normalized statistic parameter presenting the relative magnitude of the residual variance when compared to the variance of the measured data. Positive values of NSE indicate acceptable performances while the negative ones are representative for unacceptable models. If NSE is equal to 1 a good model is obtained. (5.24)</p>
Mean Absolute Error (MAE)	$MAE = \frac{1}{n} \sum_{i=1}^n  \hat{Y}_i - Y_i $	<p>MAE (Equation 5.25) and RMSE (Equation 5.26) evaluate the residual errors, describing the difference between the predicted and observed values. A better model is obtained when MAE and RMSE presents lower values. These performance indexes have the same units of the output variable. (5.25)</p>
Root Mean Squared Error (RMSE)	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}$	<p>MAE (Equation 5.25) and RMSE (Equation 5.26) evaluate the residual errors, describing the difference between the predicted and observed values. A better model is obtained when MAE and RMSE presents lower values. These performance indexes have the same units of the output variable. (5.26)</p>

Note:  $Y_i$  = output value;  $\hat{Y}_i$  = model output;  $\bar{Y}_i$  = average of the output variable.

**Table 5.1:** The statistical parameters analysed (continuation).

Statistical Parameters	Expression	Definition
Mean Bias Error (MBE)	$MBE = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)$	<p>MBE gives an idea if the observed value is over or under estimated. MBE results closest to zero are more desirable. This indice have the same units of the output variable. (5.27)</p>
Percent Bias (PBIAS)	$PBIAS = \left[ \frac{\sum_{i=1}^n (\hat{Y}_i - Y_i) \times (100)}{\sum_{i=1}^n (Y_i)} \right]$	<p>PBIAS (expressed as a percentage), similar to MBE, indicates that the simulated data is larger or smaller than the observed data. If PBIAS has a positive value it indicates that the model is overestimated, otherwise the model would be underestimated. Values equal to 0 corresponds to a good model. (5.28)</p>
RMSE-Observations Standard Deviation Ratio (RSR)	$RSR = \frac{RMSE}{STDEV_{obs}} = \frac{\left[ \sqrt{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \right]}{\left[ \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2} \right]}$	<p>RSR is not more than the standardization of RMSE by calculating the ratio between the RMSE and the standard deviation of the measured data. RSR ranges between 0 to a larger positive value. If RSR is 0, RMSE and the residual variation is also zero, then a good model simulation performance is obtained. RSR is a valuable statistical parameter once it integrates the benefits of error index statistics and a normalization factor allowing its application to various constituents (Moriasi <i>et al.</i>, 2007). (5.29)</p>

Note:  $Y_i$  = output value;  $\hat{Y}_i$  = model output;  $\bar{Y}$  = average of the output variable.

**Table 5.2:** General performance ratings (Moriassi *et al.*, 2007).

<b>Performance Rating</b>	<b>RSR</b>	<b>NSE</b>	<b>PBIAS</b>
<b>Very Good</b>	$0.00 \leq \text{RSR} \leq 0.50$	$0.75 < \text{NSE} \leq 1.00$	$\text{PBIAS} < \pm 15$
<b>Good</b>	$0.50 < \text{RSR} \leq 0.60$	$0.65 < \text{NSE} \leq 0.75$	$\pm 15 \leq \text{PBIAS} < \pm 30$
<b>Satisfactory</b>	$0.60 < \text{RSR} \leq 0.70$	$0.50 < \text{NSE} \leq 0.65$	$\pm 30 \leq \text{PBIAS} < \pm 55$
<b>Unsatisfactory</b>	$\text{RSR} > 0.70$	$\text{NSE} \leq 0.50$	$\text{PBIAS} \geq \pm 55$

Before applying this classification scheme (Table 5.2), the performance ratings presented were evaluated to verify if they were adequate to the samples in study. It was verified that the performance ratings recommended by Moriassi *et al.* (2007) were applicable to the models developed during this research once no conflicting performance ratings were obtained into the models. Normally the results of RSR, NSE and PBIAS were in accordance. However, in situations with conflicting performance ratings, e.g. if the RSR, NSE and PBIAS results were not all three in accordance and just two of them were in agreement, then the overall performance should be described with the performance rating defined by the two parameters in accordance. RSR and NSE are standardized parameters, therefore expected, as visualized, their ability to many models evaluation. The performance ratings used for PBIAS demonstrated to be, in most cases, in accordance with the other two statistical parameters and so, in this way, considered a good statistical parameter for these models evaluation. Moreover, it was considered important and appropriate using the general performance ratings developed by Moriassi *et al.* (2007) to compare simulated output with the measured.

## 5.5. Final Remarks

To predict the physical-chemical properties of the diverse petroleum fractions analysed (fuel oil, HVGO and LVGO) using as predictors NMR spectroscopy, different statistical models (principal component regression, partial least squares and artificial neural networks) were applied. The performance of each model developed was evaluated by using statistical significant regression parameters and a classification based on very good, good, satisfactory and unsatisfactory model.



# 6

## **RESULTS AND DISCUSSION**

This chapter presents the results obtained by nuclear magnetic resonance spectroscopy in the identification of some physical-chemical properties using different multivariate models. The selected models were based on: i) principal component regression; ii) partial least squares; and iii) artificial neural networks. As it will be shown, the models were ranked considering the values of several statistical indexes aiming to quantify the performance of the models.

## 6. RESULTS AND DISCUSSION

6.1. Data.....	177
6.1.1. <sup>13</sup> C NMR Spectroscopy.....	177
6.1.2. <sup>1</sup> H NMR Spectroscopy.....	181
6.1.3. <sup>1</sup> H and <sup>13</sup> C NMR Spectroscopy.....	185
6.2. <sup>1</sup> H NMR Spectroscopy in the Analysis of Petroleum Fractions.....	186
6.2.1. Fuel Oil .....	187
6.2.1.1. Fuel Oil - Final Remarks .....	221
6.2.2. LVGO .....	224
6.2.1.2. LVGO – Final Remarks.....	243
6.2.3. HVGO .....	246
6.2.1.3. HVGO – Final Remarks .....	277
6.3. LVGO, HVGO and Fuel Oil – Chemical Composition .....	281

## 6.1. Data

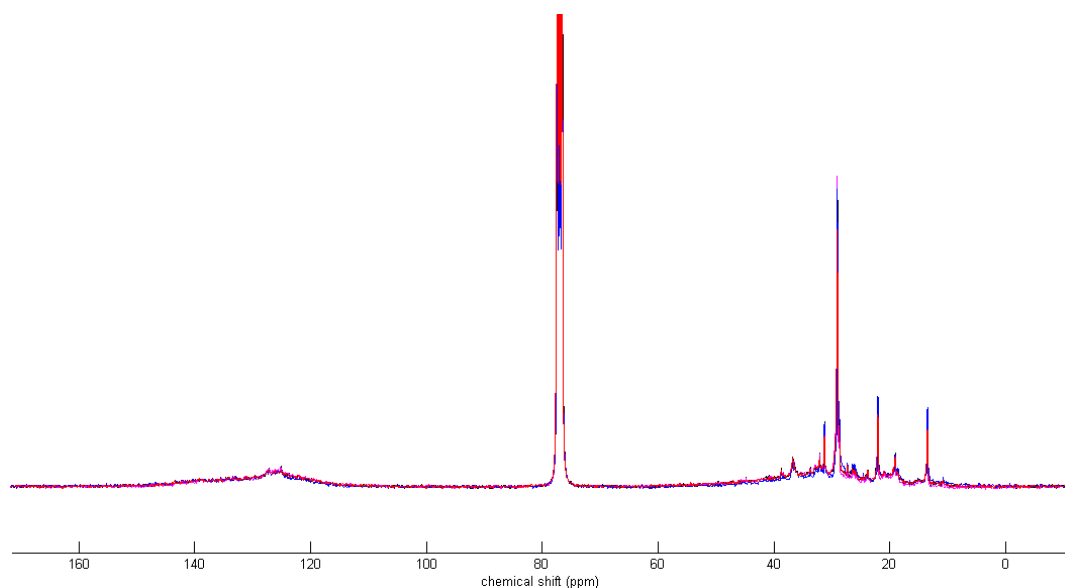
As already stated in a previous chapter, the monitoring samples considered in this study were fuel oil, HVGO and LVGO. Their physical-chemical properties were determined at the laboratory of Matosinhos refinery while the NMR spectra of such samples were determined with a laboratory spectrometer at the University of Aveiro. At the beginning of the project,  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectroscopy were used to evaluate the performance of each technique in predicting the physical-chemical properties of the samples in study. Information about the number of the first samples analysed by each technique and their properties is shown in Table 6.1. However, the analysis of each sample using both,  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectroscopy, has not always been applied. As set out in further detail below, there are some limitations associated with one of these techniques. A brief description of the results obtained using  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectroscopy techniques is presented. Finally, a detailed analysis of the results for predicting physical-chemical properties of the three different streams is presented.

**Table 6.1:** Number of samples analysed by  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectroscopy and minimum and maximum values of the properties analysed by the standard methods.

samples	properties	number of samples analysed	properties results	
			minimum	maximum
<b>Fuel Oil</b>	Kinematic viscosity 100 °C (cSt)	32	20.76	68.34
	Density (g/mL)	32	0.9665	0.9911

### 6.1.1. $^{13}\text{C}$ NMR Spectroscopy

Figure 6.1 presents some of the quantitative  $^{13}\text{C}$  NMR spectra obtained with the analysis of the fuel oil samples. By observation of the spectra it is possible to visualize that between 70 and 90 ppm no important information about the composition of the sample occurs, i.e., no important signals appears in this region. Consequently, this region was omitted to avoid including the solvent signal into the models. Hence, the spectral region considered in the developed models only included the signals between 11 and 60 ppm and between 100 and 160 ppm.



**Figure 6.1:** Quantitative <sup>13</sup>C NMR spectra of some fuel oil samples.

For the development of the models all 32 samples of fuel oil were used. From which, 12 samples constituted the training data set while the validation and test data sets were represented by 10 samples, respectively. The bin width (division of the spectrum in specified segments) used was 0.5 ppm instead of 0.02 ppm. In the case of quantitative <sup>13</sup>C NMR, it was impossible to use a bin of 0.02 ppm, due to the large chemical shift of the quantitative <sup>13</sup>C NMR spectrum. If using a bin of 0.02 ppm, 21 PCs (principal components) were select. The results obtained when using quantitative <sup>13</sup>C NMR and the multivariate models (PCR, PLS and ANN) are presented in Table 6.2.

Before applied PCR and ANN models, PCA was implemented. PCA was used as a classification method to reduce the number of variables in the models by selecting only the principal components with eigenvalues greater than or equal to 1. In this study, the PCs with eigenvalues greater than or equal to 1 incorporate more than 90% of the original data variance. These results were very important once it indicated that the PCs chosen contained significant information about the original variables and increased the confidence in the PCA results. Results presented in Table 6.2 shows that 17 PCs were used for predicting kinematic viscosity (cSt). The percentage of the original data variance contained in the selected PCs was 96.6%. For predicting density (g/mL), 17 PCs corresponding to 96.4% of the original data variance were required.



**Table 6.2:** Results obtained for different properties of fuel oil analysed when using quantitative <sup>13</sup>C NMR in multivariate models.

			bin	number PC	number LV	number nodes	total variance	R <sup>2</sup> / NSE	MAE	RMSE	MBE	PBIAS	RSR	model performance
<b>Kinematic viscosity 100 °C</b>	PCR	training and validation						0.26	7.06	8.96	0.00	0.00	0.86	unsatisfactory
		test						0.36	6.14	8.32	-1.60	-4.47	0.80	unsatisfactory
	ANN	training and validation	0.5	17			96.6%	<b>0.63</b>	<b>4.72</b>	<b>6.29</b>	<b>0.52</b>	<b>1.45</b>	<b>0.61</b>	<b>satisfactory</b>
		test				3		<b>0.52</b>	<b>5.51</b>	<b>7.18</b>	<b>-2.22</b>	<b>-6.20</b>	<b>0.69</b>	<b>satisfactory</b>
	PLS	training and validation				5		0.99	0.65	0.79	0.00	0.00	0.08	very good
		test						0.22	6.47	9.13	-1.50	-4.19	0.88	unsatisfactory
<b>Density</b>	PCR	training and validation						0.31	0.004	0.005	0.00	0.00	0.83	unsatisfactory
		test						-0.24	0.005	0.006	0.00	0.01	1.11	unsatisfactory
	ANN	training and validation	0.5	17			96.4%	<b>0.80</b>	<b>0.002</b>	<b>0.002</b>	<b>-0.001</b>	<b>-0.06</b>	<b>0.45</b>	<b>very good</b>
		test				3		<b>0.67</b>	<b>0.002</b>	<b>0.003</b>	<b>0.00</b>	<b>0.01</b>	<b>0.58</b>	<b>good</b>
	PLS	training and validation				3		0.84	0.002	0.002	0.00	0.00	0.40	very good
		test						0.50	0.003	0.004	0.00	0.04	0.71	unsatisfactory

The final models were obtained and the performance of the developed models (PCR, PLS and ANN) considering the training and validation data sets and the test data set evaluated through the determination of the performance indexes as referred in Chapter 5. When predicting fuel oil kinematic viscosity, very good performance indexes in the training and validation data sets using the PLS model with five latent variables were obtained, as visualized in Table 6.2. The model did not predict the values of the test data set. In a way to understand this behaviour, a very good prediction obtained at the training and validation data sets and an unsatisfactory prediction at the test data set, the data sets were modified to verify if no mistakes during the selection of the samples for each data set occurred. As expected, since each data set was representative of the original data and equally spread in variance, the same behaviour occurred, e.g. the model did not predict the values of the test data set independently of the selected samples composing the test data set. This lead as to conclude that the obtained performance indexes were not dependent on the choice of the samples forming each data set, but on the model applied and the number of samples composing the data sets, as better predictions, than the ones achieved with the PLS and PCR models, were obtained, at the test data set, when applying the ANN model. Comparing with PCR and PLS, it was possible to visualize that ANN model presented higher NSE value and smaller RSR result, when analysing the test data set. Furthermore, it was possible to verify that it was with the ANN model that a smaller RMSE value was obtained at the test data set. The MBE and PBIAS were negative for the test data set predicted with the ANN model, meaning that, in average, the predict kinematic viscosity of fuel oil samples was underestimated. For the training and validation data sets, the results obtained were not as good as the ones obtained with the PLS model. Nevertheless, it was with the ANN model that similar predictive results were obtained for both data sets. Using the general performance ratings developed by Moriasi *et al.* (2007), the results obtained by the ANN model were classified as satisfactory in all data sets. Accordingly, the results showed that the best model to predict the kinematic viscosity of fuel oil samples using quantitative  $^{13}\text{C}$  NMR spectroscopy was obtained with ANN model.

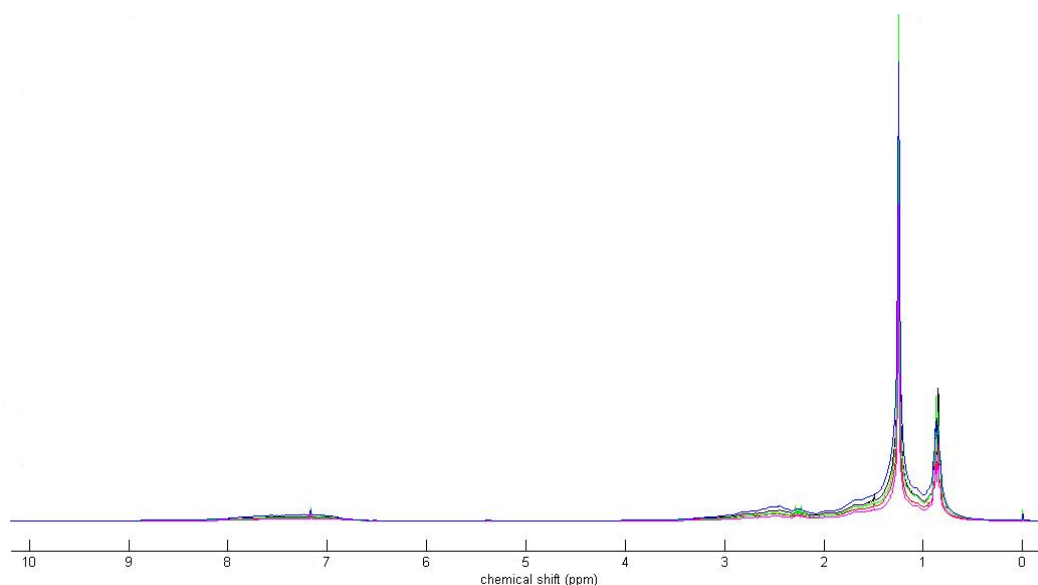
Analysing the results obtained for the prediction of fuel oil density (Table 6.2) it was possible to conclude that the worst performance occurred with PCR model, where unsatisfactory predictions for both data sets were obtained. Concerning the results obtained with the PLS model, it was verified that very good predictions at the training and

validation data sets were obtained, however, when predicting the test data set unsatisfactory results were achieved. When applying the ANN model, very good predictions at the training and validation data sets were obtained, although for the test data set only a good model was achieved. Once again, the ANN was the model that gave the best results in both data sets. High NSE value and small RSR results were indicative of a good model. Additionally, RMSE and MAE presented low values indicating that the difference between the predicted and observed density values was very small. For the training and validation data sets and for the test data set, the prediction of the density values was overestimated.

Coming back to Table 6.2, it was clear that the use of the ANN model lead to more accurate results, compared to the other models in predicting the kinematic viscosity and density of fuel oil samples by using quantitative  $^{13}\text{C}$  NMR spectroscopy. However, to improve the results obtained and the accuracy of the developed models more fuel oil samples should be included in the both data sets.

### **6.1.2. $^1\text{H}$ NMR Spectroscopy**

With respect to experimental results, the  $^1\text{H}$  NMR spectra of several fuel oil samples are presented in Figure 6.2. Even though the  $^1\text{H}$  NMR spectrum of fuel oil is very crowded and highly overlapped it is possible to assign distinct structural regions to different functional groups. The region between 0.5 to 4.5 ppm originates the aliphatic region, where it is possible to identify the aliphatic  $\text{CH}_3$  groups at 0.5 to 1.0 ppm, while the aromatic region spans from 6.5 to 10.0 ppm. From these regions some chemical shifts were eliminated to create the appropriate data for posterior application of the models. The data set comprises the chemical shift region between 0.5 to 4.5 ppm, 6.5 to 7.1 ppm and 7.4 to 10.0 ppm. While the solvent signal (7.1 to 7.4 ppm) and the regions where no important information occurred and were spectral signatures of additives or other chemical species appears (4.5 to 6.4 ppm) were avoided.



**Figure 6.2:**  $^1\text{H}$  NMR spectra of some fuel oil samples.

The 32 fuel oil samples analysed with quantitative  $^{13}\text{C}$  NMR spectroscopy were also the samples analysed by  $^1\text{H}$  NMR spectroscopy, to compare the results of both techniques. Prior to multivariate modelling, the 32 fuel oil samples constituting the data set were divided in three different sets, the training data set containing 12 fuel oil samples, the validation data set comprising 10 samples and the test data set containing 10 samples. The spectral area used was divided in constant segments of 0.02 ppm, since the bin width of 0.02 ppm procedure was preferred over other segments. Moreover, the whole procedure used to apply the multivariate models was the same as the one performed when using the quantitative  $^{13}\text{C}$  NMR spectra. The results obtained after applying multivariate modelling to the  $^1\text{H}$  NMR spectra are provided in Table 6.3.

**Table 6.3:** Results obtained for different properties of fuel oil analysed when using <sup>1</sup>H NMR in multivariate models.

			bin	number PC	number LV	number nodes	total variance	R <sup>2</sup> / NSE	MAE	RMSE	MBE	PBIAS	RSR	model performance
<b>Kinematic viscosity 100 °C</b>	PCR	training and validation						0.90	2.62	3.27	0.00	0.00	0.31	very good
		test						0.42	6.37	7.92	1.08	3.03	0.76	unsatisfactory
	ANN	training and validation	0.02	7		3	96.7%	<b>0.88</b>	<b>2.80</b>	<b>3.59</b>	<b>0.15</b>	<b>0.43</b>	<b>0.34</b>	<b>very good</b>
		test						<b>0.66</b>	<b>4.11</b>	<b>6.14</b>	<b>0.12</b>	<b>0.32</b>	<b>0.59</b>	<b>good</b>
	PLS	training and validation			12			0.99	0.21	0.26	0.00	0.00	0.02	very good
		test						0.57	5.39	6.82	1.70	4.75	0.66	satisfactory
<b>Density</b>	PCR	training and validation						0.60	0.003	0.004	0.00	0.00	0.64	satisfactory
		test						0.18	0.004	0.005	0.00	0.21	0.91	unsatisfactory
	ANN	training and validation	0.02	7		3	96.6%	0.89	0.001	0.002	0.00	0.03	0.32	very good
		test						0.82	0.002	0.002	0.00	0.12	0.42	very good
	PLS	training and validation			8			<b>0.98</b>	<b>0.001</b>	<b>0.001</b>	<b>0.00</b>	<b>0.00</b>	<b>0.12</b>	<b>very good</b>
		test						<b>0.81</b>	<b>0.00</b>	<b>0.002</b>	<b>0.00</b>	<b>0.04</b>	<b>0.43</b>	<b>very good</b>

The firsts results obtained during the prediction of the different properties of fuel oil by using  $^1\text{H}$  NMR spectroscopy in combination with multivariate modelling, demonstrated that  $^1\text{H}$  NMR spectroscopy has high potential to predict such properties.

The prediction of the kinematic viscosity was performed using 7 PCs corresponding to 96.7% of the total variance. The selection of the number of PCs was made taking into account the Kaiser criterion as referred in Section 5.2.2 and, as previously mentioned, also considering the total variance obtained when using the quantitative  $^{13}\text{C}$  NMR spectroscopy. Coming back to Table 6.3, independently of the model applied, very good predictions of fuel oil kinematic viscosity were obtained for the training and validation data sets. High NSE results, small RMSE and MAE values, and small RSR results characterized all developed models achieved for the training and validation data sets. However, for the test data set, the predictions obtained with the applied models were not as good as the ones achieved for the training and validation data sets. As visualized in Table 6.3, the PCR model presented unsatisfactory performance indexes with low quality results, as small NSE result and very high RMSE value. The PLS model also indicated that the results obtained when predicting the kinematic viscosity of fuel oil were not accurate. In this case, only satisfactory predictions were performed by the model. Even changing the number of latent variables, it was not possible to improve the PLS results. When the ANN model was applied, better performance indexes were obtained, showing that the develop model was of higher quality. This model, obtained when using 3 nodes in the hidden layer, was characterized for having higher NSE result, small RMSE and MAE values, and small RSR value, comparing with the ones obtained with the PCR and PLS models.

For the density predictions of the fuel oil samples, the results obtained during training and validation data sets and test data set using PLS and ANN models demonstrated that accurate models were achieved. It was observed in Table 6.3 that very good performance indexes were obtained for both data sets when both models, PLS and ANN models, were applied. Only the PCR model presented a significant difference, since worse predictions were obtained for both data sets, training and validation data sets and test data set. The predictions obtained with the PCR model at the training and validation data sets presented reasonable results, with a satisfactory model achieved; however, the performance indexes for the test data set were slightly worse than for the training and validation data sets. For

the PCR and ANN model, 7 PCs were used, explaining 96.6% of the total variance and 3 nodes at the hidden layer of the ANN model where needed. The prediction with PLS model was performed using 8 latent variables.

Concluding, the ANN and PLS models were preferred over PCR model to predict the properties of the fuel oil samples by using  $^1\text{H}$  NMR spectroscopy. The performance indexes indicated that the developed models were of higher quality when using the ANN model to predict the fuel oil kinematic viscosity. Relatively to the prediction of the fuel oil density, similar performance indexes than the ones achieved with the PLS model were obtained when using the ANN model. In general, higher coefficient of determination between the predicted and measured data and smaller residual variations were achieved when using the ANN model to predict both properties analysed. The ANN model demonstrated to be a very important method having the capability to construct very good models between the measured data and the predicted properties, due to their ability to learn complex non-linear and multivariate relationships between process parameters.

### 6.1.3. $^1\text{H}$ and $^{13}\text{C}$ NMR Spectroscopy

Comparing the results presented above, it was possible to conclude that both techniques,  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectroscopy, combined with multivariate statistical methods could give good predictions of the properties of fuel oil samples. In addition, it was demonstrated that both techniques could be used to predict the kinematic viscosity at 100 °C and the density of fuel oil samples instead of the traditional ASTM methods in laboratory applications. However, when the idea was to evaluate the possibility of using an industrial application of NMR spectroscopy, to predict the properties of different samples, quantitative  $^{13}\text{C}$  NMR spectroscopy was not a good choice. Besides the good results, especially when predicting fuel oil density, the main problem associated with this technique was related to the time required to obtain a quantitative  $^{13}\text{C}$  NMR spectrum of such complex samples as petroleum fractions. It was necessary eight hours to obtain a quantitative  $^{13}\text{C}$  NMR spectrum of a fuel oil sample. With less than eight hours it was not possible to obtain a good spectrum, with a good signal resolution and an adequate signal to noise ratio. The time required to acquire this spectrum was very high and even a standard ASTM method could be faster (see Table

4.2, page 126). Besides the time, the preparation of a quantitative  $^{13}\text{C}$  NMR sample could also be a restriction. As referred previously, to obtain a quantitative  $^{13}\text{C}$  NMR spectrum a paramagnetic relaxation reagent as  $\text{Cr}(\text{acac})_3$  was required.  $\text{Cr}(\text{acac})_3$  was difficult to dissolve in fuel oil samples requiring higher amounts of the solvent ( $\text{CDCl}_3$ ). Additionally, to guarantee that a perfect homogeneous solution and all  $\text{Cr}(\text{acac})_3$  was dissolved it was necessary to use the ultrasonic bath. Due to these facts and the difficulty associated with the use of the paramagnetic relaxation reagent it is not guaranteed that a perfect solution is obtainable at an online sample preparation.

The main objective when evaluating the use of an online NMR spectrometer is to verify if its implementation will improve and lead to a better control over the streams and lead to achieve optimal operations and profitability in real time. Additionally, the objective is to obtain information about different properties, from a single measurement, that contribute to immediately adjust the production variables and increase the profitability by producing well at the first time, avoiding product loss.  $^1\text{H}$  NMR spectroscopy have the benefit of giving results in just a few minutes, doesn't need complicate samples preparation, gives good predictions and contributes to develop reliable models. For this reason,  $^1\text{H}$  NMR spectroscopy was the technique used in this work for the prediction of petroleum fractions properties.

In laboratory applications the use of both  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectroscopy will be recommended. However, for industrial applications the  $^1\text{H}$  NMR spectroscopy will be the chosen technique.

## **6.2. $^1\text{H}$ NMR Spectroscopy in the Analysis of Petroleum Fractions**

As indicated above,  $^1\text{H}$  NMR has been considered a fast and viable technique to predict petroleum fractions properties. As a result, in this work,  $^1\text{H}$  NMR spectroscopy, combined with multivariate methods, has been used in the prediction of the physical-chemical properties of fuel oil, HVGO and LVGO samples. The physical-chemical properties that were measured for each stream, the number of samples analysed by  $^1\text{H}$  NMR spectroscopy and some information about the properties results are reported in Table 6.4. The results of the development models for all properties analysed in each stream are presented as follows.



Since the procedure used to obtain the predicted results was similar in all properties of each analysed stream, it is only presented in detailed for the first property, i.e., for sulfur content.

**Table 6.4:** Number of samples analysed by  $^1\text{H}$  NMR spectroscopy and minimum and maximum values of the properties analysed by the standard methods.

samples	properties	number of samples analysed	properties results		
			minimum	maximum	
<b>Fuel Oil</b>	Sulfur content (% m/m)	213	0.53	3.54	
	Kinematic viscosity 100 °C (cSt)	217	14.20	1367	
	Kinematic viscosity 50 °C (cSt)	108	103.2	890.4	
	Density (g/mL)	185	0.9487	1.019	
	Carbon residue (% m/m)	103	12.9	22.1	
	Flash point (°C)	157	49	102	
	P-value	217	1.00	1.8	
<b>LVGO</b>	Density (g/mL)	105	0.8592	0.9050	
	Distillation (°C)	5%	105	227	307
		10%	105	231	322
		50%	105	251	361
		90%	105	279	414
		95%	105	292	417
<b>HVGO</b>	Density (g/mL)	189	0.9152	0.9609	
	Kinematic viscosity 100 °C (cSt)	168	8.050	18.60	
	Carbon residue (% m/m)	142	0.08	1.03	
	Distillation (°C)	5%	116	359	410
		10%	116	374	418
		50%	116	433	466
90%		114	484	542	
	95%	99	500	558	

### 6.2.1. Fuel Oil

The purpose of analysing fuel oil with  $^1\text{H}$  NMR spectroscopy was to verify if it was possible to use this technique to predict, by only one measurement, several properties which are part of fuel oil specifications. These physical-chemical properties were: i) sulfur content; ii) density; iii) carbon residue; iv) kinematic viscosity; v) flash point; and vi) P-value. The results obtained in predicting these proprieties with  $^1\text{H}$  NMR spectroscopy will be presented below.

**i) Sulfur content (% m/m)**

To determine the sulfur content of fuel oil stream using  $^1\text{H}$  NMR spectroscopy, 213 samples were analysed. These samples exhibit a wide variety of results including fuel oil samples from cogeneration fuel oil, bunker fuel oil and some fuel oil produced at Sines refinery. The 213 samples were divided into three different groups including: i) training data set with 127 fuel oil samples; ii) the validation data set with 43 fuel oil samples; and iii) the test data set composed of 43 fuel oil samples. All three different groups were representative of the data using the procedure described in the Section 5.3. To these samples, PCR, PLS and ANN models were applied. Before applying PCR and ANN models, a PCA model was implemented. The statistical performance indexes were used to evaluate the performance of the models.

Prior to multivariate modelling, the spectral area used was divided in constant segments of 0.02 ppm. The choice of using a bin width of 0.02 ppm were acquired after analysing the results of the statistical parameters obtained with other bins (bin width 0.02, 0.03 and 0.04). As can be seen in Table 6.5, it was with a bin width of 0.02 ppm that better predictions were obtained, especially for PLS and ANN models. However, for the PCR model, it was with a bin width of 0.03 and 0.04 ppm where slightly better results at the training and validation data sets were obtained. Although, when the PCR model was applied for the training and validation data sets, satisfactory results were obtained independently of the bin used. Besides these small differences, especially for the PCR model, it was considered appropriate using a bin width 0.02 ppm in comparison with bin widths of 0.03 and 0.04 ppm. Moreover, it was observed that it was with a bin width 0.02 ppm that higher number of principal components were selected, due to the subdivision of the spectra into a higher number of regions, and it was with these 13 PCs that a higher percentage of the total variance of the original data was explained.

**Table 6.5:** Performance indexes achieved using PCR, PLS and ANN models when applying different bins (bin = 0.02, 0.03 and 0.04).

			number PC	number LV	number nodes	total variance	R <sup>2</sup> /NSE	MAE	RMSE	MBE	PBIAS	RSR	model performance
<b>bin = 0.02</b>	PCR	training and validation	13		3	97.4%	0.59	0.21	0.28	0.00	0.00	0.64	satisfactory
		test					0.47	0.23	0.32	-0.02	-2.74	0.73	unsatisfactory
	ANN	training and validation					0.96	0.06	0.09	-0.0005	-0.05	0.20	very good
		test					0.91	0.10	0.14	0.02	2.46	0.30	very good
	PLS	training and validation					0.99	0.02	0.03	0.00	0.00	0.06	very good
		test					0.92	0.06	0.12	0.003	0.34	0.28	very good
<b>bin = 0.03</b>	PCR	training and validation	10		3	96.8%	0.62	0.19	0.27	0.00	0.00	0.61	satisfactory
		test					0.51	0.21	0.31	-0.04	-4.31	0.70	satisfactory
	ANN	training and validation					0.95	0.07	0.10	-0.002	-0.28	0.23	very good
		test					0.90	0.10	0.14	0.01	1.42	0.32	very good
	PLS	training and validation					0.99	0.03	0.04	0.00	0.00	0.09	very good
		test					0.74	0.08	0.22	-0.02	-2.63	0.51	good
<b>bin = 0.04</b>	PCR	training and validation	9		3	97.0%	0.62	0.20	0.27	0.00	0.00	0.61	satisfactory
		test					0.41	0.22	0.34	-0.03	-3.6	0.77	unsatisfactory
	ANN	training and validation					0.94	0.08	0.11	0.001	0.15	0.24	very good
		test					0.83	0.11	0.18	0.03	3.65	0.41	very good
	PLS	training and validation					0.99	0.03	0.04	0.00	0.00	0.08	very good
		test					0.85	0.09	0.17	-0.0	-1.27	0.39	very good

The principal components selected, corresponding to 97.4% of the original data variance, were used as inputs of the PCR and ANN models. PCR considered that the variables which most influence the predicted sulfur content were only 6 of the 13 PCs, corresponding to parameters which present statistical significance. Using the PCR model, performance indexes for the test data set were worse than the results obtained at the training and validation data sets. Analysing the Table 6.6 and taking into account the general performance ratings developed by Moriasi *et al.* (2007) for the test data set, unsatisfactory results were obtained while for the training and validation data sets satisfactory results were achieved. An unsatisfactory result was obtained in the test data set due to the lower quality of the performance indexes. Comparing all data sets, the test data set presented small NSE value and a higher RMSE and RSR values. Analysing the PBIAS and MBE, it was demonstrated that for the test data set, the predicted results presented a significant difference when compared to the measured sulfur content, which means that the predicted results were underestimated. These results demonstrated that PCR did not predict the values of the test data set and so this model was not robust enough to predict such wide variety of results.

**Table 6.6:** Performance indexes achieved when using PCR model during training, validation and test data sets, to predict the sulfur content of fuel oil samples.

	PCR model	
	training and validation	test
R <sup>2</sup> /NSE	0.59	0.47
MAE	0.21	0.23
RMSE	0.28	0.32
MBE	0.00	-0.02
PBIAS	0.00	-2.74
RSR	0.64	0.73
Model performance	<b>satisfactory</b>	<b>unsatisfactory</b>

ANN models with different network structures (different numbers of nodes at the hidden layer) were tested to find the best network for predicting the different properties. Taking into account the information from Table 6.7 and from the training and validation data sets, the best network was achieved with 3 nodes at the hidden layer. Besides the very good

models obtained using 2, 3 and 4 nodes at the hidden layer, the best performances were obtained with 3 nodes (see Table 6.7). The network used was characterized for having the same number of inputs as the number of PCs used, one hidden layer with three nodes and one output layer with one node. The network had been trained during one thousands of attempts and the chosen network corresponded to the network which the performance indexes achieved when using 3 nodes at the hidden layer, are presented in Table 6.7.

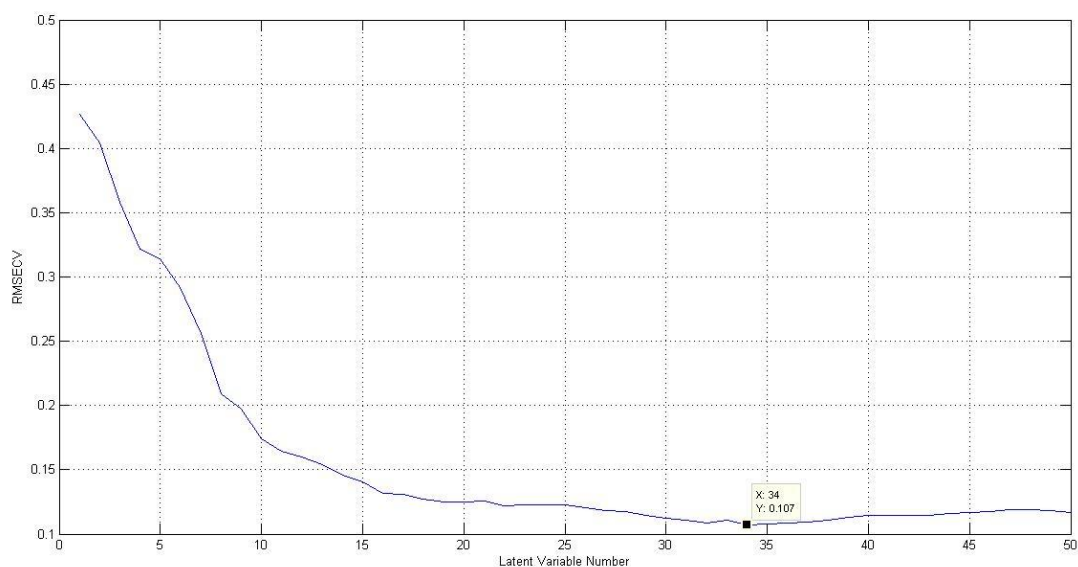
**Table 6.7:** Performance indexes achieved when changing the number of nodes at the hidden layer in the ANN model.

	ANN model					
	2 nodes		3 nodes		4 nodes	
	training and validation	test	training and validation	test	training and validation	test
R <sup>2</sup> /NSE	0.93	0.89	0.96	0.91	0.95	0.84
MAE	0.07	0.10	0.06	0.10	0.06	0.12
RMSE	0.11	0.15	0.09	0.14	0.10	0.18
MBE	-0.004	0.030	-0.001	0.020	0.000	0.040
PBIAS	-0.51	2.91	-0.05	2.46	0.34	4.24
RSR	0.26	0.34	0.20	0.30	0.22	0.40
Model performance	<b>very good</b>	<b>very good</b>	<b>very good</b>	<b>very good</b>	<b>very good</b>	<b>very good</b>

As indicated in Table 6.7, very good performances were achieved for all data sets. The performance indexes calculated were not very different in both data sets, with the exception of the PBIAS value. The PBIAS value was higher at the test data set indicating that the predicted results were overestimated, while the predicted results for the training and validation data sets were underestimated. This PBIAS difference may occur because, in the test data set, the percentage of samples was lower than that of the training and validation data sets, meaning that, besides the equivalent data sets, the number of samples encompassed in each data set may influence the PBIAS behaviour.

When using the PLS model, the original variables were used as inputs, and the calculation of the number of latent variables was required for predicting the dependent variable. The choice of the number of latent variables was made taking into account the minima of the RMSECV (root mean squared error of cross validation) as plotted in Figure 6.3. The

RMSECV reached a minimum value for 34 LVs, as visualized in Figure 6.3. However, different models were developed using different number of LVs to compare the performance of the PLS model. Apart from the 34 LVs, 16 and 37 LVs were tested (see Table 6.8). The choice of the 16 and 37 LVs were made taking into account other minimum of the RMSECV plot. It was considered important to analyse the effect of more and less LVs than the number of LVs indicated by the global minimum in the RMSECV plot. When using 16 LVs the performance indexes were for both, training and validation data sets and test data set, slightly worse than the ones obtained with a higher number of LVs. This behaviour was expected since, typically the quality of the prediction increases with the number of latent variables until a certain number of latent variables (Abdi, 2010). Comparing the performance indexes achieved when using 34 and 37 LVs, it was verified that the 34 LVs selected at the training and validation data sets contributed to the best performances at the test data set. This behaviour, a decrease of the quality of prediction with the 37 LVs, can be explained by the increased number of latent variables. When using a higher number of latent variables, if the quality of the prediction decreases this means that the model is overfitting the data, i.e., the model becomes more dependent of the data used at the training data set and will not be useful to fit new observations as the ones included at the test data set (Abdi, 2010). This fact contributes to a better performance of the model for the training and validation data sets and a worse performance for the test data set. As it was expected, for the training and validation data sets, when the NSE value increased with the increased of the latent variables, the RMSE, the MAE and the RSR results decreased.



**Figure 6.3:** RMSECV plot for predicting the number of latent variables.

**Table 6.8:** Performance indexes achieved when changing the number of LVs in the PLS model.

	PLS model					
	16 LVs		34 LVs		37 LVs	
	training and validation	test	training and validation	test	training and validation	test
$R^2$ /NSE	0.98	0.86	0.99	0.92	0.99	0.91
MAE	0.05	0.10	0.02	0.06	0.02	0.07
RMSE	0.06	0.17	0.03	0.12	0.02	0.13
MBE	0.000	-0.010	0.000	0.003	0.000	-0.002
PBIAS	0.00	-1.69	0.00	0.34	0.00	-0.26
RSR	0.15	0.38	0.06	0.28	0.05	0.30
Model performance	<b>very good</b>	<b>very good</b>	<b>very good</b>	<b>very good</b>	<b>very good</b>	<b>very good</b>

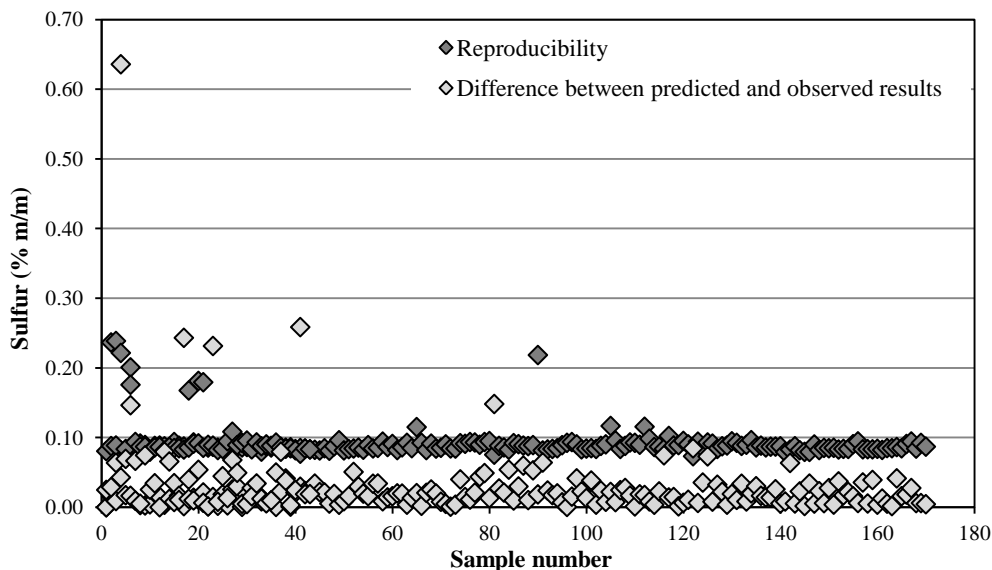
Better performance indexes were obtained when 34 LVs were used for modelling the sulfur content, as visualized in Table 6.8. As illustrated, very good models were obtained for all data sets, training and validation data sets and test data set. A combination of a very high NSE values, and a very small RMSE, MAE and RSR results were indicative of the good agreement between prediction and target values. Only PBIAS and MBE were quite different, although they presented reasonable results. The PBIAS value indicated a slight

overprediction of the predicted values at the test data set. Besides having overestimated results for the test data set, when calculating the reproducibility between all the predicted and observed values (Figure 6.4), it was verified that the difference between the values were respecting the reproducibility limits defined by the standard test method, which are derived statistically from especially designed studies. The reproducibility is a standardized term adopted by ASTM and other standardization organizations, and was a form to evaluate the precision of a test method. In this way, the difference between two single and independent results obtained by different ways, the measured and predicted results, was calculated and then compared with the reproducibility defined by the standard test method. The reproducibility of each standard test method analyses the degree of agreement between the measured and predicted results and support or challenge the validity of the assumption that both test results have been produced on the same material in a correct manner (IP Standards 336, 2004; Coleman *et al.*, 2008). As visualized in Figure 6.4, it was guaranteed that, in average, the predicted sulfur values, the ones experimentally achieved and the ones obtained with the model, were respecting the reproducibility calculated by the equation defined from the IP standard method (Equation 6.1). From all 213 predicted results only 6 results of sulfur content, obtained from the difference between the measured and predicted results, were out of reproducibility, indicating that  $^1\text{H}$  NMR spectroscopy could be used to predict the sulfur content instead of using the IP standard method. The reproducibility defined for the sulfur standard test method was determined using the following equation:

$$\text{Reproducibility} = 0.055 (m + 0.8) \quad (6.1)$$

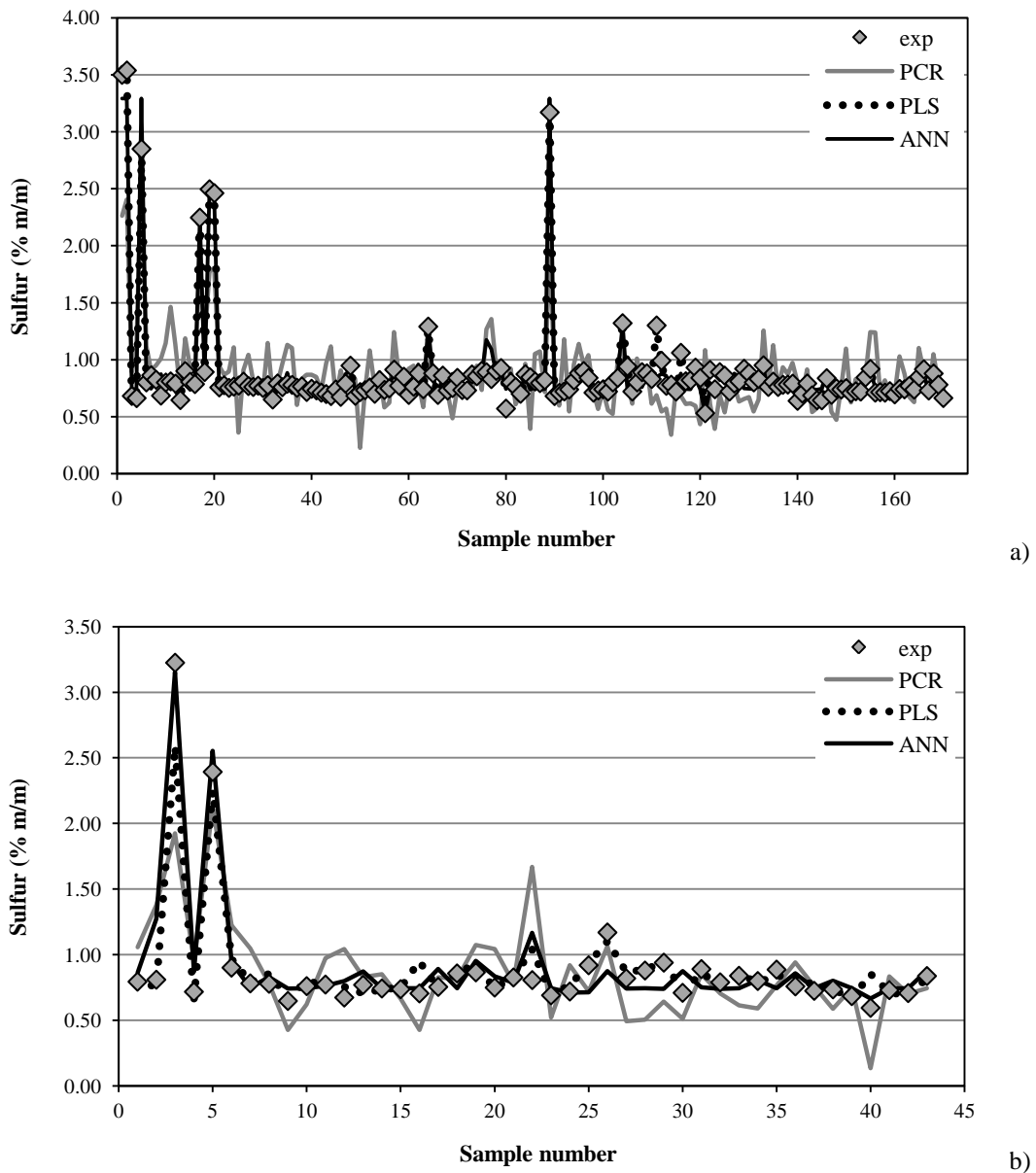
where  $m$  is the mean sulfur content.





**Figure 6.4:** Reproducibility and difference between predicted and observed sulfur content when using PLS model.

Figure 6.5 shows the predicted values obtained by the different multivariate models (PCR, PLS and ANN) and the measured data, corresponding to training and validation data sets and test data set. For the training and validation data sets (Figure 6.5a), better performance indexes were achieved when PLS model was applied. A high quality model was also obtained when the ANN model was used. The worst results were achieved with PCR model. For the test data set (Figure 6.5b), it was shown that PLS model led to slight better predictions as it occurred for the training and validation data sets, comparing to the ANN model. At the test data set, the PCR predictions were not able to predict the sulfur content.

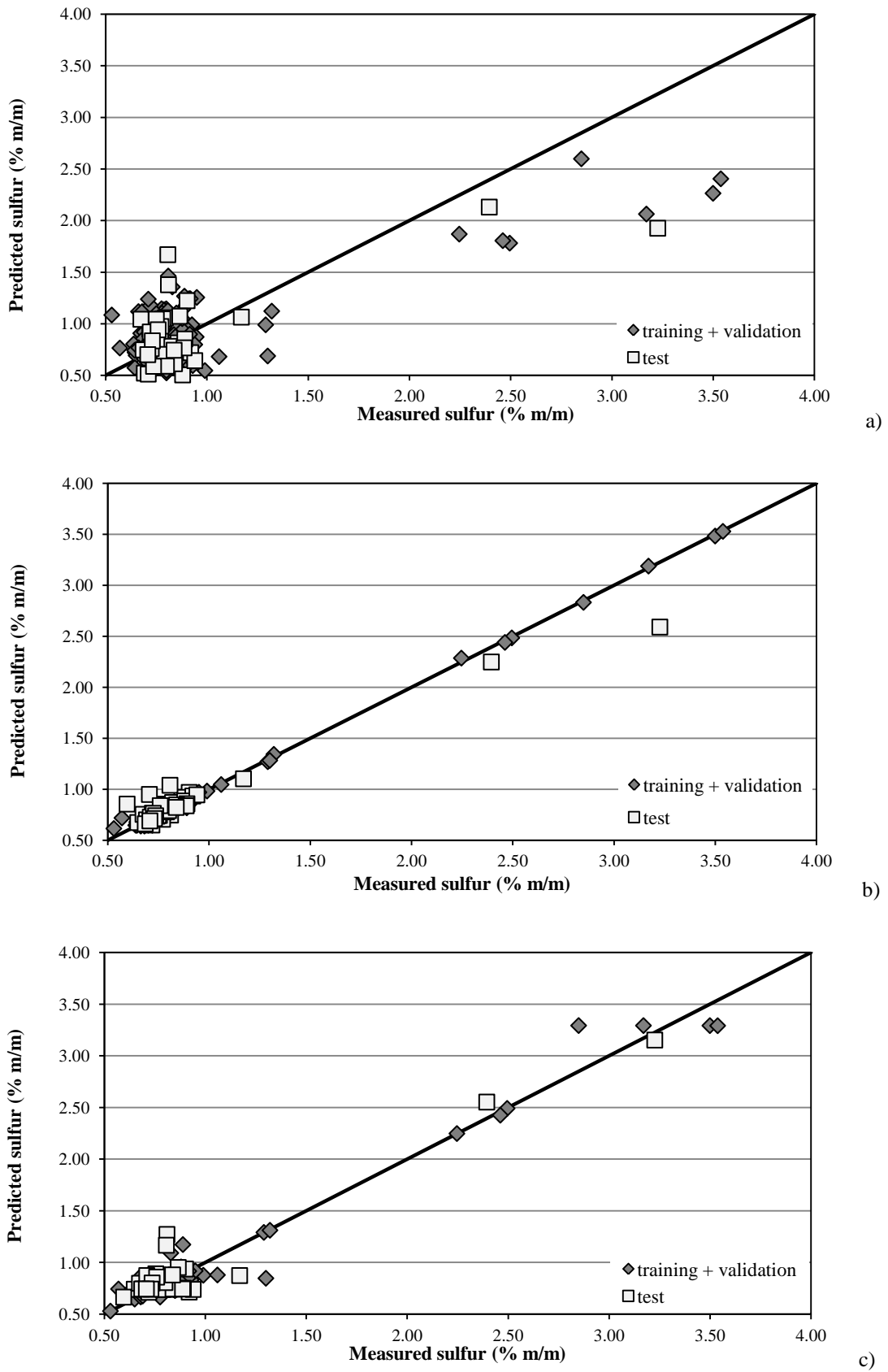


**Figure 6.5:** Prediction of sulfur content for the a) training and validation data sets and b) test data set, when using the three multivariate models.

To corroborate the results formerly presented, Figure 6.6 depicts the linear plot between the predicted sulfur content by all models and the measured sulfur content, corresponding to both, training and validation data sets and test data set. Analysing Figure 6.6 it was possible to confirm that, it was with the PLS model (Figure 6.6b) that the predictions corresponding to both data sets were very close to the ideality, this means, there was a very good linear regression relationship of the type  $y=x$ , between the predicted and the measured sulfur content. For the ANN model (Figure 6.6c), it was evidently a higher

dispersion around the diagonal line, meaning that slightly less accurate models than the ones achieved with the PLS model were obtained, although still of high quality. A low quality model was achieved when the PCR model was applied (Figure 6.6a). A significant difference between the predicted and measured values was evidently, especially when predicting high sulfur values. Consequently, PLS model is more reliable to predict the sulfur content of fuel oil samples giving more accurate results and better performance indexes.

As depicted,  $^1\text{H}$  NMR spectroscopy could be used to predict the sulfur content. The reason for this very good correlation, achieved with the PLS model, may be related with the types of sulfur compounds typically found in fuel oil samples. These compounds may consist of alkyl sulfides and thiols, thiophene, benzothiophene, and dibenzothiophene. Normally the alkyl hydrogens close to the sulfur atom in sulfides resonate in the aliphatic region between 2.1 to 3.0 ppm (Nielsen *et al.*, 2008; Silva *et al.*, 2011).



**Figure 6.6:** Correlation between the experimental and predicted sulfur results of the fuel oil samples obtained by: a) PCR, b) PLS and c) ANN models.

### **Kinematic viscosity**

Normally, the kinematic viscosity of fuel oil samples is controlled taking into account the kinematic viscosity results at standard temperatures of 50 °C and/or 100 °C. The kinematic viscosity of fuel oil samples can be analysed in terms of both temperatures but in terms of the specifications, the kinematic viscosity of a bunker fuel oil is required, for process control, at 50 °C while the kinematic viscosity of a cogeneration fuel oil required at 100 °C. It was considered important to determine the kinematic viscosity of fuel oil samples at both temperatures, independently of analysing a bunker or a cogeneration fuel oil. The idea was to verify if  $^1\text{H}$  NMR spectroscopy could be used to predict the kinematic viscosity of fuel oil samples at 50 and 100 °C and with that, evaluate the possibility of using an online NMR spectrometer. It is expected to obtain, with just one analysis, information about the kinematic viscosity of fuel oil samples at both temperatures independently of the temperature at which the NMR experiments were performed. In this work, all the NMR experiments were obtained at ambient temperature since deuteriochloroform was the solvent selected to prepare the samples. When using deuteriochloroform, the NMR spectra were always recorded at room temperature due to its characteristics, such as, its boiling point.

#### **i) Kinematic viscosity at 100 °C (cSt)**

To analyse the possibility of using an online NMR spectrometer to determine the kinematic viscosity at 100 °C, 217 fuel oil samples were analysed. These 217 samples included different fuel oils as cogeneration fuel oil, bunker fuel oil and some fuel oil from Sines refinery. As a result, with all these samples, a wide variety of kinematic viscosity results ranging from 14.20 to 1367 cSt was guaranteed. From the 217 fuel oil samples, 131 samples represented the training data set and 43 samples constituted the validation and test data sets, respectively. As stated previously, all different data sets were representative of the data and the procedure used to guarantee such division was described in Section 5.3. Before applying multivariate modelling, a bin width 0.02 ppm to divide the spectral area in constant segments was used. Subsequently, multivariate modelling and the statistical parameters indexes were used.

Table 6.9, presents the results obtained when applying the multivariate modelling, PCR, PLS and ANN, to fuel oil samples. Fourteen principal components were used as inputs of PCR and ANN. The percentage of the original data variance contained in the 14 PCs was 97.8%. For the ANN model, it was with 3 nodes at the hidden layer that the best results were produced. When using PLS model, the minimum of the RMSECV plot indicated that 35 LVs to achieve the best results were required.

**Table 6.9:** Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the viscosity at 100 °C of fuel oil samples.

	PCR		PLS		ANN	
	training and validation	test	training and validation	test	training and validation	test
R <sup>2</sup> /NSE	0.57	0.34	0.99	0.51	0.96	0.94
MAE	55.19	58.95	12.12	51.56	16.15	16.53
RMSE	104.17	85.47	15.68	74.06	31.95	24.49
MBE	0.00	3.61	0.0004	7.55	0.52	2.69
PBIAS	0.00	5.27	0.00	11.00	0.69	3.93
RSR	0.66	0.81	0.10	0.70	0.20	0.25
Model performance	<b>satisfactory</b>	<b>unsatisfactory</b>	<b>very good</b>	<b>satisfactory</b>	<b>very good</b>	<b>very good</b>

Analysing Table 6.9 it was possible to conclude that better performance indexes were achieved with ANN model for both data sets. Very good NSE values and small residual variations were obtained when using the ANN model. The RMSE and MAE values obtained, when applying the ANN model, may appear a bit high, but taking into account the measured kinematic viscosity values, these errors were acceptable. When analysing the MBE and PBIAS, it was verified that for both data sets the predicted values were positive, indicating that, in average, the kinematic viscosity was overestimated. These results showed that the models developed by the ANN model, for both data sets, were of higher quality and characterized with lower error variance.

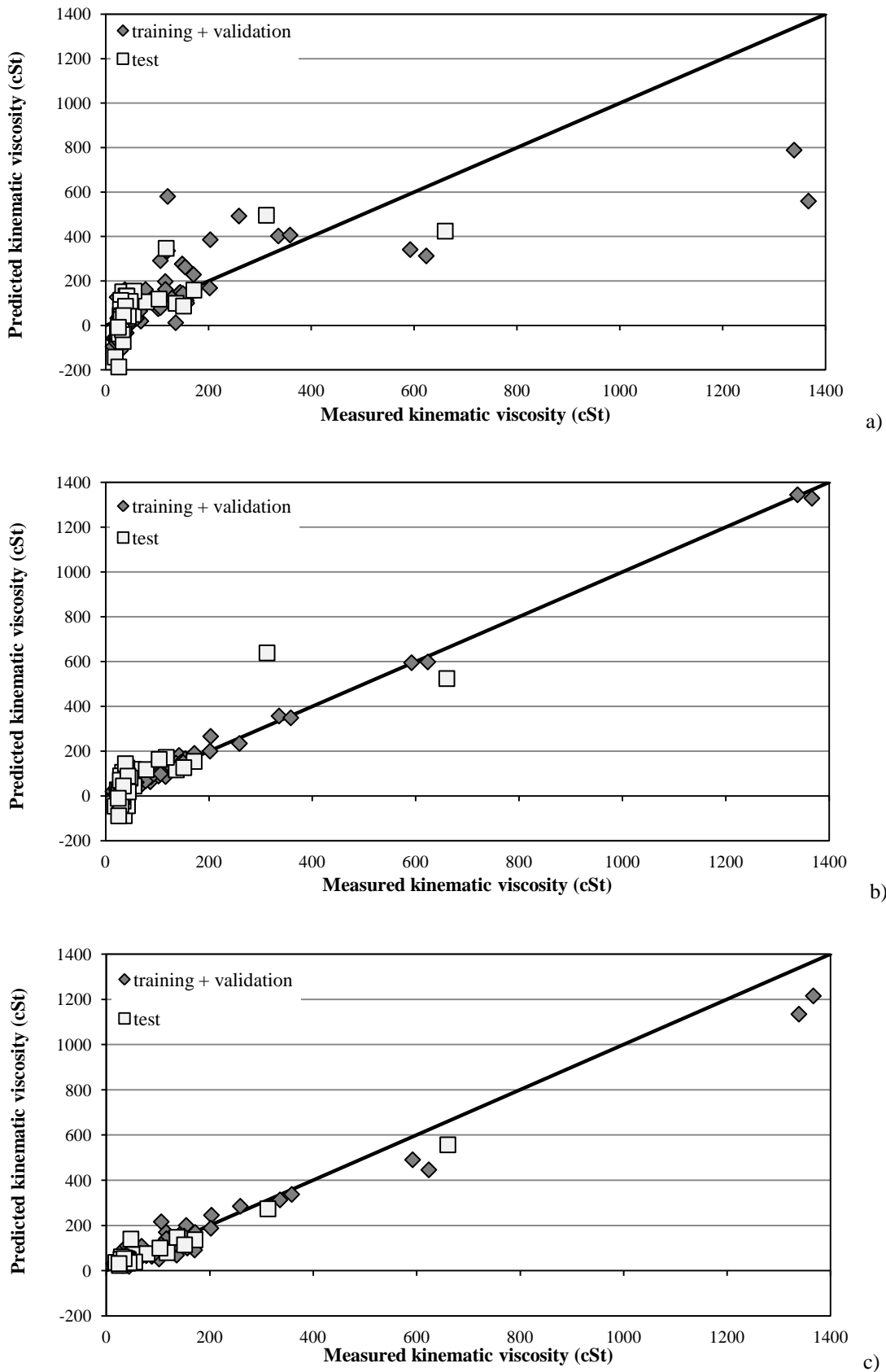
When using the PLS model, for the training and validation data sets, very good results were obtained. However, when predicting the test data set only a satisfactory model was achieved. Even the decrease of the LVs could not improve the final results of the test data

set. For this reason the PLS was not considered a possibility for predicting the kinematic viscosity of fuel oils.

The worst performance occurred when applying PCR model. The PCR model failed when predicting the results of the test data set. However, for the training and validation data sets slightly better predictions were obtained, although only a satisfactory model was achieved. Even PBIAS indicating accurate model simulation (PBIAS = 0.0%), the other performance indexes didn't performed as good predictions as the PBIAS value.

In Figure 6.7, the representation is a comparison between the prediction values obtained by the different multivariate models (PCR, PLS and ANN) and the measured data, corresponding to both, training and validation data sets and test data set. With this representation it was possible to confirm the conclusions previously made. The results showed that the use of ANN model led to a more accurate result than PCR and PLS models. PCR was the model that gave the worse performance predicting negative kinematic viscosity values (18%) for both data sets (Figure 6.7a). In the PLS model (Figure 6.7b), a high degree of correlation was evident when predicting the training and validation data sets, whereas the results from the test data set showed larger deviations, compared to kinematic viscosity values of the training and validation data sets. However, better predictions with the PLS model were obtained when compared with the PCR model. For both data sets, the PLS model produced predictions with 94% of positive values. It was possible to conclude that the agreement between the predicted kinematic viscosity results obtained when using the ANN model and the measured kinematic viscosity results was remarkable, showing a high degree of correlation (see Figure 6.7c). Only the samples with high kinematic viscosity were slightly deviated from what was expected. The ANN model was the technique which led to better predictions, and was considered more reliable to predict the kinematic viscosity of fuel oil samples with a good accuracy.

Viscosity can be used to obtain information about the composition of fuel oil samples. As example, a paraffinic fraction normally is characterized for having a low viscosity. The direct link to the compositional information contained in the NMR spectra resulted, as expected, in good kinematic viscosity predictions of the fuel oil samples.



**Figure 6.7:** Correlation between the experimental and predicted kinematic viscosity results of the fuel oil samples obtained at 100 °C by: a) PCR, b) PLS and c) ANN models.



## ii) Kinematic viscosity at 50 °C (cSt)

To determine the kinematic viscosity at 50 °C, 108 fuel oil samples were analysed. From all the samples analysed, the training data set included 64 samples while the validation and test data sets were formed by 22 fuel oil samples, respectively. Both data sets were representative of the data. With all samples analysed, bunker and cogeneration fuel oils, a wide variety of results ranging from 103.2 to 890.4 cSt were guaranteed. As occurred in the kinematic viscosity at 100 °C, a bin width 0.02 ppm, multivariate modelling and statistical parameters indexes were also used.

The results obtained when applying the multivariate modelling, PCR, PLS and ANN models, and the statistical parameters indexes are presented in Table 6.10. PCR and ANN models used the PCs as inputs whereas the PLS model was based on the original data. The number of PCs considered by principal component analysis was 15 PCs. These PCs were responsible for explaining 98.3% of the original data variance. To apply the ANN model, 3 nodes at the hidden layer were used while for the PLS model 9 LVs, based on the minimum of the RMSECV plot, were required to obtain the best results.

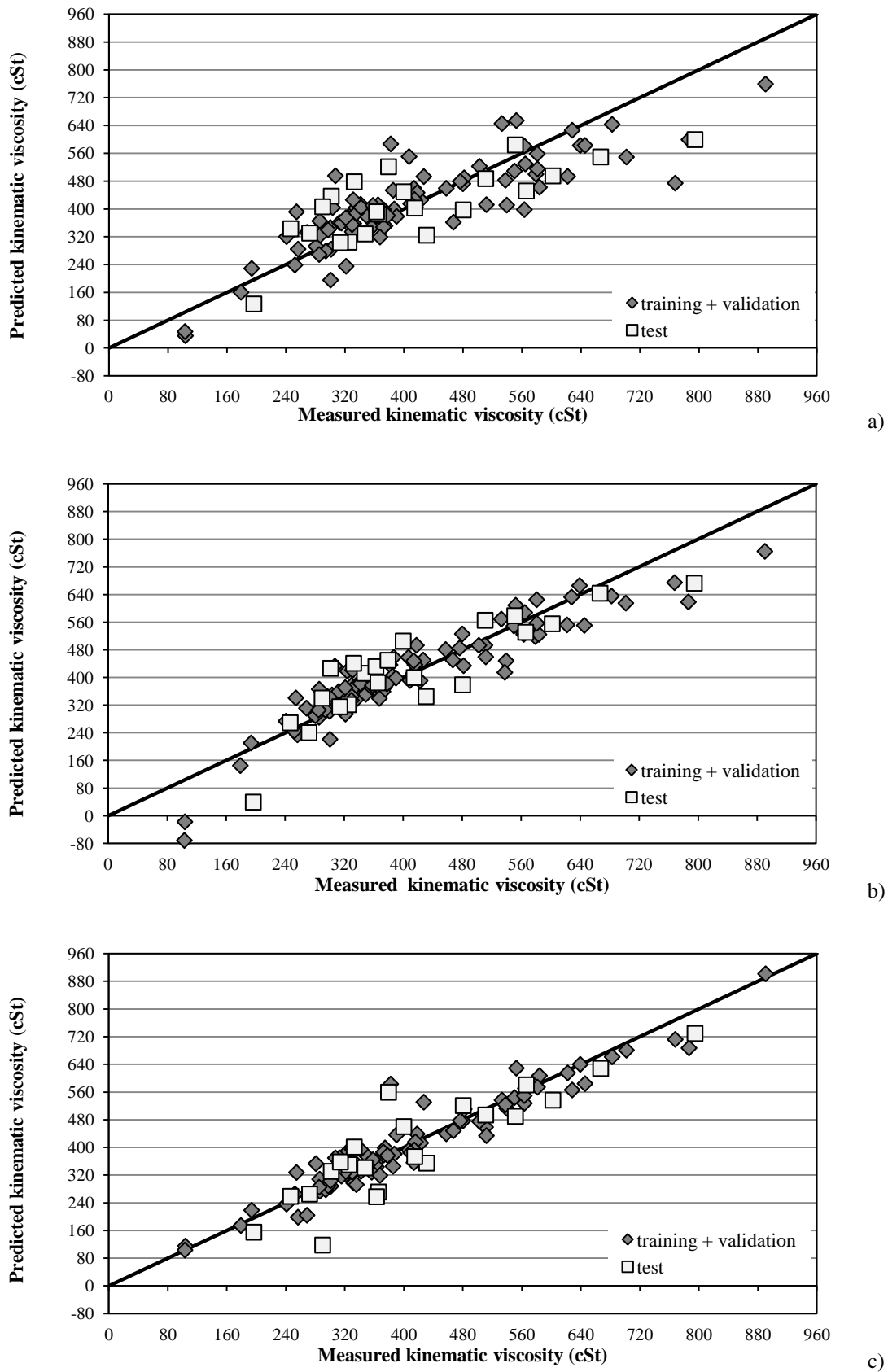
**Table 6.10:** Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the viscosity at 50 °C of fuel oil samples.

	PCR		PLS		ANN	
	training and validation	test	training and validation	test	training and validation	test
R <sup>2</sup> /NSE	0.71	0.58	0.85	0.73	0.92	0.74
MAE	58.70	77.54	42.70	60.94	28.68	57.54
RMSE	79.78	93.39	56.84	74.47	42.47	73.61
MBE	0.00	-2.72	0.00	4.16	1.08	-14.46
PBIAS	0.00	-0.65	0.00	1.00	0.26	-3.48
RSR	0.54	0.65	0.39	0.52	0.29	0.51
Model performance	<b>good</b>	<b>satisfactory</b>	<b>very good</b>	<b>good</b>	<b>very good</b>	<b>good</b>

As occurred in the prediction of the kinematic viscosity at 100 °C, the best model to predict the kinematic viscosity at 50 °C of fuel oil samples was achieved with the ANN model for both data sets, as can be visualized in Table 6.10. It was with the ANN model that smaller

residual variation and higher NSE value were achieved when compared with the respective PCR and PLS results. Also here, the results obtained for the RMSE and MAE, with the ANN model, appear to be a bit higher, but taking into account the measured viscosity values, these errors were acceptable. Analysing PBIAS and MBE results for the ANN model, it was verified that the results were positive for the training and validation data sets, meaning that, in average the predicted kinematic viscosity was overestimated, while for the test data set the predicted kinematic viscosity was underestimated. Taking into account the obtained results (Table 6.10) and following the general performance ratings presented by Moriasi *et al.* (2007) the predicted results obtained with the ANN model for the training and validation data sets were characterized as very good whereas for the test data set the results were characterized as good. The same classification was obtained when using the PLS model, a very good model was achieved for the training and validation data sets while for the test data set a good model was obtained. However, when applying the PCR model, worst predictions were obtained for both data sets. The training and validation data sets were characterized as a good model but the test data set was only classified as satisfactory. As can be concluded, independently of the model applied worst predictions were obtained for the test data set.

The conclusions previously made when analysing Table 6.10 can be confirmed by the following representations. Figure 6.8 shows the kinematic viscosity predictions with all models and the measured data, corresponding to the training and validation data sets and test data set. As visualized, it was possible to conclude that the higher degree of correlation was obtained when the ANN model was applied (see Figure 6.8c). Although, a slight dispersion of the predicted values, when using the ANN model, was notable. The PLS (Figure 6.8b) model gave slightly less accurate results than the ANN model, although still of high quality. However, the main difference between the results predicted with the ANN and the PLS models were related with the negative kinematic viscosities predicted by the PLS model (about 2% of the original data). It was with the PCR model (Figure 6.8a) that the main differences between the measured and predicted results were visualized and less accurate performance indexes were obtained, compared to PLS and ANN models predictions. The results obtained with the PCR model were surely satisfactory, but it was with the ANN model that the agreement between the predicted and measured values was of high quality.



**Figure 6.8:** Correlation between the experimental and predicted kinematic viscosity results of the fuel oil samples obtained at 50 °C by: a) PCR, b) PLS and c) ANN models.

**iv) Density (g/mL)**

To develop a model that contributes to obtain the density results 185 fuel oil samples were analysed. From the 185 samples, 37 were used for the test and validation data sets, respectively, while the training data set was formed with the remaining 111 fuel oil samples. Both data sets were representative of the data. The samples used to develop the model embraced different fuel oils, as the cogeneration fuel oils and bunker fuel oils, and a wide variety of results ranging from 0.9500 to 1.020 g/mL were included.

As occurred at the properties previously described, the spectral area was divided in constant segments of 0.02 ppm. This bin was preferred over other bin width, as already demonstrated, and it was always used for the analysis of fuel oil properties. After the preparation of the spectral data, multivariate modelling was applied and the statistical parameters indexes used to evaluate the performance of the developed models.

Before applying the multivariate models (PCR, PLS and ANN), PCA was used. With PCA the number of principal components that were used as inputs of the PCR and the ANN models was identified. For predicting density (g/mL), 15 PCs were required. These PCs were responsible for explaining 98.1% of the original data variance. The ANN model was characterized for having 3 nodes at the hidden layer. As already demonstrated the best network performance was achieved when using the 3 nodes. For all the fuel oil properties analysed, 3 nodes at the hidden layer were always used. When applying PLS model, 6 LVs were necessary to obtain better performance indexes.

Analysing Table 6.11 it was possible to conclude that very good predictions were achieved for both data sets with all multivariate models used. However, it was with ANN model that higher performance indexes were obtained for both data sets. When compared with PCR and PLS, it was with the ANN model that higher NSE values were obtained, indicating higher quality models for both data sets. As expect, due to the increase of the coefficient of determination obtained with the ANN model, smaller error indexes were achieved. A decrease of the RSR and RMSE values was indicative of a better model simulation performance. By analysing PBIAS and MBE results for the ANN model, it was found that these results were positive for both data sets, meaning that, in average the predicted density of fuel oil was overestimated. However, taking into account the results obtained for the

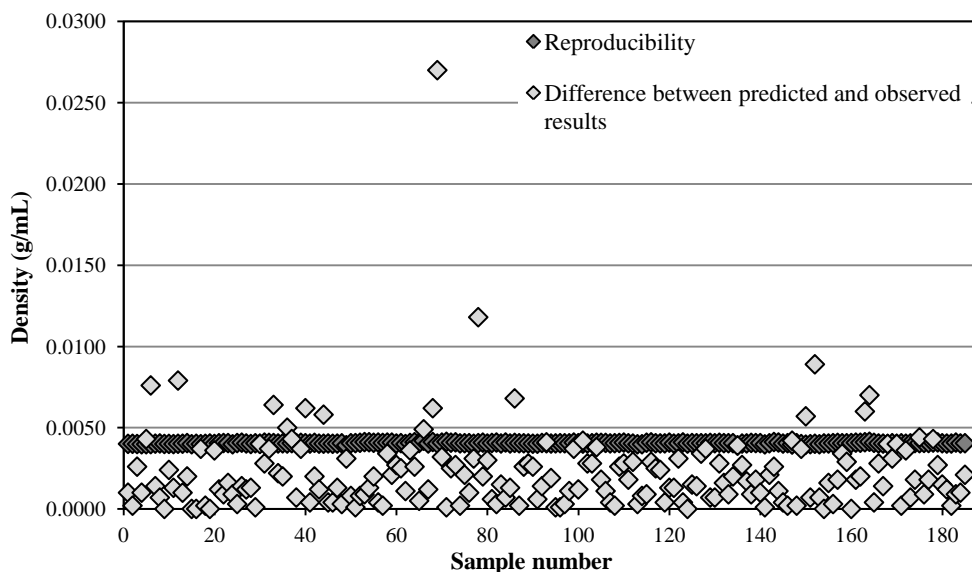
reproducibility, Figure 6.9, it was possible to conclude that almost the predicted results obtained with the ANN model respected the reproducibility established by the ASTM standard method. These results indicate that the model can be used to predict density results. The reproducibility defined by the standard test method for density results was calculated by means of Equation 6.2:

$$\text{Reproducibility} = 0.00412 m \quad (6.2)$$

where  $m$  represents the sample mean.

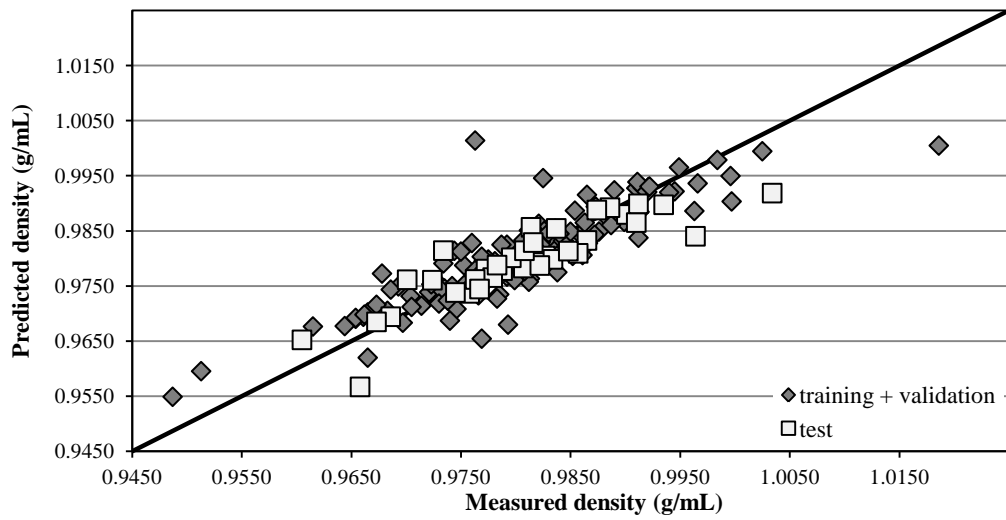
**Table 6.11:** Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the density of fuel oil samples.

	PCR		PLS		ANN	
	training and validation	test	training and validation	test	training and validation	test
R <sup>2</sup> /NSE	0.77	0.75	0.79	0.77	0.85	0.86
MAE	0.003	0.003	0.003	0.003	0.002	0.002
RMSE	0.004	0.004	0.004	0.004	0.004	0.003
MBE	0.000	-0.001	0.000	-0.001	0.000	0.000
PBIAS	0.00	-0.12	0.00	-0.11	0.03	0.02
RSR	0.48	0.50	0.45	0.48	0.38	0.37
Model performance	<b>very good</b>	<b>very good</b>	<b>very good</b>	<b>very good</b>	<b>very good</b>	<b>very good</b>

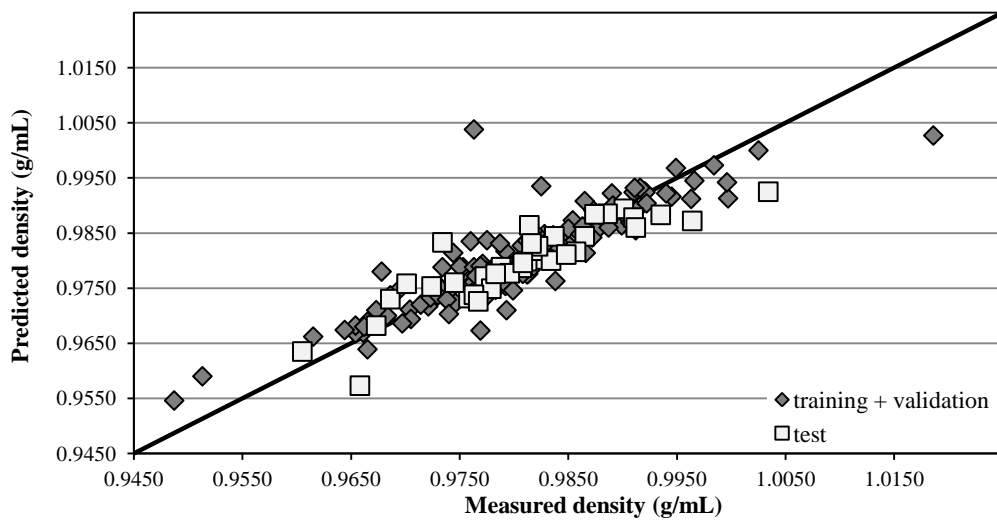


**Figure 6.9:** Reproducibility and difference between predicted and observed density when using ANN models.

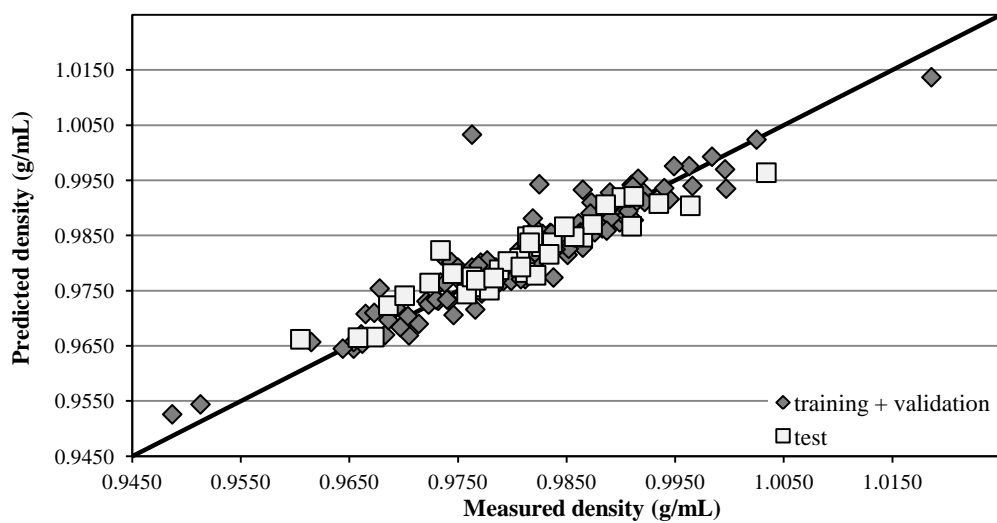
When analysing the correlation plot between the predicted and measured density results obtained with all models applied (Figure 6.10), it was possible to conclude that high quality models were achieved. However, a high degree of correlation was obtained when the ANN model was applied (Figure 6.10c). A good dispersion was notable along the  $y=x$  line and the number of results that were far from the graphic diagonal line were scarce, when using the ANN model. When applying the PCR and PLS models higher dispersion around the diagonal was notable, although very good predictions were also possible to be obtained when applying both models (see Figure 6.10a,b). The conclusion was that the NMR combined with the ANN model could be used to predict acceptable and very good results of fuel oil density. Once, the density of fuel oil samples was correlated with chemical composition. Highly aromatic fuel oils are predicted to have a high density while fuel oils containing a high fraction of aliphatic compounds tend to have a lower density.



a)



b)



c)

**Figure 6.10:** Correlation between the experimental and predicted density results of the fuel oil samples obtained by: a) PCR, b) PLS and c) ANN models.

**v) Carbon residue (% m/m)**

To evaluate the possibility of analysed carbon residue of fuel oils by using an online NMR spectrometer, 103 samples were studied. With these samples a wide variety of results was guaranteed with the carbon residue values ranging from 12.9 to 22.1% m/m. With all the samples analysed, three different groups were developed as the training data set including 61 samples and the validation and test data sets formed with 21 samples, respectively.

The multivariate modelling and the statistical performance indexes were applied. PCR and ANN were based on the PCs while PLS was based on the original data. Considering PCA, 15 PCs were select and were responsible for 98.4% of the total variance. To apply the ANN model, the best architecture was achieved with 3 nodes at the hidden layer. When using the PLS model, better predictions were obtained when 7 LVs were selected.

Table 6.12 presents the values of the performance indexes using PCR, PLS and ANN models for both training and validation data sets and test data set. The results obtained demonstrated that when using the ANN model better performance indexes were obtained for both data sets. Additionally, good predictions were obtained for the training and validation data sets when PCR and PLS were applied. However, the problem associated with these models was found when applied to the test data set, where unsatisfactory results were achieved. The test data set of both models, PCR and PLS, was characterized for having a small NSE value and high errors. On the other hand, when using the ANN model a higher NSE value was obtained, for both data sets, and a decrease of the RMSE, MAE and RSR values was notable when compared with PCR and PLS (see Table 6.12). The results obtained with the ANN model for the training and validation data sets were underestimated while for the test data set the results were overestimated. Although, the results obtained (Figure 6.11) were, in majority, respecting the reproducibility of the ASTM standard method for determining the carbon residue (Equation 6.3):

$$\text{Reproducibility} = 4.68 \times 10^{-2} (\% \text{ Micro Carbon Residue} + 3) \quad (6.3)$$



**Table 6.12:** Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the carbon residue of fuel oil samples.

	PCR		PLS		ANN	
	training and validation	test	training and validation	test	training and validation	test
R <sup>2</sup> /NSE	0.65	0.37	0.75	0.31	0.77	0.70
MAE	0.85	1.28	0.74	1.30	0.54	0.94
RMSE	1.09	1.56	0.93	1.63	0.88	1.07
MBE	0.00	-0.49	0.00	-0.49	-0.06	0.26
PBIAS	0.00	-2.87	0.00	-2.85	-0.36	1.50
RSR	0.59	0.79	0.50	0.83	0.48	0.55
Model performance	<b>good</b>	<b>unsatisfactory</b>	<b>very good</b>	<b>unsatisfactory</b>	<b>very good</b>	<b>good</b>

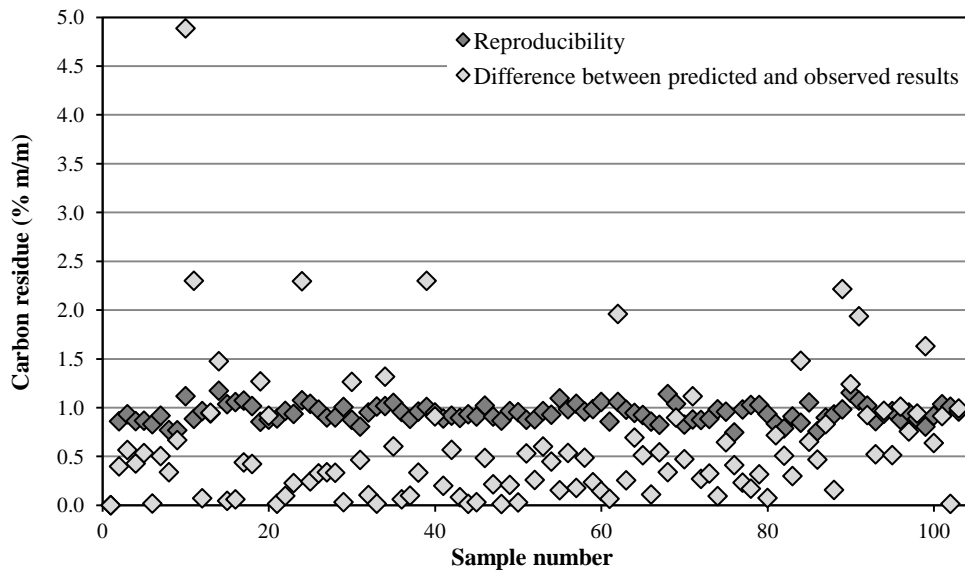
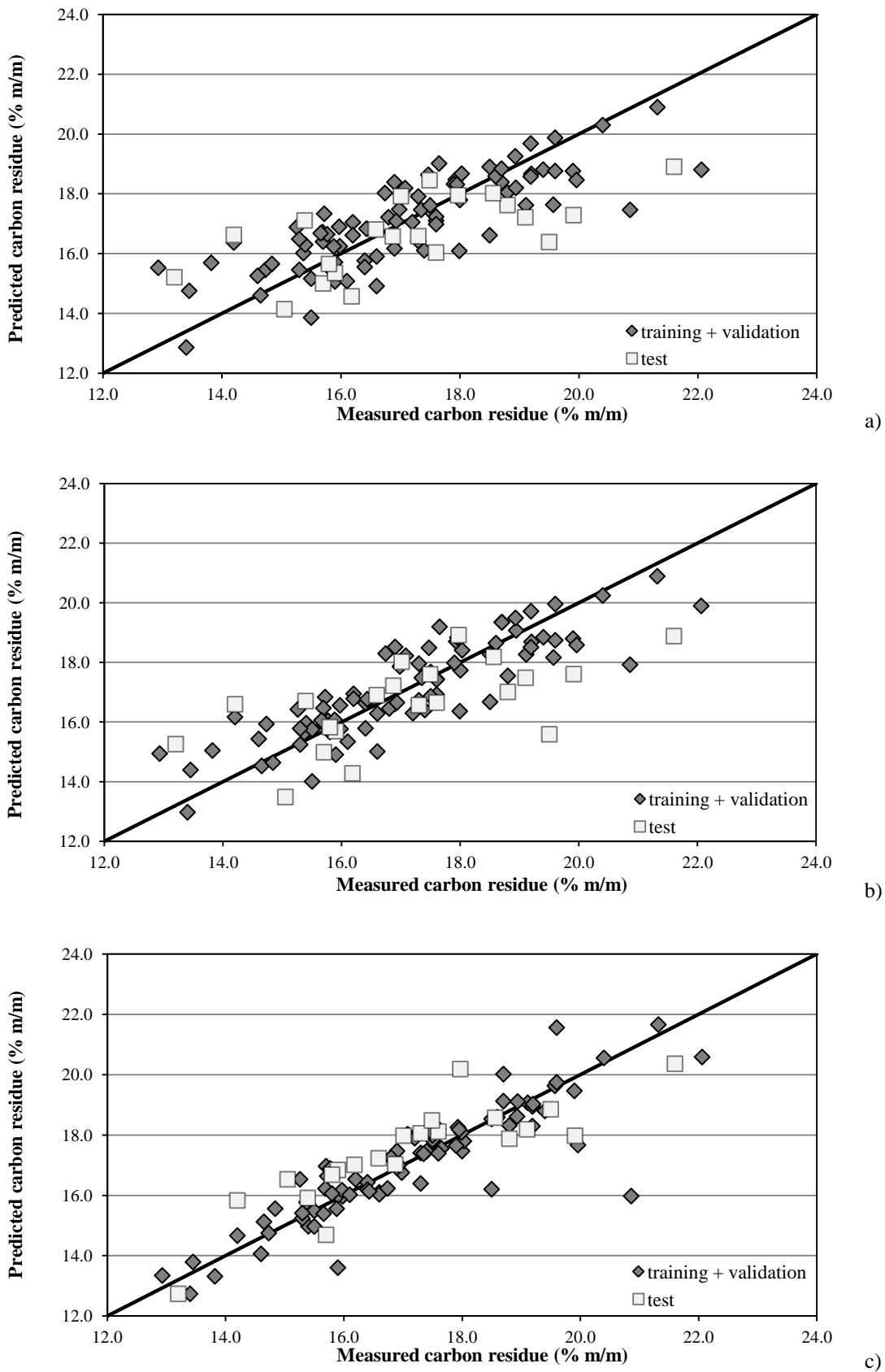
**Figure 6.11:** Reproducibility and difference between predicted and observed carbon residue when using ANN models.

Figure 6.12, showing the measured versus predicted results obtained for the training and validation data sets and test data set when applying the different multivariate models, illustrates the range of the dispersion. It is observed that the ANN model offers the best performance indexes when predicting the carbon residue of fuel oil samples. It was with the ANN model that the predicted results were closest from the diagonal symmetric line, when compared against experimental data (see Figure 6.12c). However, some dispersion around the diagonal line, especially for the higher carbon residue values, was notable.

Although, this dispersion around the diagonal line was more evident when applying the PCR and PLS models (see Figure 6.12a,b), with the models always overestimating or underestimating the results. Then, it is possible to conclude that the  $^1\text{H}$  NMR spectroscopy when combined with the ANN model could be used to predict the carbon residue of fuel oil samples.

NMR spectroscopy was always considered a good technique to predict the carbon residue once the carbon residue, related to the propensity of the fuel oil samples to form carbon deposits, was also related with the chemical composition (Nielsen *et al.*, 2008). The propensity to form carbon deposits depends on the  $H/C$  ratio and on the content of asphaltenes. The asphaltenes are characterized for having aromatic and aliphatic compounds in its structure, which are different types of hydrocarbons quantified by NMR spectroscopy. The smaller is the  $H/C$  ratio, the higher is the percentage of carbon residue.



**Figure 6.12:** Correlation between the experimental and predicted carbon residue results of the fuel oil samples obtained by: a) PCR, b) PLS and c) ANN models.

**vi) Flash point (°C)**

The study performed to analyse the prediction of the flash point of fuel oil samples by  $^1\text{H}$  NMR spectroscopy was developed considering 157 samples including cogeneration fuel oils and bunker fuel oils. The training data set was constituted by 95 samples while the validation and test data sets included 31 samples, respectively. The flash point of the fuel oil samples analysed varied between 49 to 102 °C.

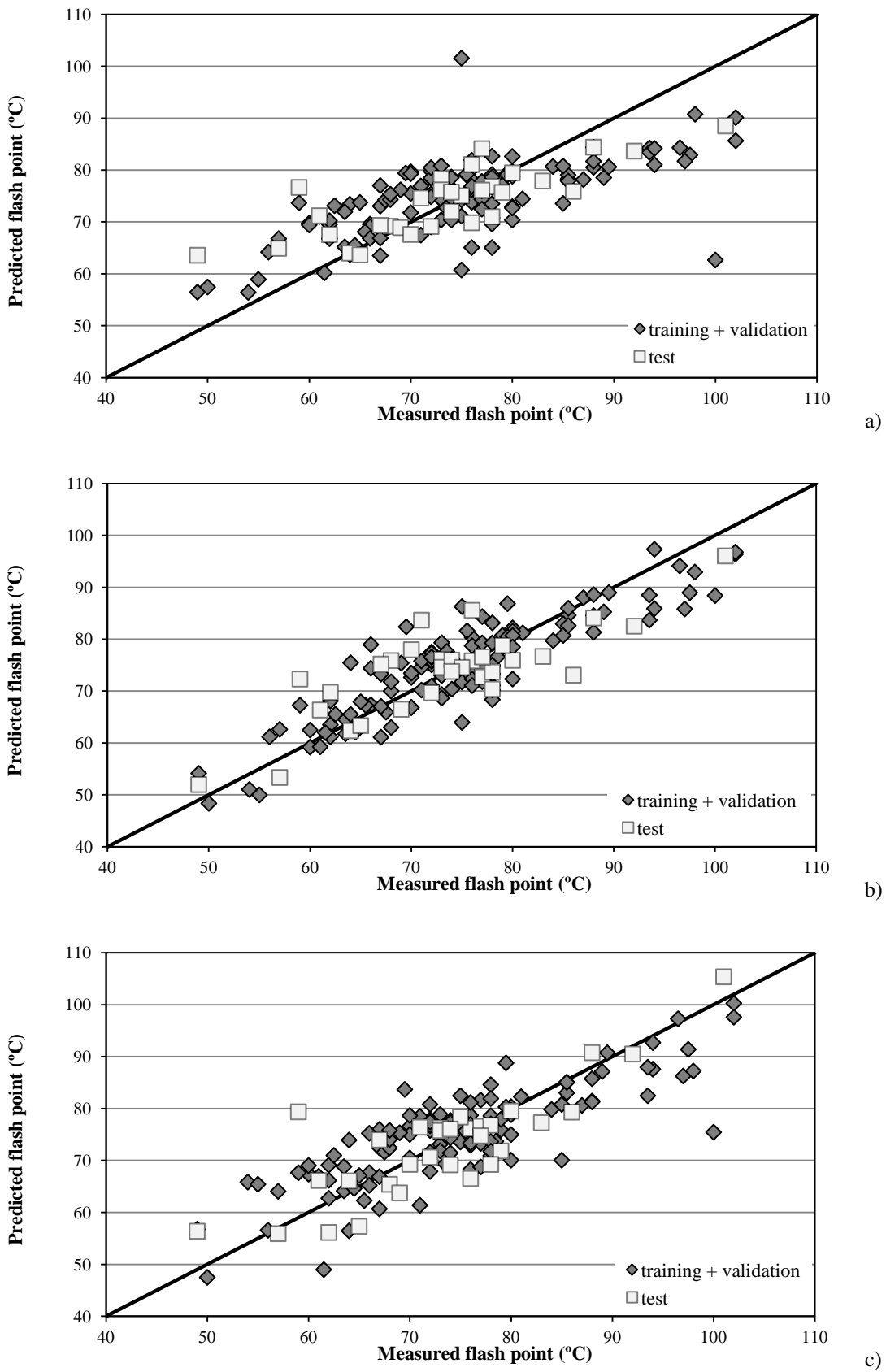
PCR, PLS and ANN models were used to predict the flash point of the fuel oil samples analysed and the statistical performance indexes were applied to evaluate the performance of the developed models. PCR and ANN models were based on the PCs whereas for the PLS model the original data was used as inputs. After applying the PCA, 15 PCs were selected corresponding to 98.3% of the original data variance. Considering the ANN model the best performance indexes were obtained when using 3 nodes at the hidden layer. Taking into account the PLS model, the best model was obtained when using 17 LVs.

The results obtained when applying the statistical performance indexes for the different multivariate models (PCR, PLS and ANN) and for both data sets, training and validation data sets and test data set are presented in Table 6.13. The worst performance occurred when PCR was applied for both data sets. A slight improvement was notorious in the test data set, but small NSE value and high errors continued to characterize both data sets, when the PCR model was applied. When using the PLS model better predictions were obtained, when compared with PCR results. A very good model was obtained for the training and validation data sets indicating that high NSE value and small errors as RMSE, MAE and RSR occurred. However, for the test data set only satisfactory results were obtained. Comparing the three models, it was with ANN model that the best predictions for both data sets were obtained. Good coefficients of determination and small errors characterized the performance indexes obtained for both data sets, when ANN was applied. The MBE and PBIAS were positive for the training and validation data sets indicating that the predicted flash point was overestimated, while, for the test data set were negative, meaning that, in average, the predicted flash point was underestimated.

**Table 6.13:** Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the flash point of fuel oil samples.

	PCR		PLS		ANN	
	training and validation	test	training and validation	test	training and validation	test
R <sup>2</sup> /NSE	0.44	0.60	0.79	0.64	0.67	0.67
MAE	5.97	4.93	3.86	4.95	4.66	4.50
RMSE	7.90	6.59	4.86	6.26	6.06	5.95
MBE	0.00	0.57	0.00	0.35	0.31	-0.22
PBIAS	0.00	0.77	0.00	0.48	0.42	-0.30
RSR	0.75	0.63	0.46	0.60	0.58	0.57
Model performance	<b>unsatisfactory</b>	<b>satisfactory</b>	<b>very good</b>	<b>satisfactory</b>	<b>good</b>	<b>good</b>

In Figure 6.13 a comparison between all measured and predicted flash point results obtained at the training, validation and test data sets when applying all models is represented. It is shown that the ANN model was considered the best model to predict the flash point of fuel oil samples, showing a good degree of correlation for the training and validation data sets and test data set, simultaneously. The correlation obtained illustrated a good agreement between the predicted and measured flash point results, although, as visualized in Figure 6.13c, a slight dispersion in the results around the diagonal line was notorious. This dispersion between the predicted and measured flash point results was verified independently of the model applied, however it was mainly pronounced when the PCR model was used. Figure 6.13a shows that the PCR model was not a good alternative to estimate the flash point of fuel oil samples. When applying the PLS model, a better agreement between the predicted and measured flash point results was obtained (Figure 6.13b), although for the test data set larger deviations were achieved and consequently low quality predictions. Therefore, the <sup>1</sup>H NMR spectroscopy combined with ANN model was considered the more reliable approach to predict the flash point of fuel oil samples, giving the idea that good performance indexes could be obtained. The flash point is also another property that depends on the chemical composition of fuel oil samples. Normally, samples with a higher aromatic content present a higher flash point.



**Figure 6.13:** Correlation between the experimental and predicted flash point results of the fuel oil samples obtained by: a) PCR, b) PLS and c) ANN models.

**vii) P-value**

To determine the P-value, 217 fuel oil samples were studied. From all variables, the training data set included 131 samples whereas the validation and test data sets were constituted by 43 fuel oil samples, respectively. A wide variety of results were covered with the P-value ranging from 1.00 to 1.8.

To predict the P-value of the fuel oil samples, PCR, PLS and ANN models were applied. To evaluate the performance of the developed models the statistical performance indexes were used. PCR and ANN models used the PCs as inputs, while PLS was based on the original data. The number of PCs considered by the principal component analysis was 14 PCs, which were responsible for explaining 97.7% of the original data variance. Using these PCs as inputs, the best ANN architecture was achieved when using 3 nodes at the hidden layer. When using the PLS model, 26 LVs were considered the most adequate for the PLS model to perform good predictions.

Table 6.14 presents the values of the performance indexes using all multivariate models for both training and validation data sets and test data set. The results obtained indicated that PCR was not a good model to predict the P-value, once unsatisfactory models were obtained for both data sets. However, when using the PLS model, very good predictions were obtained for the training and validation data sets but when predicting the test data set the performance indexes were not as good. Only satisfactory model was obtained for the test data set, indicating that small NSE value and high errors were achieved. Different numbers of latent variables were used, but better predictions were obtained when using the 26 LVs. Additionally, when using the ANN model the same occurred. A good model was obtained for the training and validation data sets but when analysing the test data set worst results were achieved. Nevertheless, the model developed for the test data set was classified as satisfactory. The predictions obtained for the test data set when using the ANN model were better when compared with the ones obtained with PCR and PLS models. Higher NSE value and small errors were obtained with the ANN model for the test data set.

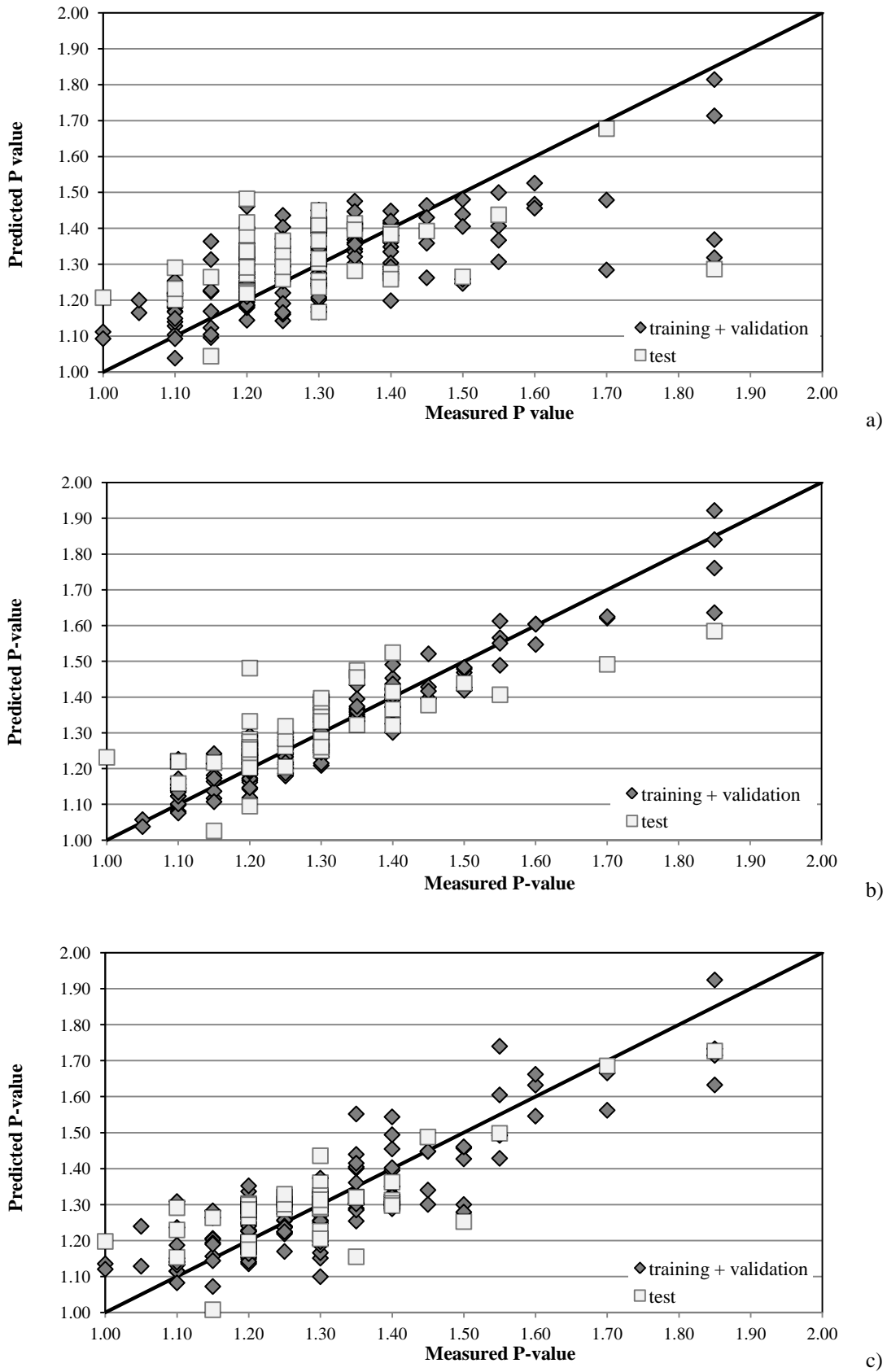
**Table 6.14:** Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the P-value of fuel oil samples.

	PCR		PLS		ANN	
	training and validation	test	training and validation	test	training and validation	test
R <sup>2</sup> /NSE	0.50	0.10	0.91	0.53	0.74	0.62
MAE	0.07	0.11	0.03	0.08	0.06	0.07
RMSE	0.11	0.14	0.04	0.10	0.08	0.09
MBE	0.00	0.03	0.00	0.02	0.004	0.009
PBIAS	0.00	2.29	0.00	1.84	0.27	0.67
RSR	0.71	0.95	0.30	0.68	0.51	0.62
Model performance	<b>unsatisfactory</b>	<b>unsatisfactory</b>	<b>very good</b>	<b>satisfactory</b>	<b>good</b>	<b>satisfactory</b>

As depicted in Figure 6.14 which shows the linear plot between the measured versus predicted P-value results obtained for all data sets with all models applied, a very wide deviation between the predicted and measured data was notable when the PCR model was applied. It was observed in Figure 6.14a that there wasn't an agreement between the predicted and measured values and hence, low quality predictions were obtained. When applying the ANN model a better dispersion around the diagonal line, compared to the one obtained when using the PCR model, was evident (see Figure 6.14c) and consequently, better predictions were achieved. Nevertheless, P-value results smaller than 1.10 and higher than 1.70 continued to be predicted very poorly, when the ANN model was applied. Better performance indexes for the training and validation data sets were obtained when the PLS model was used. As visualized in Figure 6.14b, when using the PLS model, a small dispersion around the diagonal line was achieved, compared to the ones obtained when applying the PCR and ANN models. It was notorious in Figure 6.14b, that better agreement between the predicted and measured results were achieved, especially for the training and validation data sets. However, for the test data set less accurate model was obtained and hence, the performance indexes obtained were not as good as the ones achieved for the training and validation data sets. Taking into account the given information and the results obtained, especially for the test data set, NMR combined with ANN model could be considered a possible technique to predict the P-value of fuel oil samples.



P-value is a property that gives information about the stability of fuel oils samples. Such stability is related with the composition of the fuel oil, more specifically, it is related with the quantity of asphaltenes. The exact structure of asphaltenes is not known but it is known that asphaltenes molecules are constituted by carbon and hydrogen atoms and it is rich in aromatic and aliphatic compounds. This information leads to conclude that NMR is a good technique to predict the P-value content of fuel oil samples, since NMR spectroscopy is a technique that gives information about the chemical composition. In consequence, if the P-value is related with the chemical composition, it should be possible to use NMR to predict the P-value content. The results obtained for the training and validation data sets especially when using the PLS and ANN models confirmed that it was possible to obtain good predictions with those approaches. The worst predictions obtained at the test data set may be justified by the number of samples used. The test data set was composed with 43 samples and maybe for this property more samples to increase the quality of the developed models were necessary.



**Figure 6.14:** Correlation between the experimental and predicted P-value results of the fuel oil samples obtained by: a) PCR, b) PLS and c) ANN models.

### 6.2.1.1. Fuel Oil - Final Remarks

$^1\text{H}$  NMR spectroscopy combined with different multivariate models (PCR, PLS and ANN) were used to predict the properties of fuel oil samples. At the same time, statistical performance indexes were applied to evaluate the performance of each model. As visualized in Table 6.15, the ANN model present better performance indexes in almost all properties analysed. For the sulfur content it was with the PLS model that better predictions were obtained. However, when using the ANN model to predict sulfur content very good performance indexes were also achieved. In some of the properties studied, PLS model also gave very good results. When applying the PCR model, worst predictions were obtained. To develop models with more accuracy more fuel oil samples should be analysed. Therefore, more experimental data will be required in an attempt to improve the final results.

Concluding, the final results presented in Table 6.15 demonstrated that  $^1\text{H}$  NMR spectroscopy is a good technique to obtain information about the fuel oil properties such as kinematic viscosity, sulfur content, carbon residue, density, P-value and flash point. The final models highlights that reasonable correlations were achieved when using  $^1\text{H}$  NMR spectroscopy, demonstrating that it can contribute to determine some properties with acceptable accuracy. The great advantage is that  $^1\text{H}$  NMR spectroscopy can give good predictions in just a few minutes which lead to control and avoid sample losses. Taking into account the obtained results, the possibility of acquiring an online NMR spectrometer for Matosinhos refinery is evaluated. It is expected that the use of an online NMR spectrometer can contribute in the process control and product quality evaluation, providing fuel oil samples with the desired properties and respecting the specifications. On the other hand, the online monitoring of such properties would also be very important to control the quantities of cutter stocks used in fuel oil production. An example of a simple economic evaluation of the implementation of an online NMR spectrometer at fuel oil stream is presented in Appendix A.

**Table 6.15:** Results obtained for different properties of fuel oil analysed when using <sup>1</sup>H NMR in multivariate models.

			number PC	number LV	total variance	R <sup>2</sup> /NSE	MAE	RMSE	MBE	PBIAS	RSR	model performance
<b>Sulfur content</b>	PCR	training and validation	13		97.4%	0.59	0.21	0.28	0.000	0.00	0.64	satisfactory
		test				0.47	0.23	0.32	-0.020	-2.74	0.73	unsatisfactory
	ANN	training and validation				0.96	0.06	0.09	-0.001	-0.05	0.20	very good
		test				0.91	0.10	0.14	0.020	2.46	0.30	very good
	PLS	training and validation		34		<b>0.99</b>	<b>0.02</b>	<b>0.03</b>	<b>0.000</b>	<b>0.00</b>	<b>0.06</b>	<b>very good</b>
		test			<b>0.92</b>	<b>0.06</b>	<b>0.12</b>	<b>0.003</b>	<b>0.34</b>	<b>0.28</b>	<b>very good</b>	
<b>Kinematic viscosity 100 °C</b>	PCR	training and validation	14		97.8%	0.57	55.19	104.17	0.00	0.00	0.60	satisfactory
		test				0.34	58.95	85.47	3.61	5.27	0.81	unsatisfactory
	ANN	training and validation				<b>0.96</b>	<b>16.15</b>	<b>31.95</b>	<b>0.52</b>	<b>0.69</b>	<b>0.20</b>	<b>very good</b>
		test				<b>0.94</b>	<b>16.50</b>	<b>26.49</b>	<b>2.69</b>	<b>3.93</b>	<b>0.25</b>	<b>very good</b>
	PLS	training and validation		35		0.99	12.12	15.68	0.00	0.00	0.10	very good
		test			0.51	51.56	74.06	7.55	11.00	0.70	satisfactory	
<b>Kinematic viscosity 50 °C</b>	PCR	training and validation	15		98.3%	0.71	58.70	79.78	0.00	0.00	0.54	good
		test				0.58	77.54	93.39	-2.72	-0.65	0.65	satisfactory
	ANN	training and validation				<b>0.92</b>	<b>28.68</b>	<b>42.47</b>	<b>1.08</b>	<b>0.26</b>	<b>0.29</b>	<b>very good</b>
		test				<b>0.74</b>	<b>57.54</b>	<b>73.61</b>	<b>-14.46</b>	<b>-3.48</b>	<b>0.51</b>	<b>good</b>
	PLS	training and validation		9		0.85	42.70	56.84	0.00	0.00	0.39	very good
		test			0.73	60.94	74.47	4.16	1.00	0.52	good	

**Table 6.15:** Results obtained for different properties of fuel oil analysed when using <sup>1</sup>H NMR in multivariate models (continuation).

			number PC	number LV	total variance	R <sup>2</sup> /NSE	MAE	RMSE	MBE	PBIAS	RSR	model performance
<b>Carbon residue</b>	PCR	training and validation	15		98.4%	0.65	0.85	1.09	0.000	0.00	0.59	good
		test				0.37	1.28	1.56	-0.490	-2.87	0.79	unsatisfactory
	ANN	training and validation				<b>0.77</b>	<b>0.54</b>	<b>0.88</b>	<b>-0.061</b>	<b>-0.36</b>	<b>0.48</b>	<b>very good</b>
		test				<b>0.70</b>	<b>0.94</b>	<b>1.07</b>	<b>0.260</b>	<b>1.50</b>	<b>0.55</b>	<b>good</b>
	PLS	training and validation				0.75	0.74	0.93	0.000	0.00	0.50	very good
		test				0.31	1.30	1.63	-0.490	-2.85	0.83	unsatisfactory
<b>Flash point</b>	PCR	training and validation	15		98.3%	0.44	5.97	7.90	0.00	0.00	0.75	unsatisfactory
		test				0.60	4.93	6.59	0.57	0.77	0.63	satisfactory
	ANN	training and validation				<b>0.67</b>	<b>4.66</b>	<b>6.06</b>	<b>0.31</b>	<b>0.42</b>	<b>0.58</b>	<b>good</b>
		test				<b>0.67</b>	<b>4.50</b>	<b>5.95</b>	<b>-0.22</b>	<b>-0.30</b>	<b>0.57</b>	<b>good</b>
	PLS	training and validation				0.79	3.86	4.86	0.00	0.00	0.46	very good
		test				0.64	4.95	6.26	0.35	0.48	0.60	satisfactory
<b>P-value</b>	PCR	training and validation	14		97.7	0.50	0.07	0.11	0.000	0.00	0.71	unsatisfactory
		test				0.10	0.11	0.14	0.030	2.29	0.95	unsatisfactory
	ANN	training and validation				<b>0.74</b>	<b>0.06</b>	<b>0.08</b>	<b>0.004</b>	<b>0.27</b>	<b>0.51</b>	<b>good</b>
		test				<b>0.62</b>	<b>0.07</b>	<b>0.09</b>	<b>0.009</b>	<b>0.67</b>	<b>0.62</b>	<b>satisfactory</b>
	PLS	training and validation				0.91	0.03	0.04	0.000	0.00	0.30	very good
		test				0.53	0.08	0.10	0.020	1.84	0.68	satisfactory
<b>Density</b>	PCR	training and validation	15		98.1%	0.77	0.003	0.004	0.000	0.00	0.48	very good
		test				0.75	0.003	0.004	-0.001	-0.12	0.50	very good
	ANN	training and validation				<b>0.85</b>	<b>0.002</b>	<b>0.004</b>	<b>0.000</b>	<b>0.03</b>	<b>0.38</b>	<b>very good</b>
		test				<b>0.86</b>	<b>0.002</b>	<b>0.003</b>	<b>0.000</b>	<b>0.02</b>	<b>0.37</b>	<b>very good</b>
	PLS	training and validation				0.79	0.003	0.004	0.000	0.00	0.45	very good
		test				0.77	0.003	0.004	-0.001	-0.11	0.48	very good

### 6.2.2. LVGO

$^1\text{H}$  NMR spectroscopy was used to determine some LVGO properties such as density (ASTM Standards D4052, 2011) and the distillation curves at atmospheric pressure (ISO Standards 3405, 2011). The main objective of analysing this stream was to evaluate its fractionation and to ensure that a good separation into the vacuum column was obtained. It was for this reason that these two properties were considered the most important to analyse.

As occurred in the analysis of the fuel oil samples, PCR, PLS and ANN were the models used to predict the LVGO properties. PCR and ANN models were based on the principal components while PLS model was based on the original data. The statistical performance indexes were also applied to evaluate the performance of the developed models. The spectral area was divided in constant segments of 0.02 ppm. It was with this bin that better predictions were achieved and consequently a bin width of 0.02 ppm was always used for the analysis of LVGO properties. The procedure used to determine the best network architecture when applying the ANN model was the same as presented for the fuel oil samples. The best architecture was achieved with different number of neurons in the input layer depending on the number of PCs used, three neurons in the hidden layer and one neuron in the output layer. When using the PLS model, the original variables were used as inputs and the choice of the number of latent variables was made taking into account the minima of the RMSECV (root mean squared error of cross validation) plot. As referred in Table 6.4 (page 187), the study was performed using 105 LVGO samples for each property analysed. Thereby, the training data set included 63 LVGO samples whereas the validation and test data sets were constituted by 21 LVGO samples, respectively. All three different groups were representative of the data using the procedure described in the Section 5.3.

#### vi) Density

Table 6.16 presents the performance indexes achieved when applying the different multivariate models for the data sets, training and validation data sets and test data set, when predicting the density of LVGO samples. The ANN and PCR models were based on the PCs, which were determined by principal component analysis. Fifteen PCs were

selected, explaining 98.1% of the original data variance. The PLS model was based on the original data and 13 latent variables to predict the best performance indexes were used.

The results obtained (see Table 6.16) demonstrated that the best performance indexes were achieved with the ANN model for both data sets. With the ANN model, high coefficients of determination or high NSE values were obtained and very RMSE, MAE and RSR low values were achieved. These results were indication of very good predictions for both data sets. Taking into account the PBIAS and MBE values, for the training and validation data sets the predicted results were overestimated, while for the test data set the predicted values were underestimated. Less accurate predictions were obtained when the PLS model was applied, although still of high quality. Comparing the PLS and the ANN models, the main emphasized differences were the smaller NSE values and higher RSR results obtained with the PLS model. When applying the PCR model, very good performance indexes were achieved for the test data set, whereas for the validation and training data sets only satisfactory predictions were obtained.

**Table 6.16:** Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the density of LVGO samples.

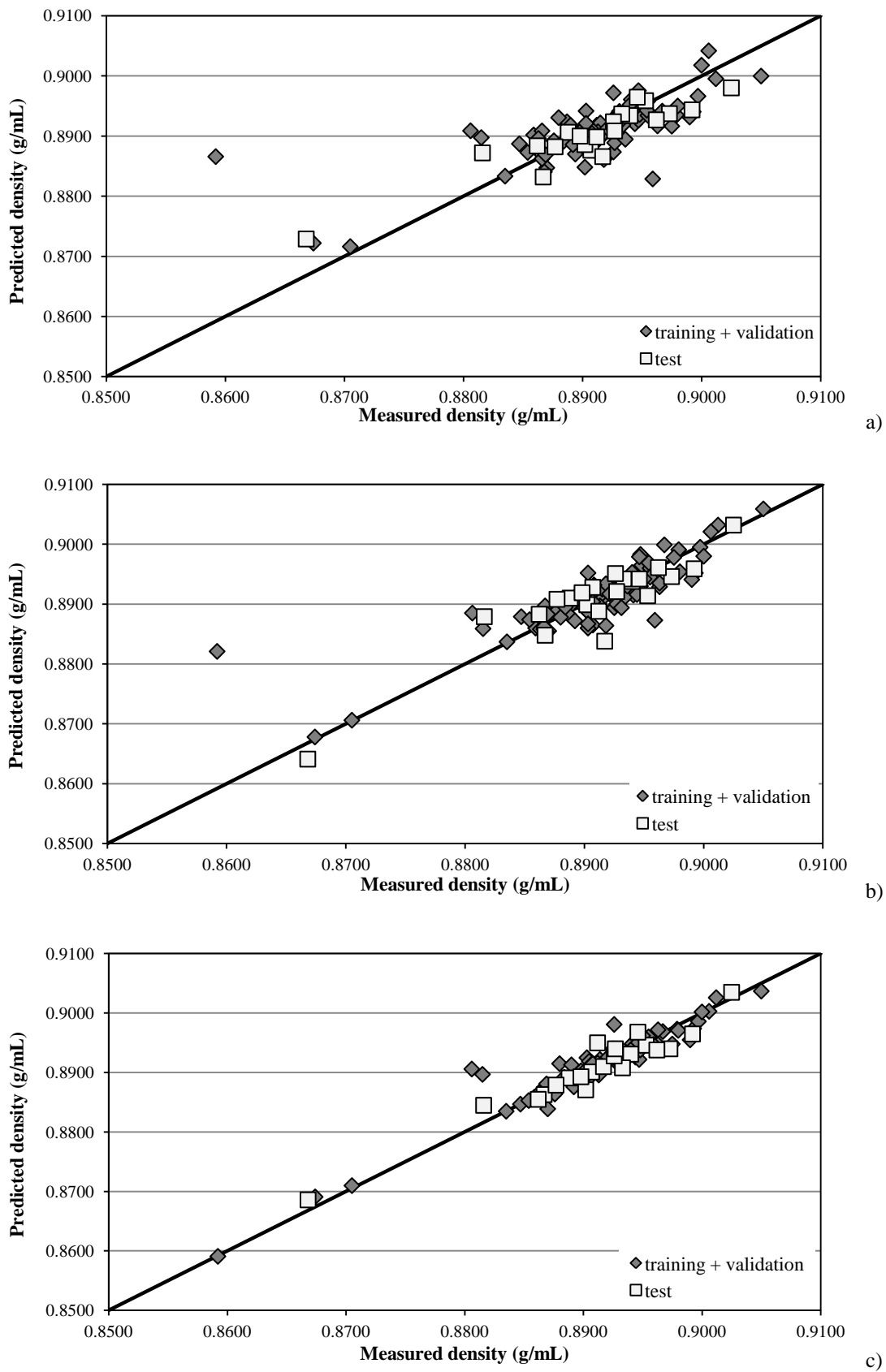
	PCR		PLS		ANN	
	training and validation	test	training and validation	test	training and validation	test
R <sup>2</sup> /NSE	0.54	0.80	0.72	0.82	0.92	0.93
MAE	0.003	0.003	0.002	0.002	0.001	0.002
RMSE	0.004	0.003	0.003	0.003	0.002	0.002
MBE	0.0000	-0.0007	0.0000	-0.0002	0.0002	-0.0002
PBIAS	0.00	-0.08	0.00	-0.02	0.02	-0.03
RSR	0.68	0.44	0.53	0.42	0.29	0.27
Model performance	<b>satisfactory</b>	<b>very good</b>	<b>good</b>	<b>very good</b>	<b>very good</b>	<b>very good</b>

Comparing the measured and predicted density results obtained for the training, validation and test data sets when using all the multivariate models (Figure 6.15) it was possible to conclude that a high quality model was achieved when the ANN model was applied. As visualized in Figure 6.15c the agreement between the predicted and measured results was remarkable, showing that the developed model was of high quality. The correlation

between the predicted and measured values obtained when applying the PLS model gave slightly less accurate performance than the former although still of high quality, as visualized in Figure 6.15b. When using the PCR model, higher dispersion around the diagonal line was notorious (see Figure 6.15a). As illustrated in Figure 6.15a, the PCR model had more difficulty to predict the lower density results of the LVGO samples. Consequently, low quality predictions were obtained especially for the training and validation data sets.

In general, the results obtained indicate that the use of  $^1\text{H}$  NMR spectroscopy and ANN models is a good alternative to determine the density of LVGO samples.





**Figure 6.15:** Correlation between the experimental and predicted density results of the LVGO samples obtained by: a) PCR, b) PLS and c) ANN models.

## **Distillation**

The distillation curve of LVGO samples is a very important source of information which needs to be controlled to confirm that a good separation between LVGO and HVGO, in the vacuum column, is achieved. The distillation curve of LVGO was determined by defining the initial boiling point (IBP) and endpoint (EP). The cuts of the distillation curve that were analysed were 5, 10, 50, 90 and 95% (v/v) distillation cuts. These cuts were considered the most important to evaluate how the separation of LVGO occurred.

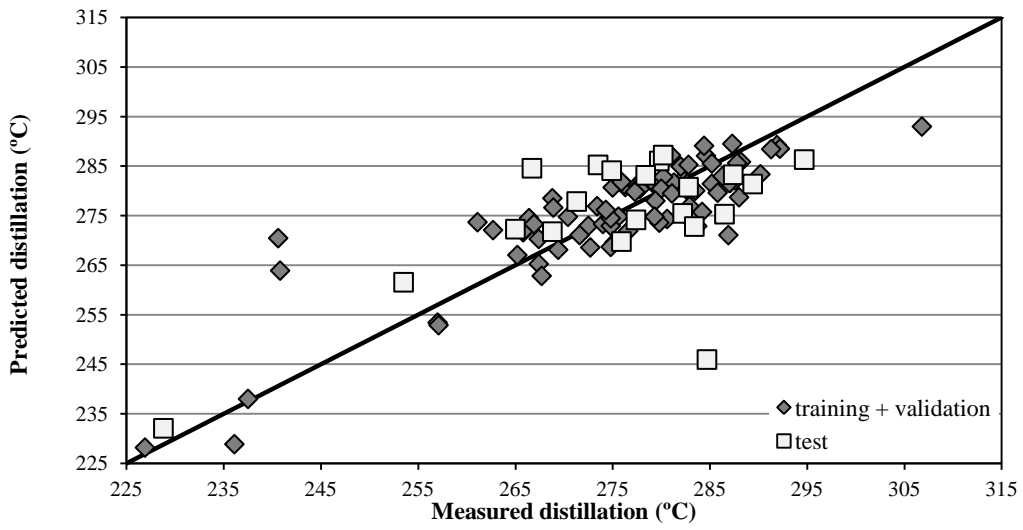
### **vi) 5% Distillation**

The results obtained when 5% distillation of LVGO samples were predicted (see Table 6.17) demonstrated that performances were particularly very good, for both data sets, when the ANN model was applied. Before applying the ANN model, 16 PCs were used as inputs of this model, explaining 98.5% of the total variance. The results obtained (Table 6.17) demonstrated that the NSE values or coefficients of determination varied from 0.87 to 0.79 for the training and validation data sets and test data set, respectively. The performance indexes determined, such as, RMSE, MAE, and RSR were smaller for both data sets compared with the results obtained for the PCR and PLS models. These coefficients were the indicative for a very good model performance. As indicated by PBIAS and MBE the results obtained for the test data set were underestimated while the predicted results obtained for the training and validation data sets were overestimated. When applying the PCR and PLS models the performance indexes were not as good as the performances obtained with the ANN model. However good correlations were achieved especially for the training and validation data sets. Models with low quality were obtained when predicting the test data set. Unsatisfactory predictions were achieved for the test data set of the PCR and PLS models. For this reason, PCR and PLS models were not considered a good alternative to predict the 5% distillation results of a distillation curve.

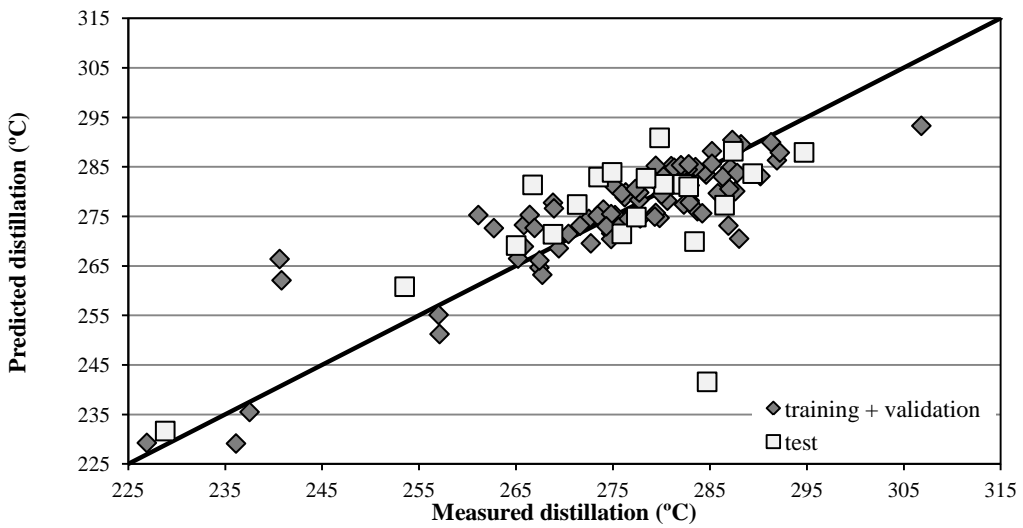
**Table 6.17:** Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the 5% distillation of LVGO samples.

	PCR		PLS		ANN	
	training and validation	test	training and validation	test	training and validation	test
R <sup>2</sup> /NSE	0.74	0.30	0.76	0.29	0.87	0.79
MAE	4.77	8.77	4.51	7.68	3.24	5.39
RMSE	6.64	11.59	6.35	11.72	4.65	6.41
MBE	0.00	-0.70	0.00	-0.68	0.53	-0.17
PBIAS	0.00	-0.25	0.00	-0.25	0.19	-0.06
RSR	0.51	0.84	0.49	0.84	0.36	0.46
Model performance	<b>good</b>	<b>unsatisfactory</b>	<b>very good</b>	<b>unsatisfactory</b>	<b>very good</b>	<b>very good</b>

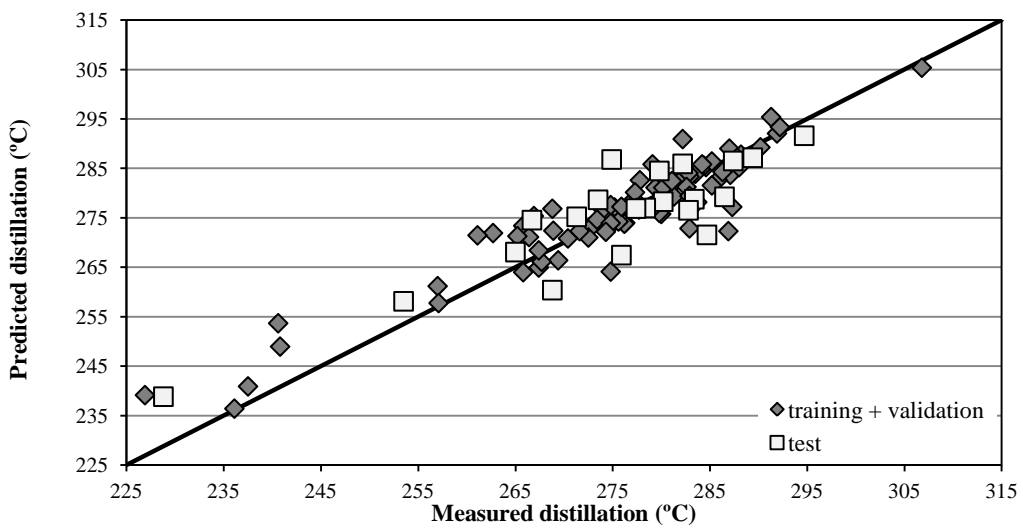
The linear fits of calibration for all the multivariate models used are illustrated in Figure 6.16. As visualized in Figure 6.16c, it was with the ANN model that a better agreement between the predicted and measured 5% distillation results was obtained. Consequently, the ANN model predicted the 5% distillation of LVGO samples with the higher accuracy. However, some dispersion between the predicted and measured results was notable and some difficult in predicting the 5% distillation temperature of samples with a distillation temperature smaller than 255 °C was found. Samples with higher distillation temperature were easily predicted. Comparing the agreement between the predicted and measured results for each data set, training and validation data sets and test data set, when using the ANN model, it was visualized that the values of the test data set were not as well predicted as the ones from the training and validation data sets. Consequently, low quality model was obtained for the test data set, although still with very good performance. Models with high quality for both data sets were not achieved when applying the PCR and PLS models. Figure 6.16a and Figure 6.16b clearly demonstrated that there was some dispersion around the diagonal line and hence, the agreement between the predicted and measured results was not remarkable, especially for the test data set. It was for this reason that the performance indexes obtained for the test data set when using the PCR and PLS models were of low quality. Concluding, the ANN model is more reliable to predict the 5% distillation temperatures of LVGO samples.



a)



b)



c)

**Figure 6.16:** Correlation between the experimental and predicted 5% distillation of the LVGO samples obtained by: a) PCR, b) PLS and c) ANN models.

**vi) 10% Distillation**

Table 6.18 demonstrated that the prediction of 10% distillation of LVGO samples through the use of  $^1\text{H}$  NMR spectroscopy could be made when applying all multivariate models. Very good performance indexes were achieved for both data sets when PLS and ANN models were applied. When using the PCR model, good predictions were obtained for the training and validation data sets, whereas for the test data set very good predictions were achieved. PCR and ANN models used as inputs 15 PCs, explaining 98.3% of the original data variance. PLS used the original data as inputs and it required 21 LVs to perform good predictions.

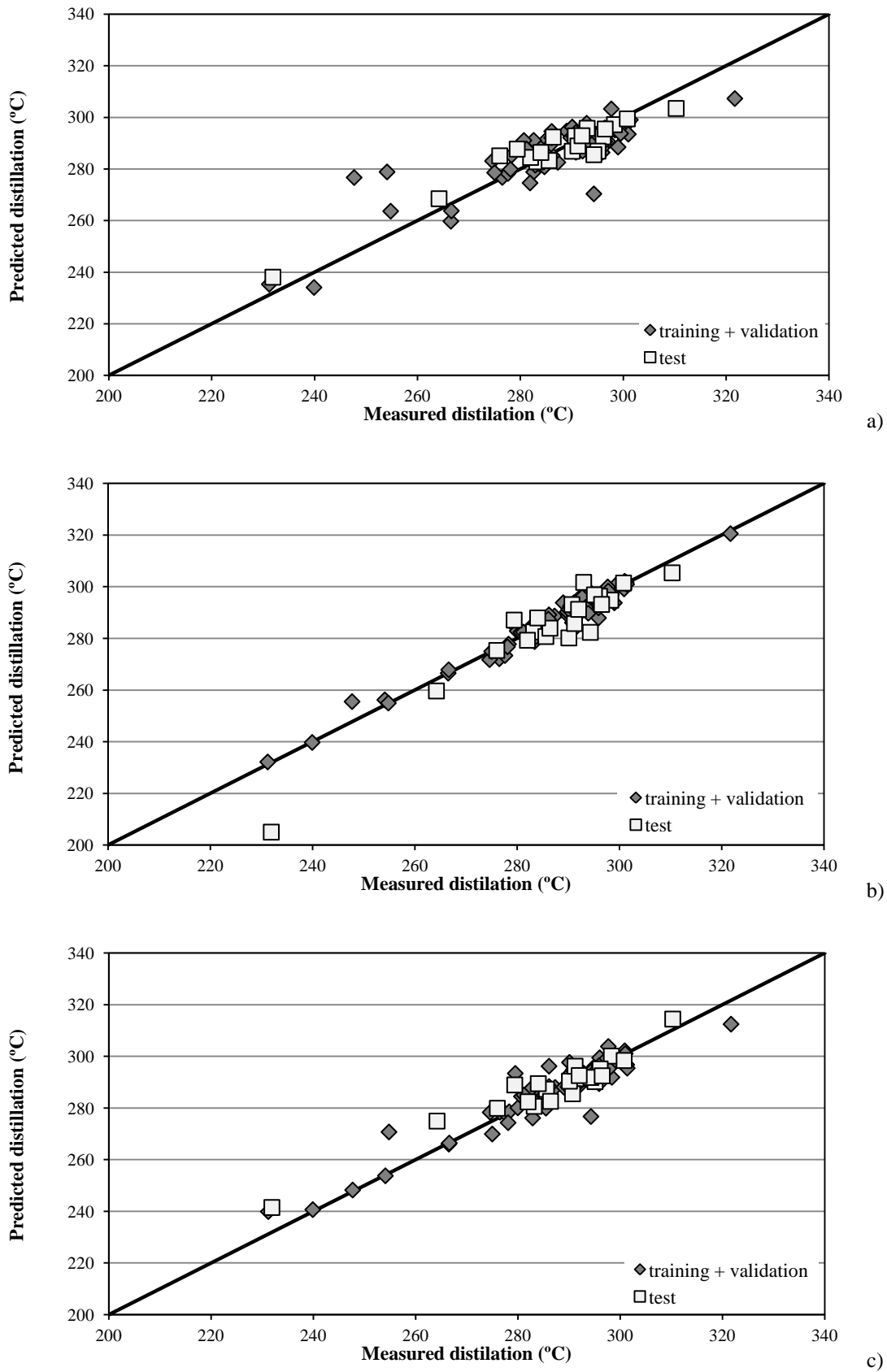
**Table 6.18:** Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the 10% distillation of LVGO samples.

	PCR		PLS		ANN	
	training and validation	test	training and validation	test	training and validation	test
R <sup>2</sup> /NSE	0.73	0.90	0.96	0.75	0.89	0.90
MAE	4.99	3.98	1.99	5.14	2.86	3.86
RMSE	7.02	4.86	2.65	7.75	4.42	4.86
MBE	0.00	0.24	0.00	-2.75	-0.11	1.24
PBIAS	0.00	0.08	0.00	-0.96	-0.04	0.43
RSR	0.52	0.31	0.20	0.50	0.33	0.31
Model performance	<b>good</b>	<b>very good</b>	<b>very good</b>	<b>very good</b>	<b>very good</b>	<b>very good</b>

When comparing the predicted results obtained for the training and validation data sets using all different multivariate models (Table 6.18), it was confirmed that good performance indexes were achieved for all models. However, it was with the PLS model that better predictions were obtained. The predicted results were characterized for presenting high NSE values and the smallest errors when compared with PCR and ANN predictions. As indicated by PBIAS and MBE, accurate model predictions were obtained. The performance indexes obtained when using PLS model were indicative of very good performance indexes.

As illustrated in Table 6.18, it was with the PCR and ANN models that better predictions for the test data set were achieved. The performance indexes obtained with PCR and ANN were quite similar, with the predicted results characterized for having high coefficients of determination or NSE values and small predicted errors. The main difference between these models was the MBE and PBIAS results. MBE and PBIAS indicated that the predicted results were overestimated for the two models, however it was with the PCR model that predicted results were more accurate.

Taking into account the predicted and measured results obtained with all models for both training and validation data sets and test data set, it was considered that the ANN model was better to predict the 10% distillation of LVGO samples, as visualized in Figure 6.17c. The ANN model was considered the best model taking into account the performance indexes obtained for both data sets and especially for the test data set. It was with the test data set that the response of the model with new situations was evaluated. As illustrated in Figure 6.17, the agreement between the predicted and measured values, when applying the PLS and ANN models, was remarkable. When applying the PCR model, Figure 6.17a, larger deviations, compared to the ones achieved with the other two models, were found especially for the training and validation data sets. In general, the ANN model predicted the 10% distillation temperatures with higher quality and accuracy and consequently was considered the most reliable for such prediction.



**Figure 6.17:** Correlation between the experimental and predicted 10% distillation of the LVGO samples obtained by: a) PCR, b) PLS and c) ANN models.

**vi) 50% Distillation**

The prediction of 50% distillation cut of LVGO samples was performed using 16 PCs as inputs of the PCR and ANN models. These PCs were responsible for 98.5% of the total variance. Considering the PLS model, 21 LVs were necessary to performed good predictions.

Analysing the results presented in Table 6.19 it was concluded that when predicting the 50% distillation cut, independently of the multivariate models used, very good predictions were obtained for both data sets. High coefficients of determination or NSE values and low model errors characterize all predicted results obtained for both data sets.

**Table 6.19:** Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the 50% distillation of LVGO samples.

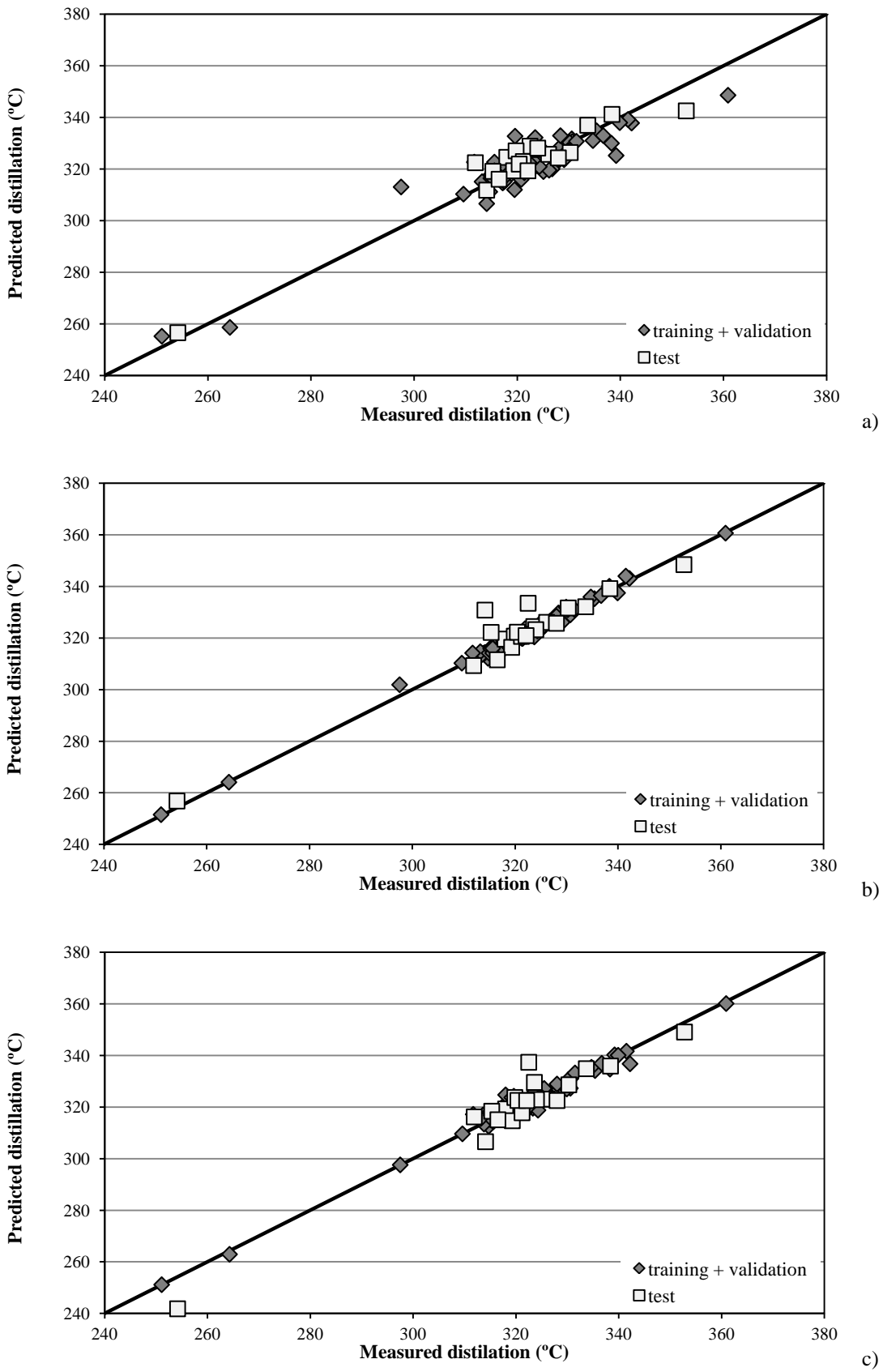
	<b>PCR</b>		<b>PLS</b>		<b>ANN</b>	
	training and validation	test	training and validation	test	training and validation	test
R <sup>2</sup> /NSE	0.87	0.92	0.99	0.92	0.97	0.90
MAE	3.56	3.78	0.91	3.14	1.63	4.04
RMSE	4.78	4.74	1.26	5.04	2.19	5.42
MBE	0.00	1.50	0.00	1.14	-0.03	-0.43
PBIAS	0.00	0.47	0.00	0.36	-0.01	-0.13
RSR	0.36	0.27	0.10	0.29	0.17	0.31
Model performance	<b>very good</b>	<b>very good</b>	<b>very good</b>	<b>very good</b>	<b>very good</b>	<b>very good</b>

When comparing the predicted results obtained for the training and validation data sets using all different multivariate models (Table 6.19) it was possible to conclude that it was with the PLS model that better performance indexes were achieved. However, as already stated and as demonstrated in Table 6.19, very good predictions were also obtained with ANN and PCR models. The predicted results obtained with the PLS model were characterized for displaying a very high NSE values, and the smallest RMSE, MAE and RSR results when comparing with the other models. In addition, PBIAS and MBE results indicated that accurate model predictions were achieved with the PLS model.



The predicted results obtained for the test data set, when using the different multivariate models, were as good as the results obtained for the training and validation data sets. As visualized in Table 6.19, better predictions were obtained when PLS and PCR were applied. The performance indexes calculated for the PLS and PCR models were quite similar, although some differences at the PBIAS and MBE values were found. The PBIAS and MBE results were positive for both models indicating a slight overprediction. However, when using the PLS model the results were similar with the measure data, as illustrated in Figure 6.18b. Figure 6.18 shows the measured versus predicted values obtained for both data sets when applying all the multivariate models, illustrating the range of the dispersion. As illustrated, independently of the model used, the agreement between the predicted and measured values was remarkable, showing that the models were of high quality and accurate models were achieved. Some dispersion around the diagonal line was notable when the PCR model was used (see Figure 6.18a), although this model is still of high quality.

Considering the results obtained to predict the 50% distillation of LVGO samples through the use of  $^1\text{H}$  NMR spectroscopy, better performance indexes were achieved when the PLS model with 21 LVs was used. PLS model was more reliable to predict the 50% distillation of LVGO samples for both data sets. This meant that samples with smaller or higher temperature, especially samples of the training and validation data sets, were very well predicted. Only a few samples of the test data set were slight deviated from the diagonal line. This could be the reason why the predicted results at the test data set presented slightly smaller coefficient of determination when compared with the one obtained from the training and validation data sets.



**Figure 6.18:** Correlation between the experimental and predicted 50% distillation of the LVGO samples obtained by: a) PCR, b) PLS and c) ANN models.

### vi) 90% Distillation

Table 6.20 presents the values of the performance indexes using PCR, PLS and ANN models for both, training and validation data sets and test data set. Before applying the PCR and ANN models, PCA was used to determine the number of PCs that were used as inputs of these models. With PCA, 15 PCs responsible for explaining 98.1% of the total variance were selected. To develop the PLS model, 13 LVs were used.

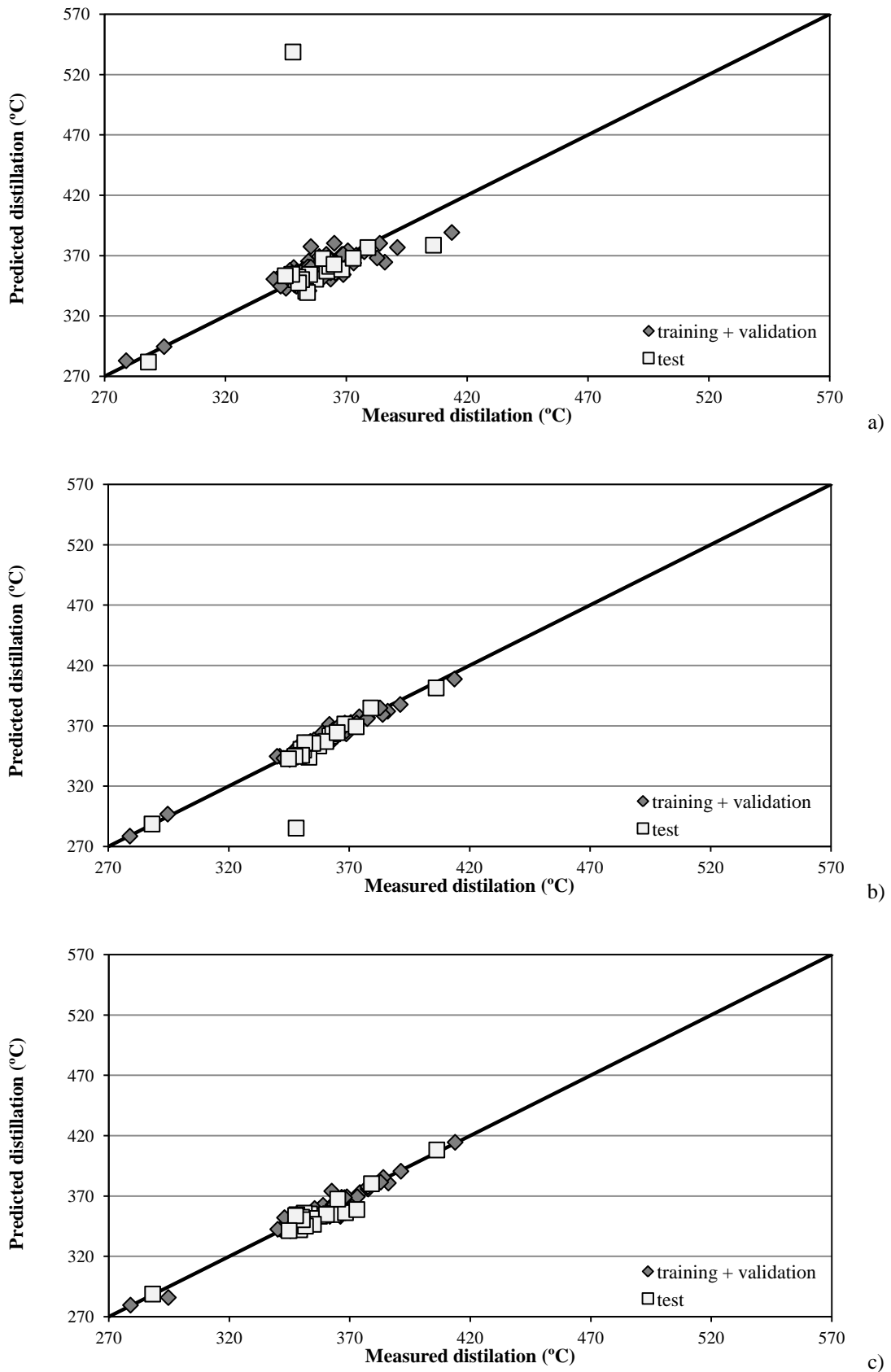
Better performance indexes were achieved with the ANN model for both data sets (see Table 6.20). As illustrated in Table 6.20, it was with the ANN model that higher coefficients of determination or NSE results were obtained. In addition, small RMSE, MAE and RSR values were also achieved. The values of MBE and PBIAS were negative, indicating a slight underprediction. It was with the ANN that models of high quality were achieved for both data sets, training and validation data sets and test data set. When applying the PCR and PLS models, to predict the training and validation data sets, very good predictions were obtained, although both models failed when predicting the test data set. Hence, models with low quality were achieved when using the PCR and PLS models to predict the test data set.

**Table 6.20:** Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the 90% distillation of LVGO samples.

	PCR		PLS		ANN	
	training and validation	test	training and validation	test	training and validation	test
R <sup>2</sup> /NSE	0.77	-3.36	0.97	0.52	0.94	0.91
MAE	6.00	15.12	2.28	5.73	2.68	5.09
RMSE	7.79	42.50	2.86	14.15	3.87	6.21
MBE	0.00	5.61	0.00	-4.12	-0.29	-2.38
PBIAS	0.00	1.57	0.00	-1.16	-0.08	-0.67
RSR	0.48	2.09	0.18	0.70	0.24	0.31
Model performance	<b>very good</b>	<b>unsatisfactory</b>	<b>very good</b>	<b>satisfactory</b>	<b>very good</b>	<b>very good</b>

To confirm the conclusions previously made, Figure 6.19 depicts the measured versus predicted 90% distillation results with all models (PCR, PLS and ANN models) and for

both data sets, training and validation data sets and test data set. As illustrated in Figure 6.19c, the ANN model was most reliable to predict the 90% distillation results of LVGO samples, since a very good agreement between the predicted and measured results existed. A remarkable agreement between the predicted and measured results was also obtained when using the PLS model to predict the training and validation data sets. Although, the model failed when predicting the test data set. As visualized in Figure 6.19b, the larger deviation of some predicted values was one of the factors that contributed for such low quality. The higher dispersion of the predicted results was obtained when the PCR model was applied, as illustrated in Figure 6.19a. Consequently, less accurate models were obtained and even low quality predictions, for the test data set, were achieved. For these reasons, the ANN model was considered the best model to predict the 90% distillation results of LVGO samples.



**Figure 6.19:** Correlation between the experimental and predicted 90% distillation of the LVGO samples obtained by: a) PCR, b) PLS and c) ANN models.

**vi) 95% Distillation**

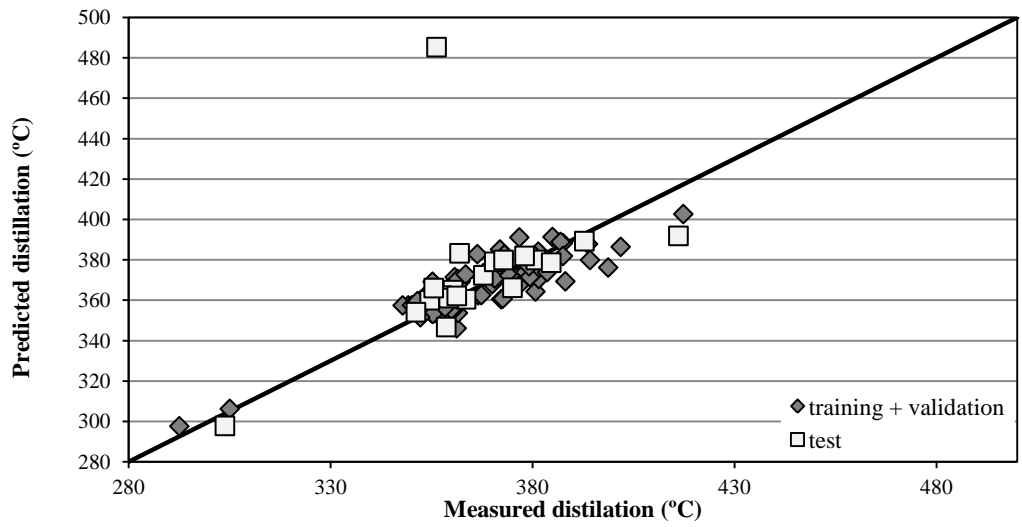
In most data sets, prediction of 95% distillation of LVGO samples through the use of  $^1\text{H}$  NMR spectroscopy when applying the different multivariate models, resulted in very good models performance. With the exception of the predicted results obtained for the test data set, when applying the PCR model, that unsatisfactory results were achieved (Table 6.21). For the PCR and ANN models 16 PCs corresponding to 98.4% of the total variance were used as inputs.

**Table 6.21:** Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the 95% distillation of LVGO samples.

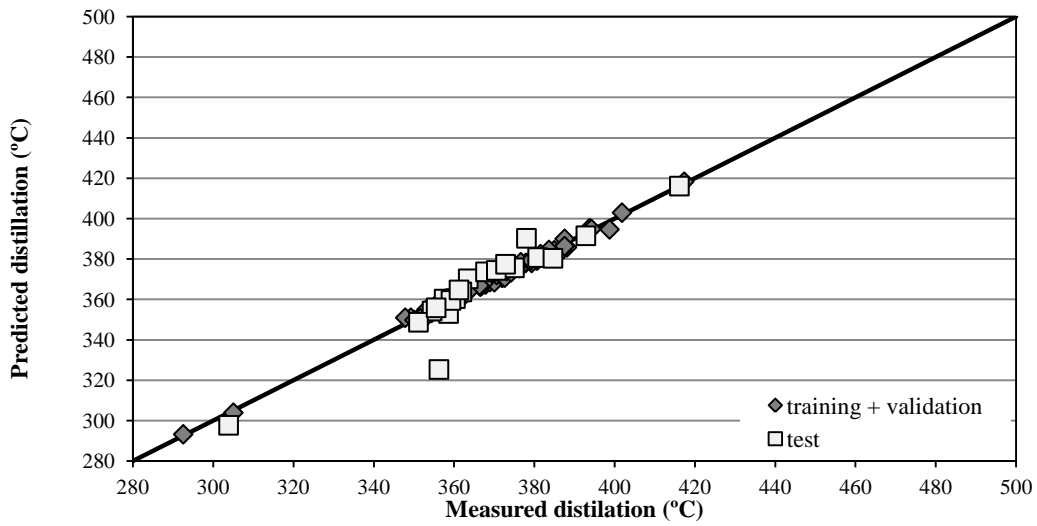
	PCR		PLS		ANN	
	training and validation	test	training and validation	test	training and validation	test
R <sup>2</sup> /NSE	0.77	-1.10	0.99	0.85	0.94	0.81
MAE	6.62	12.86	1.13	4.45	2.43	7.35
RMSE	8.12	29.58	1.38	7.98	4.12	8.96
MBE	0.00	6.71	0.00	-0.38	-0.82	-0.22
PBIAS	0.00	1.83	0.00	-0.11	-0.22	-0.06
RSR	0.48	1.45	0.08	0.39	0.24	0.44
Model performance	<b>very good</b>	<b>unsatisfactory</b>	<b>very good</b>	<b>very good</b>	<b>very good</b>	<b>very good</b>

Analysing the linear plot of calibration between the measured and predicted results for all models applied and for both training and validation data sets and test data set (Figure 6.20) it was possible to conclude that PLS model was the more reliable to predict the 95% distillation values (Figure 6.20b). The PLS model developed with 22 LVs was characterized for having the highest coefficient of determination or NSE values and the smallest RMSE, MAE and RSR results, when comparing with the predictions obtained with the other models for both data sets. In addition, as indicated by the PBIAS and MBE results an accurate model was achieved when using the PLS model to predict the training and validation data sets, while a slight underprediction, for the test data set, was found. As illustrated in Figure 6.20c, a good agreement between the predicted and measured values was also found when the ANN model was applied. A slight deviation of the predicted results, from the test data set, was notorious (Figure 6.20c), compared with the PLS model,

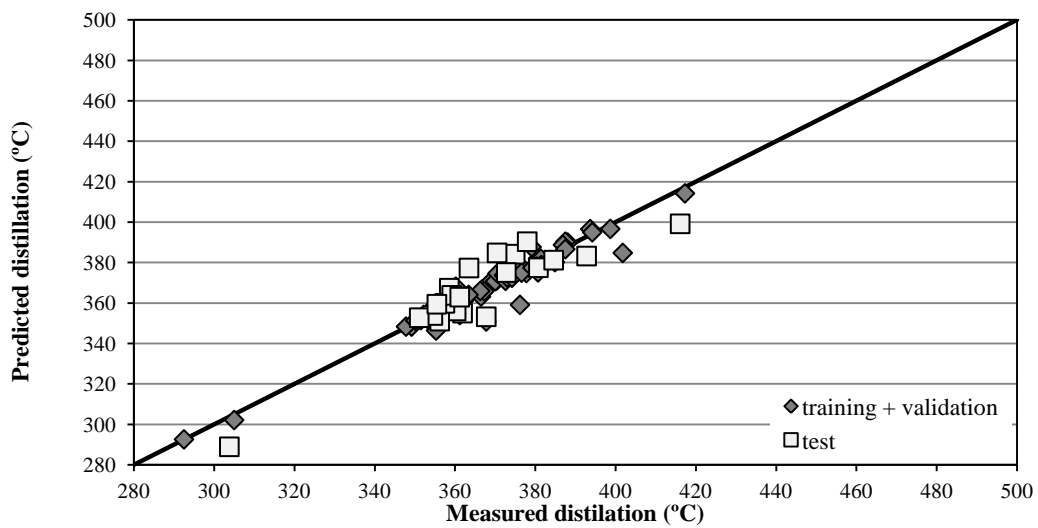
although the performance indexes obtained were still of high quality. Figure 6.20a shows a larger deviation of the predicted results, especially for the test data set, when the PCR model was applied. Consequently, a low quality model for the test data set was found. For the training and validation data sets, the agreement between the predicted and measured results was not so evident and a slightly less accurate model was obtained, although still of high quality and characterized as a very good model. Concluding, the PLS model could be considered the most reliable to determine the 95% distillation of LVGO samples.



a)



b)



c)

**Figure 6.20:** Correlation between the experimental and predicted 95% distillation of the LVGO samples obtained by: a) PCR, b) PLS and c) ANN models.



### 6.2.1.2. LVGO – Final Remarks

The prediction of LVGO properties were made through the use of  $^1\text{H}$  NMR spectroscopy combined with different multivariate methods (PCR, PLS and ANN). The performance developed by each model was evaluated by applying some statistical performance indexes. Table 6.22 presents the values of performance indexes calculated for all developed models for both, training and validation data sets and test data set. As visualized in Table 6.22, it was possible to conclude that, in almost all properties analysed, it was with the ANN model that better predictions were obtained. Even, for the properties where better predictions were achieved with the PLS model, very good predictions were also obtained with the ANN model. PCR model was not considered such a good model to predict LVGO properties as ANN and PLS models. However, when predicting some properties, PCR also gave good predictions. As expected, the ANN model performed better predictions due to its ability to develop models around non-linear relationship between the measured properties and the predicted results.

Concluding,  $^1\text{H}$  NMR spectroscopy is a good technique to obtain information about the LVGO properties such as density and the distillation at atmospheric pressure.

**Table 6.22:** Results obtained for different properties of LVGO analysed when using <sup>1</sup>H NMR in multivariate models.

			number PC	number LV	total variance	R <sup>2</sup> /NSE	MAE	RMSE	MBE	PBIAS	RSR	model performance
<b>Density</b>	PCR	training and validation	15		98.1%	0.54	0.003	0.004	0.0000	0.00	0.68	satisfactory
		test				0.80	0.003	0.003	-0.0007	-0.08	0.44	very good
	ANN	training and validation				<b>0.92</b>	<b>0.001</b>	<b>0.002</b>	<b>0.0002</b>	<b>0.02</b>	<b>0.29</b>	<b>very good</b>
		test				<b>0.93</b>	<b>0.002</b>	<b>0.002</b>	<b>-0.0002</b>	<b>-0.03</b>	<b>0.27</b>	<b>very good</b>
	PLS	training and validation				0.72	0.002	0.003	0.0000	0.00	0.53	good
		test				0.82	0.002	0.003	-0.0002	-0.02	0.42	very good
<b>Distillation (5%)</b>	PCR	training and validation	16		98.5%	0.74	4.77	6.64	0.00	0.00	0.51	good
		test				0.30	8.77	11.59	-0.70	-0.25	0.84	unsatisfactory
	ANN	training and validation				<b>0.87</b>	<b>3.24</b>	<b>4.65</b>	<b>0.53</b>	<b>0.19</b>	<b>0.36</b>	<b>very good</b>
		test				<b>0.79</b>	<b>5.39</b>	<b>6.41</b>	<b>-0.17</b>	<b>-0.06</b>	<b>0.46</b>	<b>very good</b>
	PLS	training and validation				0.76	4.51	6.35	0.00	0.00	0.49	very good
		test				0.29	7.68	11.72	-0.68	-0.25	0.84	unsatisfactory
<b>Distillation (10%)</b>	PCR	training and validation	15		98.3%	0.73	4.99	7.02	0.00	0.00	0.52	good
		test				0.90	3.98	4.86	0.24	0.08	0.31	very good
	ANN	training and validation				<b>0.89</b>	<b>2.86</b>	<b>4.42</b>	<b>-0.11</b>	<b>-0.04</b>	<b>0.33</b>	<b>very good</b>
		test				<b>0.90</b>	<b>3.86</b>	<b>4.86</b>	<b>1.24</b>	<b>0.43</b>	<b>0.31</b>	<b>very good</b>
	PLS	training and validation				0.96	1.99	2.65	0.00	0.00	0.20	very good
		test				0.75	5.14	7.75	-2.75	-0.96	0.50	very good

**Table 6.22:** Results obtained for different properties of LVGO analysed when using <sup>1</sup>H NMR in multivariate models (continuation).

			number PC	number LV	total variance	R <sup>2</sup> /NSE	MAE	RMSE	MBE	PBIAS	RSR	model performance
<b>Distillation (50%)</b>	PCR	training and validation	16		98.5%	0.87	3.56	4.78	0.00	0.00	0.36	very good
		test				0.92	3.78	4.74	1.50	0.47	0.27	very good
	ANN	training and validation				0.97	1.63	2.19	-0.03	-0.01	0.17	very good
		test				0.90	4.04	5.42	-0.43	-0.13	0.31	very good
	PLS	training and validation				<b>0.99</b>	<b>0.91</b>	<b>1.26</b>	<b>0.00</b>	<b>0.00</b>	<b>0.10</b>	<b>very good</b>
		test				<b>0.92</b>	<b>3.14</b>	<b>5.04</b>	<b>1.14</b>	<b>0.36</b>	<b>0.29</b>	<b>very good</b>
<b>Distillation (90%)</b>	PCR	training and validation	15		98.1%	0.77	6.00	7.79	0.00	0.00	0.48	very good
		test				-3.36	15.12	42.50	5.61	1.57	2.09	unsatisfactory
	ANN	training and validation				<b>0.94</b>	<b>2.68</b>	<b>3.87</b>	<b>-0.29</b>	<b>-0.08</b>	<b>0.24</b>	<b>very good</b>
		test				<b>0.91</b>	<b>5.09</b>	<b>6.21</b>	<b>-2.38</b>	<b>-0.67</b>	<b>0.31</b>	<b>very good</b>
	PLS	training and validation				0.97	2.28	2.86	0.00	0.00	0.18	very good
		test				0.52	5.73	14.15	-4.12	-1.16	0.70	satisfactory
<b>Distillation (95%)</b>	PCR	training and validation	16		98.4%	0.77	6.62	8.12	0.00	0.00	0.48	very good
		test				-1.10	12.86	29.58	6.71	1.83	1.45	unsatisfactory
	ANN	training and validation				0.94	2.43	4.12	-0.82	-0.22	0.24	very good
		test				0.81	7.35	8.96	-0.22	-0.06	0.44	very good
	PLS	training and validation				<b>0.99</b>	<b>1.13</b>	<b>1.38</b>	<b>0.00</b>	<b>0.00</b>	<b>0.08</b>	<b>very good</b>
		test				<b>0.85</b>	<b>4.45</b>	<b>7.98</b>	<b>-0.38</b>	<b>-0.11</b>	<b>0.39</b>	<b>very good</b>

### 6.2.3. HVGO

The purpose of analysing the HVGO stream is to evaluate its fractionating and to ensure that a good separation into the vacuum column between LVGO, MVGO, HVGO and residue is obtained. One of the most important properties used to evaluate the HVGO samples is the carbon residue. With this property the quantity of carbon deposits found in HVGO samples are controlled. As the determination of the carbon residue using the ASTM standard method is time consuming, the use of  $^1\text{H}$  NMR spectroscopy became a possible alternative to evaluate the HVGO samples. Among carbon residue, kinematic viscosity, distillation and density were also considered important parameters to be investigated in order to evaluate the HVGO stream.

The HVGO samples were analysed by  $^1\text{H}$  NMR spectroscopy and to exploit the full information content of the NMR spectra various multivariate data methods have been used. Principal component regression, partial least squares and artificial neural networks were the methods used to predict the HVGO properties, such as occurred when predicting the fuel oil and LVGO properties. Principal component analysis was also used to reduce the number of predictive variables and to determine the principal components. These principal components were used as inputs of the PCR and ANN models. However, the PLS model was based on the original data. At the same time, the models behaviour was evaluated calculating the statistical performance indexes.

Prior to multivariate modelling, the spectral area was divided in constant segments of 0.02 ppm. As occurred when analysing the fuel oil and LVGO samples, the choice of the bin was made taking into account the predictions achieved when evaluating the possibility of using other bins. The data pre-processing was the same independently of the samples analysed. When applying the ANN model, the procedure used to determine the best network architecture was the same as presented for the fuel oil and LVGO samples. The best architecture was achieved with different numbers of neurons in the input layer depending on the number of PCs used, three neurons in the hidden layer and one neuron in the output layer. To apply the PLS model, the number of latent variables was determined taking into account the minima of the RMSECV (root mean squared error of cross validation) plot.

Table 6.4 (shown in page 187) presents the number of samples analysed for each property. Independently of the number of samples analysed, the data was divided into three different groups such as the training, the validation and the test data sets. All three different groups were representative of the data using the procedure described in Section 5.3.

The results obtained for each property are discussed as followed and the values of performance indexes calculated for both data sets, training and validation data sets and test data set, and for all properties analysed when using the different multivariate models, presented.

#### **i) Density (g/mL)**

The study performed to analyse the prediction of the density of the HVGO samples by  $^1\text{H}$  NMR spectroscopy was developed considering 189 samples. The training data set was constituted by 113 samples while the validation and test data sets included 38 samples, respectively. The density of the HVGO samples analysed varied between 0.9200 to 0.9600 g/mL. To develop the PCR and ANN models 14 PCs responsible for 98.1% of the total variance were used as inputs while for the PLS model, 11 LVs were necessary.

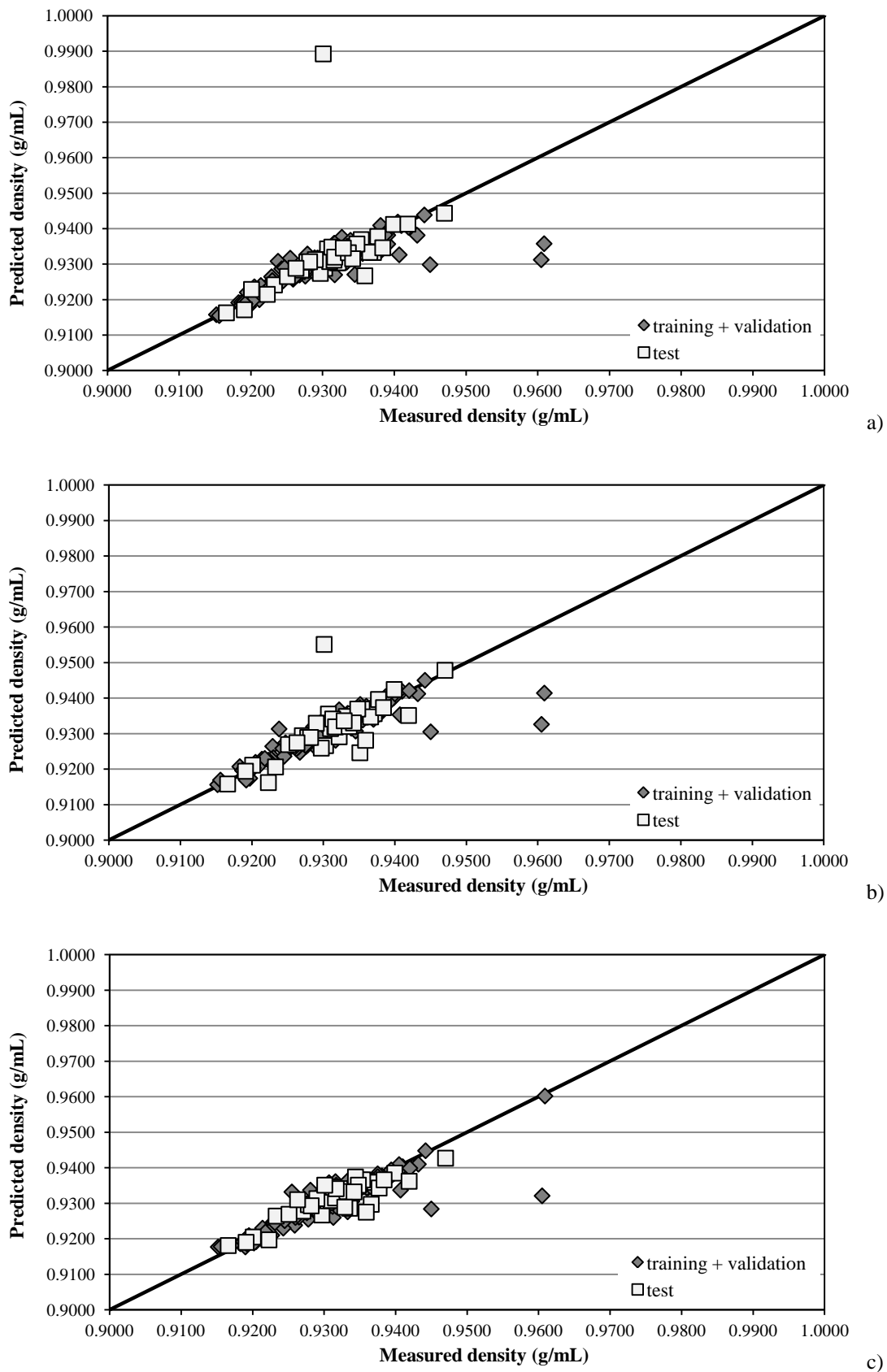
Table 6.23 presents the values of the performance indexes using PCR, PLS and ANN models for both, training and validation data sets and test data set. The results obtained demonstrated that only the ANN model presented better predictions for both data sets. When using the other models, PCR and PLS models, worst predictions were obtained especially for the test data set.

**Table 6.23:** Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the density of HVGO samples.

	PCR		PLS		ANN	
	training and validation	test	training and validation	test	training and validation	test
R <sup>2</sup> /NSE	0.64	-0.48	0.73	0.28	0.76	0.76
MAE	0.002	0.003	0.002	0.003	0.002	0.002
RMSE	0.004	0.007	0.003	0.005	0.003	0.003
MBE	0.0000	0.0008	0.0000	0.0002	0.0000	-0.0003
PBIAS	0.00	0.09	0.00	0.02	0.00	-0.03
RSR	0.60	1.22	0.52	0.84	0.49	0.49
Model performance	<b>satisfactory</b>	<b>unsatisfactory</b>	<b>good</b>	<b>unsatisfactory</b>	<b>very good</b>	<b>very good</b>

The results presented in Table 6.23 show that the use of the ANN model led to more accurate results than the PCR and PLS models, for both data sets. It was with the ANN model that higher coefficients of determination or NSE values were obtained for both data sets. In addition, it was also with the ANN model that smaller errors were achieved, such as smaller RMSE, MAE and RSR results. PBIAS and MBE were negatives for the ANN model and both data sets, meaning that the predicted density values were underestimated. When analysing the linear plot between the measured and the predicted density by the different multivariate models and for both data sets (Figure 6.21), the same conclusions were made. As illustrated, it was with the ANN model that the agreement between the predicted and measured results were most remarkable (see Figure 6.21c). Some dispersion around the diagonal line was notorious, especially for higher density results. However, samples with smaller density values were very well predicted. Consequently, the models obtained, for both data sets, when using the ANN model were of very good quality. Less accurate models were achieved when the PCR and PLS models were applied. As visualized in Figure 6.21a and Figure 6.21b, larger deviations between the predicted and measured values were obtained showing that these developed models were of low quality, especially when predicting the samples of the test data set.

Concluding, <sup>1</sup>H NMR spectroscopy combined with the ANN model was the most reliable to predict the density of HVGO samples.



**Figure 6.21:** Correlation between the experimental and predicted density of the HVGO samples obtained by: a) PCR, b) PLS and c) ANN models.

## ii) Kinematic viscosity at 100°C (cSt)

To predict the kinematic viscosity at 100 °C, 168 HVGO samples were analysed. From the 168 HVGO samples, 100 samples represented the training data set while 34 samples constituted the validation and test data sets, respectively. With all these samples, a wide variety of kinematic viscosity results ranging from 8.050 to 18.60 cSt were guaranteed.

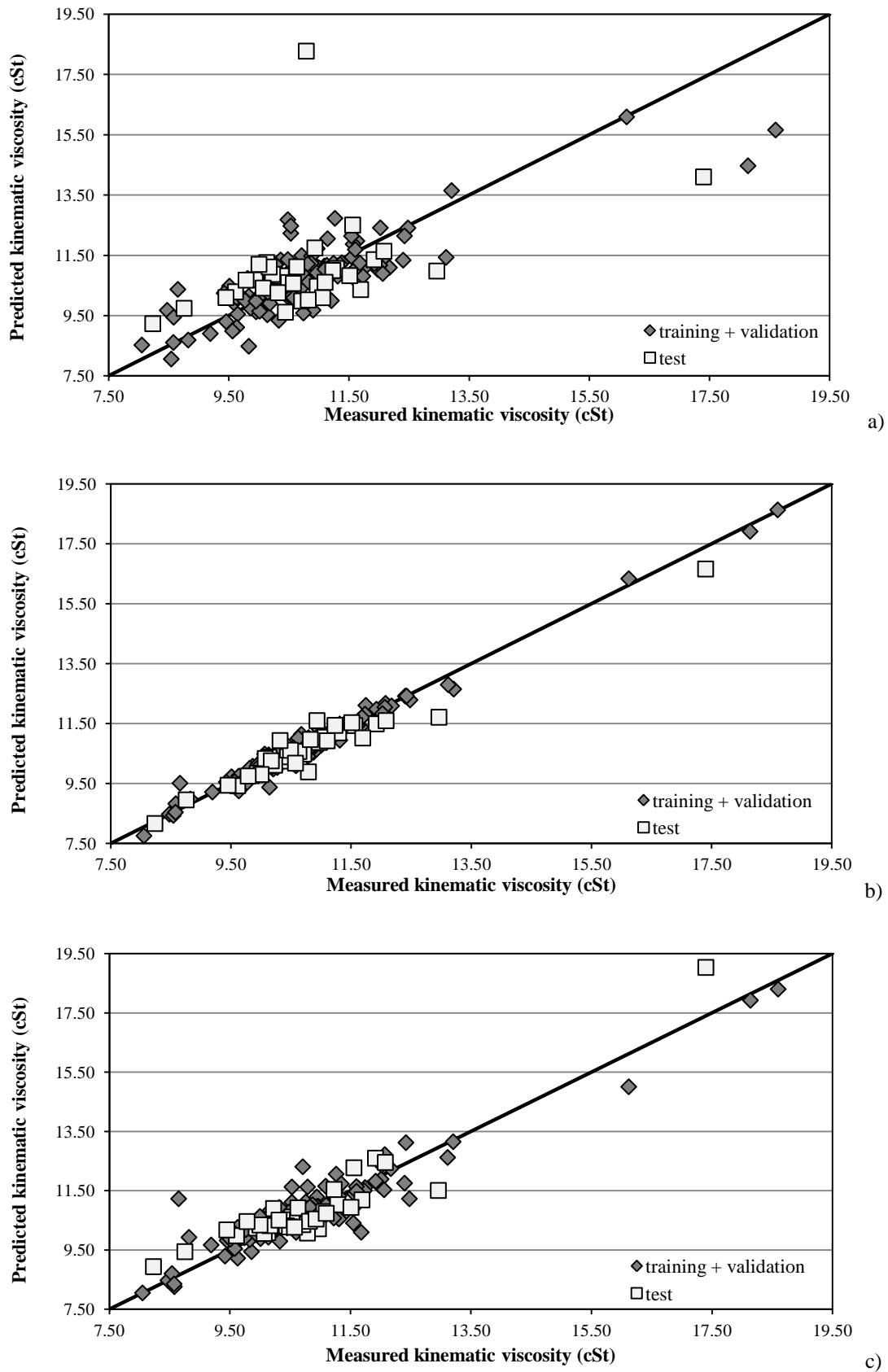
Table 6.24 presents the performance indexes achieved when applying the different multivariate models for both data sets, training and validation data sets and test data set, when predicting the kinematic viscosity of HVGO samples. Sixteen principal components, explaining 98.3% of the original data variance were used as inputs of the PCR and ANN models while for the PLS model 20 LVs to perform good PLS models were required. As visualized in Table 6.24, better performance indexes were achieved with the PLS and ANN models for both data sets. However, when applying the PCR model good predictions were obtained for the training and validation data sets, whereas for the test data set unsatisfactory results were achieved. PCR was not considered a good model to predict the kinematic viscosity of HVGO samples due to the performance indexes obtained. This model was characterized for having insignificant coefficient of determination or negative NSE results which indicated that the mean observed value was better measured than predicted, meaning that unacceptable performances were obtained. Analysing the errors obtained, it was verified that higher RMSE, MAE and RSR values were achieved with the PCR model.

**Table 6.24:** Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the kinematic viscosity at 100 °C of HVGO samples.

	PCR		PLS		ANN	
	training and validation	test	training and validation	test	training and validation	test
R <sup>2</sup> /NSE	0.68	-0.20	0.97	0.93	0.87	0.84
MAE	0.54	0.96	0.16	0.27	0.34	0.46
RMSE	0.78	1.60	0.22	0.39	0.50	0.58
MBE	0.00	0.19	0.00	-0.12	0.02	0.07
PBIAS	0.00	1.75	0.00	-1.12	0.14	0.64
RSR	0.56	1.10	0.16	0.27	0.36	0.40
Model performance	<b>good</b>	<b>unsatisfactory</b>	<b>very good</b>	<b>very good</b>	<b>very good</b>	<b>very good</b>



As visualized in Table 6.24, when using the PLS model, better predictions were obtained for both data sets. It is important to point out that when using the PLS model very good NSE values were obtained for both data sets. It was with this model that the smallest errors were obtained, such as, smaller RMSE, MAE and RSR results, when compared with the ones obtained with the PCR and ANN models. Analysing the PBIAS and MBE values, it was possible to conclude that accurate model simulation was obtained for the training and validation data sets, whereas for the test data set the predicted results were underestimated. As illustrated in Figure 6.22b, which shows the measured and predicted values of the kinematic viscosity determined by PLS model for both data sets, the predicted results obtained presented a very good agreement with the experimental values. Small and high HVGO kinematic viscosities were well predicted when using the PLS model and consequently high quality models, for both data sets, were achieved. Less accurate models were obtained when the ANN model was applied, although still of high quality with both data sets characterized with very good predictions. As indicated in Figure 6.22c larger deviations compared to Figure 6.22b were found as well as some difficulties in predicting higher kinematic viscosities of the HVGO samples. However, the ANN model predict more than 80% of the data. The same predictions were not possible to obtain when using the PCR model. As illustrated in Figure 6.22a a weak agreement between the predicted and measured results were found when the PCR model was applied, especially for the test data set. Consequently, unsatisfactory predictions were achieved. Figure 6.22 clearly demonstrated that  $^1\text{H}$  NMR spectroscopy combined with the PLS model could be a very good alternative to predict the kinematic viscosity of HVGO samples.



**Figure 6.22:** Correlation between the experimental and predicted kinematic viscosity at 100 °C of the HVGO samples obtained by: a) PCR, b) PLS and c) ANN models.

### iii) Carbon residue (% m/m)

As already stated, carbon residue is a very important property that provides information on coking propensity of HVGO samples during combustion or conversion process in hydrotreatment units. To determine this property using  $^1\text{H}$  NMR spectroscopy, 142 HVGO samples were analysed. The training data set was composed by 86 samples while the test and the validation data sets were formed with 28 samples each. With all samples analysed, a wide variety of results ranging from 0.08 to 1.03% m/m were guaranteed. To develop the PCR and ANN models 15 principal components, explaining 98.2% of the original data variance, were used as inputs. For the PLS model, 17 LVs were required.

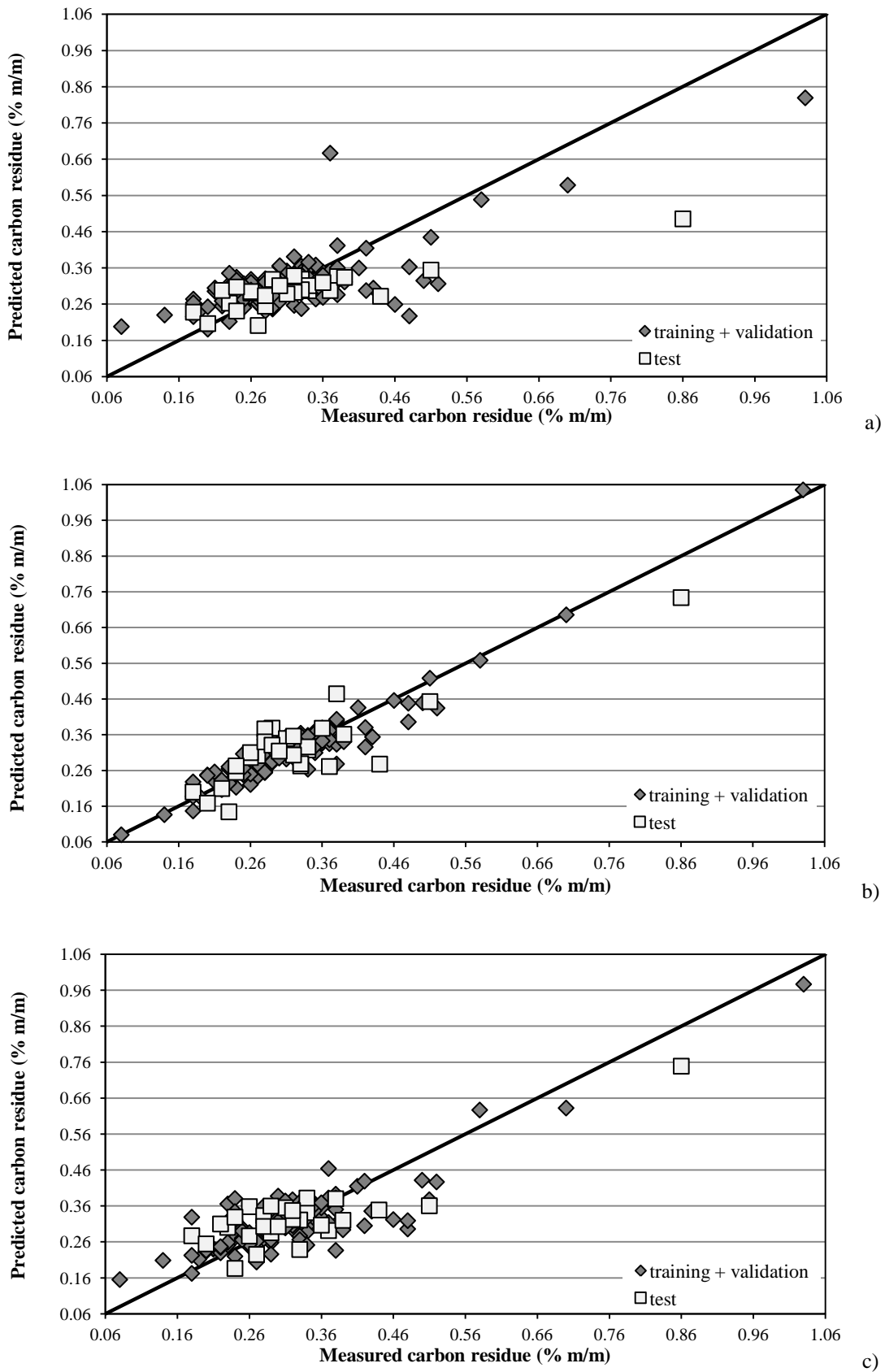
As visualized in Table 6.25, the performance indexes calculated for the training and validation data sets and for the test data set demonstrated that the PLS model performed very good predictions. When applying the ANN model good predictions were obtained for both data sets. On the other hand, worst performance occurred, for both data sets, with the PCR model.

**Table 6.25:** Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the carbon residue of HVGO samples.

	PCR		PLS		ANN	
	training and validation	test	training and validation	test	training and validation	test
R <sup>2</sup> /NSE	0.55	0.48	0.91	0.76	0.72	0.71
MAE	0.05	0.06	0.02	0.05	0.04	0.06
RMSE	0.07	0.09	0.03	0.06	0.06	0.07
MBE	0.000	-0.030	0.000	-0.003	-0.001	0.001
PBIAS	0.00	-8.01	0.00	-0.80	-0.21	0.35
RSR	0.67	0.72	0.30	0.50	0.53	0.54
Model performance	<b>satisfactory</b>	<b>unsatisfactory</b>	<b>very good</b>	<b>very good</b>	<b>good</b>	<b>good</b>

Once again, the PLS model demonstrated to be the best alternative when predicting the carbon residue of HVGO samples (see Table 6.25). As demonstrated, it was with the PLS model that better performance indexes were obtained for both data sets. Although, for the test data set, the predictions were not as high as occurred for the training and validation

data sets. It was with the PLS model, for both data sets, that higher NSE values were obtained. In addition, smaller RMSE, MAE and RSR results also characterized the developed PLS model. PBIAS and MBE indicated that for the training and validation data sets an accurate model was obtained. However, for the test data set the results were underestimated. Since very good predictions were obtained when using the PLS model, this model was considered the best one to predict the carbon residue of HVGO samples. As analysed in Figure 6.23b, the developed model gave accurate results, although some dispersion between the measured and predicted results were notorious, especially for the test data set. The larger deviation between the predicted and measured values of the test data set could be the reason why the performance indexes obtained for the test data set was not as high as the ones obtained for the training and validation data sets. Although, for both data sets, the developed models were of very good quality predicting 91 and 76% of the data for the training and validation data sets and test data set, respectively. Smaller correlations were obtained when the PCR and ANN model were used. As visualized in Figure 6.23a and Figure 6.23c, the PCR and ANN models presented some difficulties in predicting smaller and higher carbon residue values. Moreover, larger deviations between the predicted and experimental carbon residue results, when compared to Figure 6.23b, were found. Though, these larger deviations were more pronounced when the PCR model was applied demonstrating that the agreement between the predicted and measured carbon residue was not remarkable. For these reasons, PCR model, which haven't predicted more than 48% of the test data set, were not considered a good model to predict the carbon residue of HVGO samples. On the other hand, the ANN model gave slightly more accurate results than the PCR model and a better agreement between the predicted and measured results were achieved, as illustrated in Figure 6.23c. Consequently, models with better quality, for both data sets, than the ones achieved with the PCR model were obtained. ANN model predicted more than 70% of the data. Figure 6.23 demonstrated that  $^1\text{H}$  NMR spectroscopy combined with the PLS model could be an alternative to the ASTM standard test method to determine the carbon residue of the HVGO samples.



**Figure 6.23:** Correlation between the experimental and predicted carbon residue of the HVGO samples obtained by: a) PCR, b) PLS and c) ANN models.

## **Distillation**

As occur for LVGO samples, the cuts of the distillation curve that were analysed to evaluate the HVGO samples were the 5, 10, 50, 90 and 95% distillation cuts. Analysing these cuts of the distillation curve it was evaluate how the separation of the HVGO stream occurred. The idea was to verify if there were no losses of HVGO and if some residue were collected in the HVGO stream.

$^1\text{H}$  NMR spectroscopy was used to verify if it can described, with good precision, the distillation curve of HVGO samples.

### **iv) 5% Distillation**

When predicting 5% distillation cut of HVGO samples through the use of  $^1\text{H}$  NMR spectroscopy 116 samples were used, from which, 70 samples composed the training data set while 23 samples formed the validation and test data sets, respectively. This distillation cut included a wide variety of results ranging from 374 to 418 °C. Thirteen principal components, responsible for explaining 98.3% of the total variance, were used as inputs of the PCR and ANN models. Considering the PLS model, 10 LVs were required.

Table 6.26 clearly demonstrated the results of the performance indexes obtained when using PCR, PLS and ANN models to predict both data sets, training and validation data sets and test data set. As visualized, the predictions obtained were not as good as expected. It was expected better predictions with all applied models. Only the PLS model was able to predict the 5% distillation cut of the HVGO samples. When using the ANN model only satisfactory results were obtained for both data sets. Worse predictions were achieved when using the PCR model.

**Table 6.26:** Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the 5% distillation cut of HVGO samples.

	PCR		PLS		ANN	
	training and validation	test	training and validation	test	training and validation	test
R <sup>2</sup> /NSE	0.18	0.29	0.75	0.66	0.62	0.55
MAE	8.08	7.86	4.06	5.54	4.55	5.49
RMSE	10.03	9.63	5.53	6.67	6.85	7.64
MBE	0.00	-1.97	0.00	-2.80	-0.61	-2.93
PBIAS	0.00	-0.51	0.00	-0.73	-0.16	-0.76
RSR	0.90	0.84	0.50	0.58	0.62	0.67
Model performance	<b>unsatisfactory</b>	<b>unsatisfactory</b>	<b>very good</b>	<b>good</b>	<b>satisfactory</b>	<b>satisfactory</b>

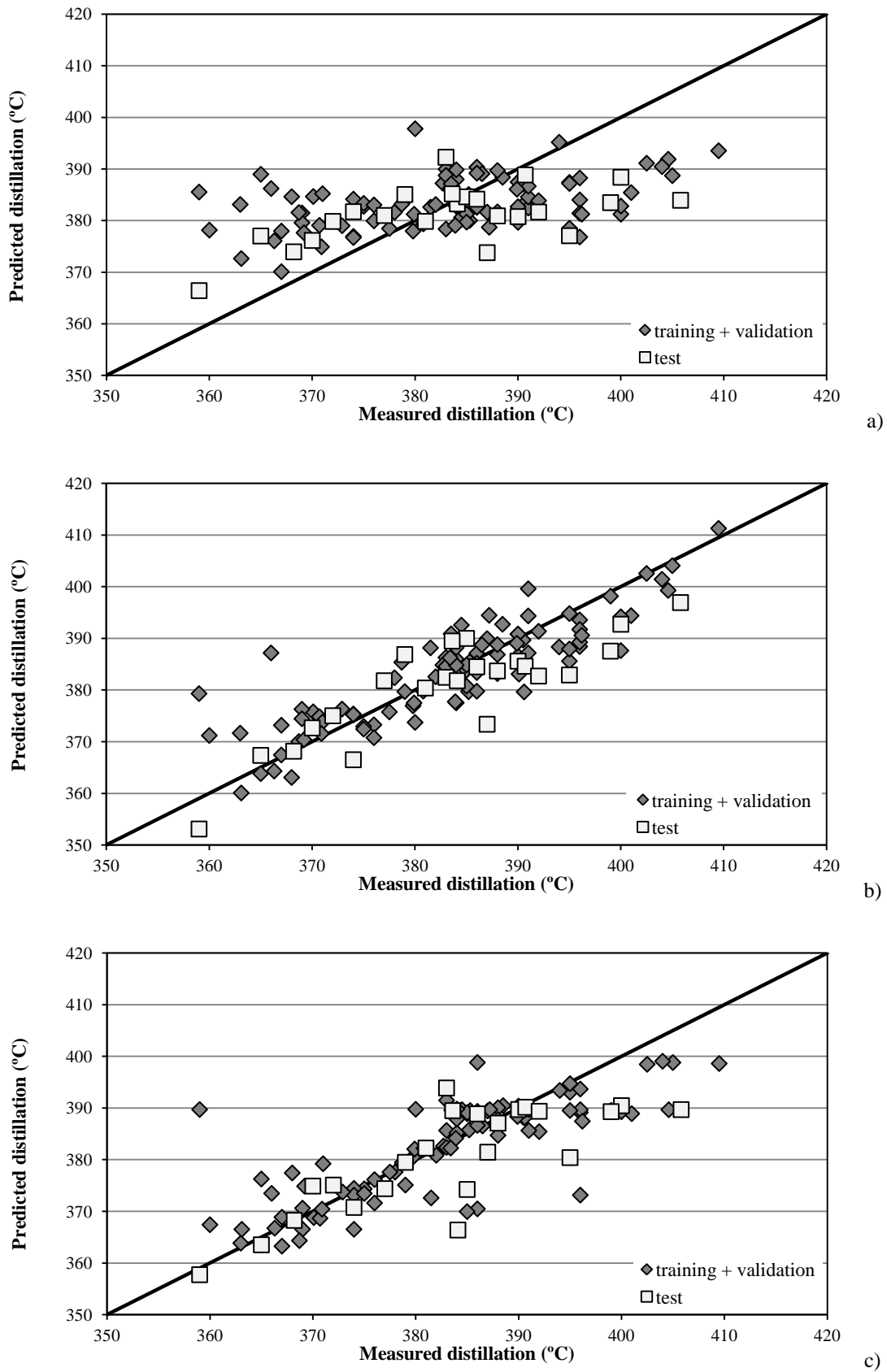
Comparing the predictions with all models and the measured data, corresponding only to the training and validation data sets (see Table 6.26), it was possible to visualize that worse predictions were obtained when using the PCR model. When applying the ANN model a small improvement was achieved and therefore, the model classified as satisfactory. On the other hand, when applying the PLS model better predictions were obtained. The PLS model, classified as very good, was characterized for predicting 75% of the data. It was with the PLS model that high NSE value, for the training and validation data sets, was obtained. Moreover, small errors, such as smaller RMSE, MAE and RSR values, were also achieved for the training and validation data sets comparing with those obtained with the PCR and ANN models. The results obtained with PBIAS and MBE were indicative that an accurate model simulation was achieved with the PLS model.

When analysing the results with all models only for the test data set, it was verified that worse predictions were obtained when comparing with those achieved for the training and validation data sets. As visualized in Table 6.26, the PCR model was not a good choice to determine the 5% distillation cut of HVGO samples. The model obtained was classified as unsatisfactory. The results obtained with the ANN model were also not so good. As occurred for the training and validation data sets, the best model to predict the 5% distillation cut of the HVGO samples, for the test data set, was the PLS model. With this model, a good NSE value was achieved and when comparing with the other results obtained for the test data set, it was with the PLS model, that small errors were obtained.

The results obtained with the PLS model for PBIAS and MBE were negative, meaning that, the predicted 5% distillations were underestimated.

Figure 6.24 depicts a correlation plot between the predicted and measured results obtained when applying the multivariate models and the standard test method, respectively. As illustrated in Figure 6.24b, it was with the PLS model that better agreement between the predicted and measured results was obtained and hence best performance, for both data sets, were achieved. Although, some dispersion between the predicted and measured results was visualized as well as some difficulties to predict lower temperatures of the 5% distillation cut of HVGO samples. The larger deviations between the predicted and experimental results was more notorious at the test data set. These larger deviations of the samples from the test data set could be one of the reasons that justify the lower predictions at the test data set, compared to the training and validation data sets. Despite these results, it was with the PLS model that the most accurate models to predict the 5% distillation cut were achieved. When using the ANN model less accurate models than the ones obtained with the PLS model were found. Larger deviations between the predicted and measured results were visualized, for both data sets, when applying the ANN model. Moreover, difficulties in predicting higher values of the 5% distillation cut was also illustrated in Figure 6.24c. Figure 6.24a clearly demonstrates that very low quality models were obtained when the PCR model was applied. No agreement between the predicted and measured results were visualized in Figure 6.24a and consequently the PCR model was not considered an alternative to predict the 5% distillation cut of HVGO samples. Concluding, the PLS model was the most reliable to predict the 5% distillation cut of the HVGO samples.





**Figure 6.24:** Correlation between the experimental and predicted 5% distillation cut of the HVGO samples obtained by: a) PCR, b) PLS and c) ANN models.

#### iv) 10% Distillation

To determine the cut of 10% distillation, 116 HVGO samples were studied. With all these samples, a wide variety of results was guaranteed with the 10% distillation cut ranging from 374 to 418 °C. From all variables, three data sets were formed, and the training data set composed of 70 samples while the validation and test data sets were constituted by 23 samples, respectively. Before applying PCR and ANN models, principal component analysis was used to determine the number of PCs used as inputs of these models. Twelve principal components, explaining 98.1% of the total variance, were selected. To apply the PLS model, the number of latent variables were also selected and better predictions were obtained when 10 LVs were used.

The performance indexes obtained when applying all multivariate models for both training and validation data sets and test data set were presented in Table 6.27. As visualized, only the PLS model was able to predict the 10% distillation cut, for both data sets. With the PLS model very good predictions were obtained for the training and validation data sets and for the test data set, good predictions were achieved. When applying the ANN models, very good predictions were obtained for the training and validation data sets. However, the same doesn't occur for the test data set, which was classified as an unsatisfactory model. In addition, unsatisfactory predictions were also obtained for both data sets when the PCR model was applied. PCR model was not considered a good model, for both data sets, to predict the 10% distillation cut of HVGO samples.

**Table 6.27:** Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the 10% distillation cut of HVGO samples.

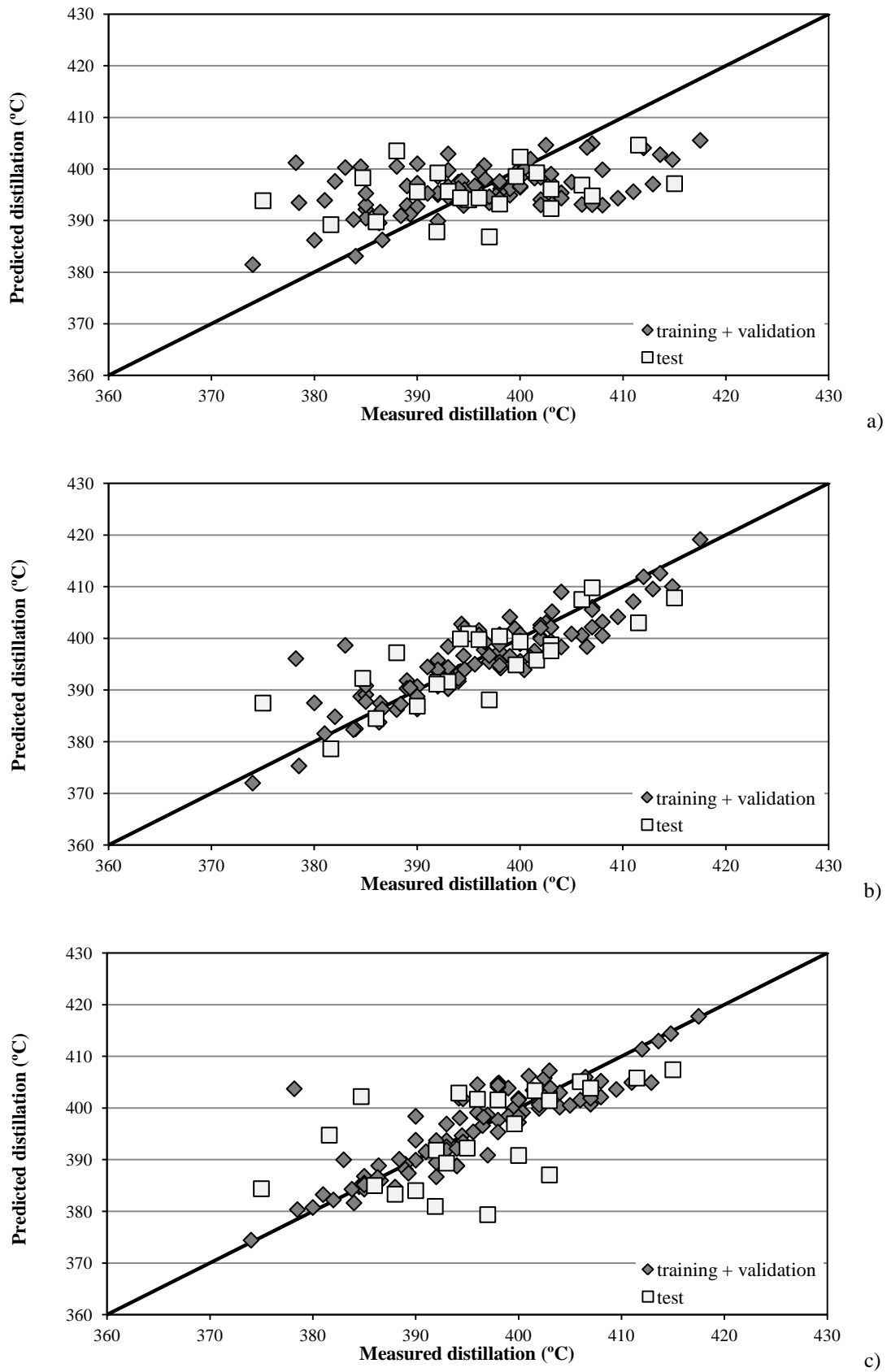
	PCR		PLS		ANN	
	training and validation	test	training and validation	test	training and validation	test
R <sup>2</sup> /NSE	0.24	0.07	0.78	0.66	0.76	0.18
MAE	6.00	7.21	3.09	4.65	2.83	6.66
RMSE	7.78	9.00	4.21	5.62	4.38	8.47
MBE	0.00	-0.51	0.00	-0.20	0.41	-1.46
PBIAS	0.00	-0.13	0.00	-0.05	0.10	-0.37
RSR	0.87	0.96	0.47	0.60	0.49	0.91
Model performance	<b>unsatisfactory</b>	<b>unsatisfactory</b>	<b>very good</b>	<b>good</b>	<b>very good</b>	<b>unsatisfactory</b>

Analysing the predictions with all models, corresponding to the training and validation data sets (Table 6.27), it was possible to conclude that the worst performance occurred with PCR model. The application of PCR model was not considered a good way to predict the 10% distillation cut of HVGO samples. However, the use of ANN and PLS models led to more accurate results than the PCR model. Very good predictions were obtained when using the ANN and PLS models. Although the ANN and PLS models presented similar results, for the training and validation data sets, the PLS model was considered to be the best. The main difference, when analysing the performance indexes achieved (Table 6.27) were found in PBIAS and MBE results. For the PLS model the PBIAS and MBE results were indicative that an accurate model simulation was achieved, while, for the ANN model, PBIAS and MBE were negative, meaning that, the predicted 10% distillation results were underestimated.

Such good results, as those obtained for the training and validation data sets, were not achieved for the test data set. Table 6.27 shows that the application of the PLS model was considered better than using PCR and ANN models to predict the 10% distillation cut of the HVGO samples. Considering the PCR and ANN models, it was observed that the predicted results were not adjusted to the measured data. Both these methods were not able to predict, for the test data set, the 10% distillation cut of the HVGO samples. The predictions obtained with the PLS model for the test data set were classified as good, meaning that, a good NSE value and small errors, compared with those obtained for the ANN and PCR models, were attained. Taking into account the results obtained with the PBIAS and MBE, the predicted 10% distillation were underestimated.

The results obtained indicated that  $^1\text{H}$  NMR spectroscopy could predict the 10% distillation cut of the HVGO samples by using the PLS model. As visualized in Figure 6.25b, which shows the predicted and measured values and the goodness of the fit obtained, it was with the PLS model that a better agreement between the predicted and measured results was achieved. Some dispersion around the calibration line was also illustrated, especially for the test data set, indicating that less quality performance indexes were obtained for the test data set, compared to the ones obtained for the training and validation data sets. In spite of the performance indexes obtained for both data sets when using the PLS model were not of higher accuracy, it was with the PLS model that the best

predictions were achieved. With the ANN model, as illustrated in Figure 6.25c, the training and validation values mostly coincide with the calibration line, whereas the test data set showed larger deviations, compared with the Figure 6.25b. These larger deviations from the test data set contributed to a less quality model and consequently lower performance indexes. Moreover, low quality models were also obtained when the PCR model was applied. As illustrated in Figure 6.25a, no agreement between the predicted and measured results, for both data sets, were found when the PCR model was used. These results leads us to conclude that PLS model was the most reliable to predict the 10% distillation cut of the HVGO samples.



**Figure 6.25:** Correlation between the experimental and predicted 10% distillation cut of the HVGO samples obtained by: a) PCR, b) PLS and c) ANN models.

### v) 50% Distillation

The prediction of 50% distillation of HVGO samples was made using 116 samples. From the 116 HVGO samples, 70 were used for the training data set while 23 composed the test and validation data sets, respectively. The samples used to develop the model included a wide variety of results ranging from 433 to 466 °C. Thirteen principal components, responsible for 98.2% of the total variance, were used as inputs of the PCR and ANN models. Thirteen latent variables were required to apply the PLS model.

The results obtained when predicting the 50% distillation of HVGO samples, presented in Table 6.28, were not as good as expected. From all models applied only the ANN model gave similar and good predictions for both data sets. The PCR model was not able to predict the 50% distillation cut of the HVGO samples and unsatisfactory predictions were obtained for both data sets. Using the PLS model, very good performance indexes were obtained for the training and validation data sets, however, the performance indexes for the test data set were slightly worse, where only satisfactory predictions were obtained.

**Table 6.28:** Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the 50% distillation cut of HVGO samples.

	PCR		PLS		ANN	
	training and validation	test	training and validation	test	training and validation	test
R <sup>2</sup> /NSE	0.35	0.41	0.90	0.54	0.66	0.72
MAE	3.52	3.58	1.41	2.95	2.73	2.58
RMSE	4.52	4.40	1.82	3.86	3.44	3.00
MBE	0.00	0.91	0.00	1.08	0.23	1.00
PBIAS	0.00	0.20	0.00	0.24	0.05	0.22
RSR	0.80	0.77	0.32	0.68	0.61	0.52
Model performance	<b>unsatisfactory</b>	<b>unsatisfactory</b>	<b>very good</b>	<b>satisfactory</b>	<b>good</b>	<b>good</b>

Considering only the results obtained and the measured data during training and validation data sets (Table 6.28), it was demonstrated that only PCR was not able to predict the 50% distillation cut of the HVGO samples. Very small NSE value and high errors were obtained when applying the PCR model. Only the MBE and PBIAS values presented a very good

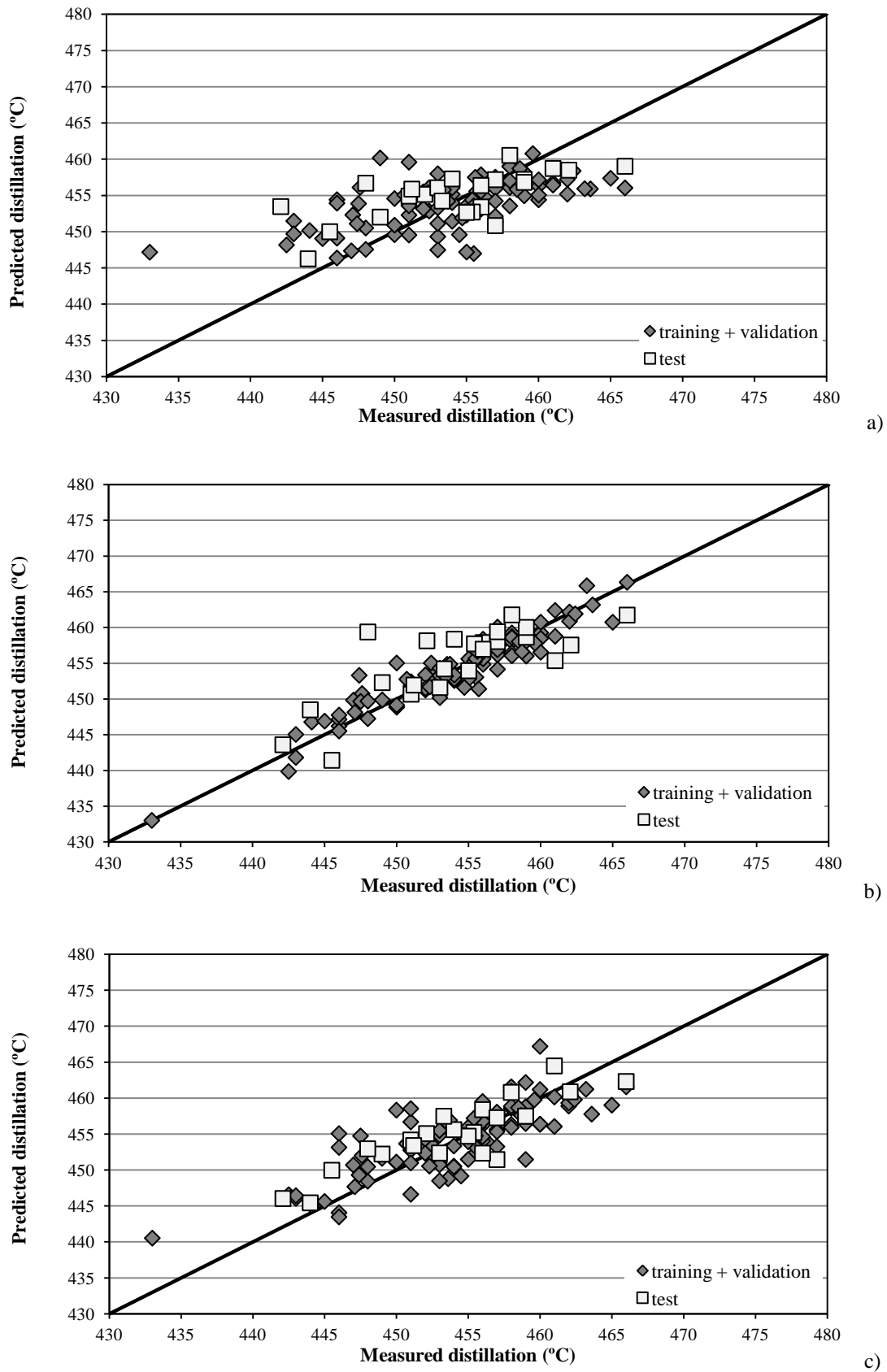
result for the PCR model, which was expected since these statistical parameters tend to minimize the real value to the average one. On the other hand, a good model was obtained when applying the ANN model and when using the PLS model very good predictions were achieved. For the training and validation data sets, the best performance indexes were obtained when the PLS model was applied. The PLS model was characterized for presenting a high NSE value and small errors, such as a smaller RMSE, MAE and RSR results, comparing with those obtained with PCR and ANN models. The PBIAS and MBE obtained with the PLS model were indicative that an accurate model simulation was achieved.

Taking into account the predicted results obtained for the test data set, when using all different multivariate models, the results were not as good as the results obtained for the training and validation data sets. As visualized in Table 6.28, the PCR model was not able to predict the 50% distillation cut of the HVGO samples and unsatisfactory predictions were obtained. Better results than the ones obtained with the PCR model were achieved with the PLS model, nevertheless, only satisfactory predictions were obtained. It was only possible to obtain good predictions when using the ANN model. This model was characterized for presenting good performance indexes, taking into account the performance ratings developed by Moriasi *et al.* (2007) (see Section 5.4). Also it was with the ANN model that smaller errors were obtained for the test data set. The MBE and PBIAS values indicated that the model developed by the ANN was overestimated.

Figure 6.26, which depicts a correlation plot of the measured and predicted results with all models, confirmed that the ANN model is the most reliable to predict the 50% distillation cut of HVGO samples. As illustrated in Figure 6.26c, it was with the ANN model that a better agreement between the predicted and experimental results, especially for the test data set, was found. It was visualized some deviation of the predicted results from the correlation line, although it was with the ANN model that the performance indexes of both data sets were in accordance and high quality model for the test data set was achieved. When applying the PLS model (Figure 6.26b), it was visualized that the training and validation results almost coincide with the calibration line, although the same doesn't occur for the test data set. As visualized, larger deviations of the test data set from the calibration line, compared to Figure 6.26c, were achieved. To confirm the results obtained

when using the PLS model, the samples included in the test data set were changed with the samples encompassed at the training and validation data sets, although the same behaviour occurs. The PLS model was unable to predict the test data set with high accuracy as occurs for the training and validation data sets. This behaviour could indicate that the selection criterion of taking the number of LVs that corresponds with the minimal RMSECV value might result in overfitting the data, although, as expected, the use of less LVs contributed to less accurate models for both data sets. The PCR model was also unable to predict the 50% distillation cut of HVGO samples and low quality models were obtained for both data sets. As illustrated in Figure 6.26a there was no agreement between the predicted and experimental results. Concluding, the ANN model was considered the best model, with good agreement between both data sets, to predict the 50% distillation cut of HVGO samples.





**Figure 6.26:** Correlation between the experimental and predicted 50% distillation cut of the HVGO samples obtained by: a) PCR, b) PLS and c) ANN models.

**vi) 90% Distillation**

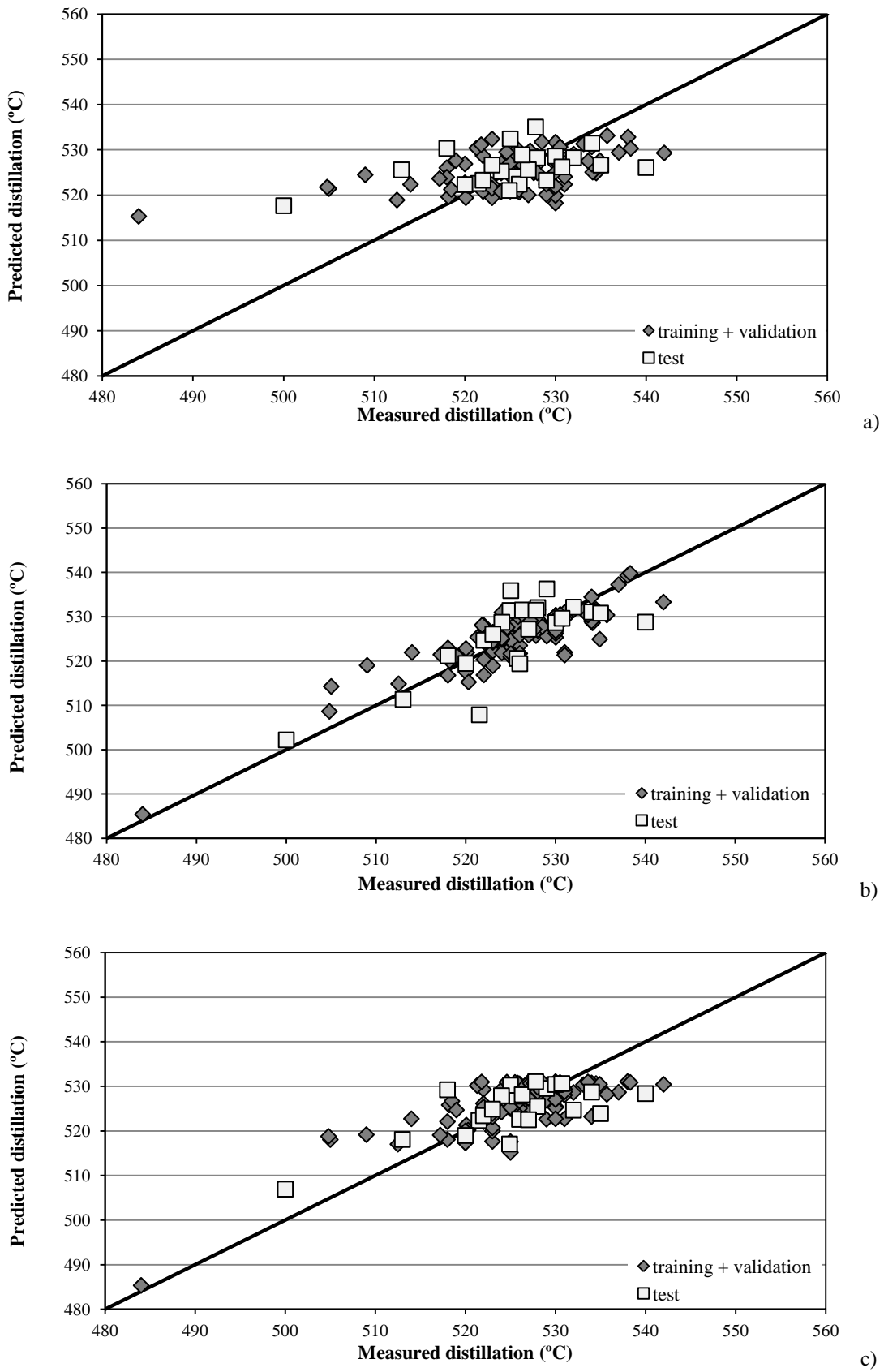
To develop a model that contributes to obtain the 90% distillation results, 114 HVGO samples were analysed. From the 114 samples, 68 composed the training data set while 23 were used for the validation and test data sets, respectively. The samples used to develop the model included a wide variety of results ranging from 484 to 542 °C. Before applying the multivariate models the principal component analysis was used. With the PCA, the number of principal components used as inputs of the PCR and the ANN models were selected. For predicting the 90% distillation cut, 12 PCs, responsible for explaining 98.1% of the original data variance, were required. When applying PLS model, 12 LVs were select.

Table 6.29 demonstrates the results of the performance indexes obtained after applying all different multivariate models for both training and validation data sets and test data set. As analysed, bad predictive performance were obtained when predicting the 90% distillation cut of the HVGO samples. Independently of the model applied, the PCR, PLS and ANN models, unsatisfactory predictions were obtained for the test data set, meaning that, neither models were able to predict the measured values from the test data set. For the training and validation data sets better predictions were achieved when using the ANN and the PLS models. However, it was with the PLS model that the best performance were obtained. It was with the PLS model that the predicted results were closer from the measured data and higher NSE value was obtained, compared with those achieved for the PCR and ANN models. The smaller errors, such as, a smaller RMSE, MAE and RSR, were also found at the training and validation data sets when using the PLS model. The PBIAS and MBE results indicated that an accurate model simulation was achieved when using the PLS model.

**Table 6.29:** Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the 90% distillation cut of HVGO samples.

	PCR		PLS		ANN	
	training and validation	test	training and validation	test	training and validation	test
R <sup>2</sup> /NSE	0.25	0.20	0.75	0.48	0.58	0.50
MAE	4.78	5.23	2.92	4.44	3.81	4.32
RMSE	6.70	7.01	3.85	5.67	5.02	5.55
MBE	0.00	0.81	0.00	0.25	0.25	-0.40
PBIAS	0.00	0.15	0.00	0.05	0.05	-0.08
RSR	0.86	0.89	0.50	0.72	0.64	0.71
Model performance	<b>unsatisfactory</b>	<b>unsatisfactory</b>	<b>good</b>	<b>unsatisfactory</b>	<b>satisfactory</b>	<b>unsatisfactory</b>

In Figure 6.27, which shows the linear plot between the predicted and measured results obtained with all models applied (PCR, PLS and ANN models) as well as the standard test method, it was possible to visualize that some dispersion around the calibration line existed independently of the model applied. Low quality predictions were achieved when the PCR model was used. As illustrated in Figure 6.27a, the developed model was unable to predict samples with lower and higher values of the 90% distillation cut. Consequently, unsatisfactory predictions characterized both data sets when the PCR model was used. Surely satisfactory predictions, for the training and validation data sets, were obtained when the ANN model was applied, as visualized in Figure 6.27c. Although, for the test data set larger deviations between the predicted and measured values were achieved. Consequently, the ANN model was also not considered an alternative to predict the 90% distillation cut of HVGO samples. Better agreement between the predicted and measured values was obtained, for the training and validation data sets, when the PLS model was used. As illustrated in Figure 6.27b, the training and validation data sets were closest from the calibration line, compared to Figure 6.27a and Figure 6.27c, although the test data set continued to show larger deviations. Consequently, none of the models were able to predict the 90% distillation cut of the HVGO samples and low quality models were achieved. To improve these bad predictions more HVGO samples should be analysed. Only 23 samples at the test data set were not enough to obtain a good model. It is believed that in increasing the number of samples, especially at the test data set, that better predictions will be achieved since improvements were obtained when increasing the database from 19 samples to 23 samples.



**Figure 6.27:** Correlation between the experimental and predicted 90% distillation cut of the HVGO samples obtained by: a) PCR, b) PLS and c) ANN models.

**vii) 95% Distillation**

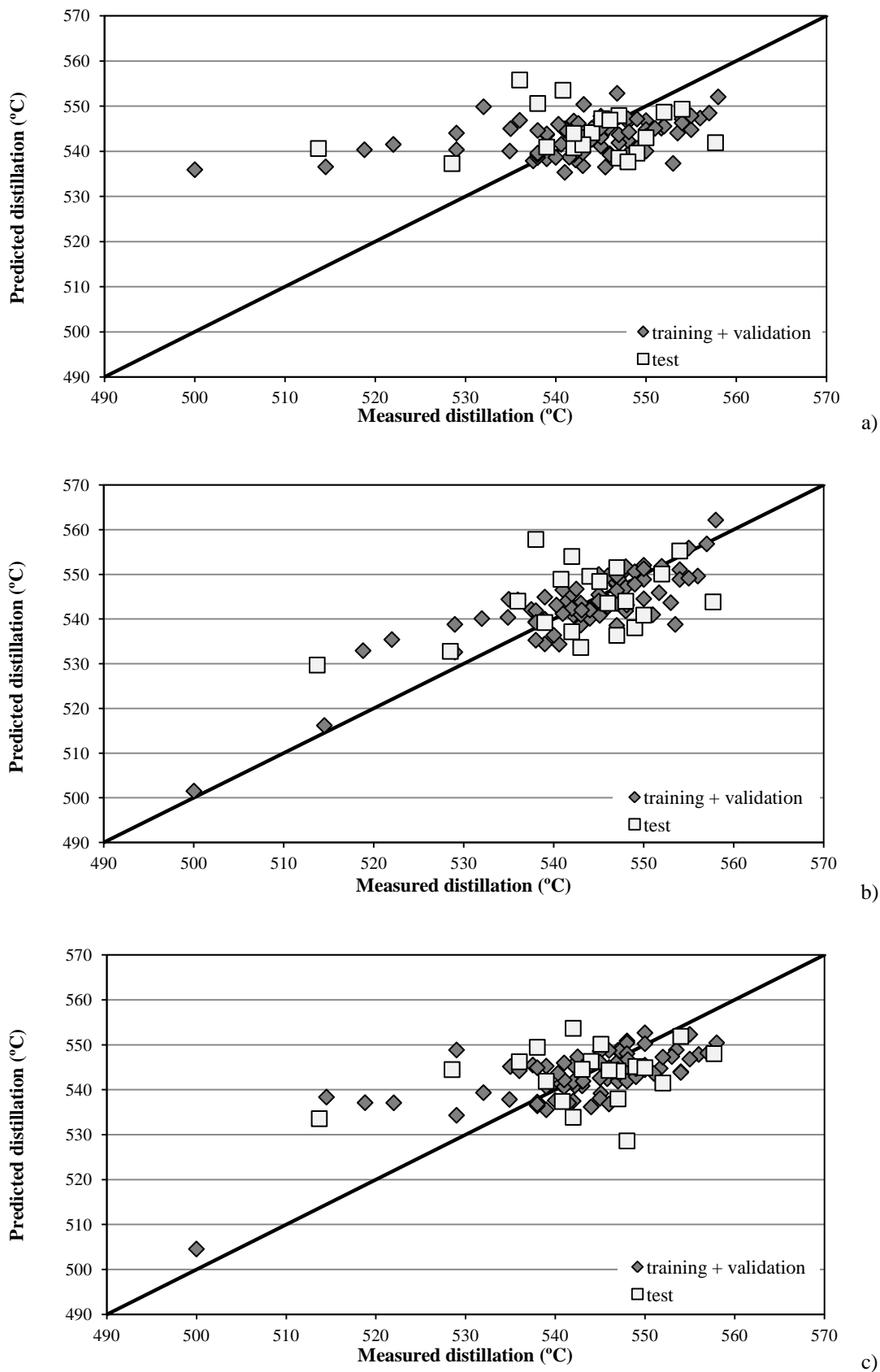
To determine the distillation cut at 95%, 99 HVGO samples were analysed. From all different 99 samples analysed, the training data set included 59 HVGO samples whereas the validation and test data sets were composed by 20 HVGO samples, respectively. The 95% distillation of all samples analysed varied between 500 to 558 °C. From the multivariate models used to predict the 95% distillation of the HVGO samples, PCR and ANN models used the PCs as inputs while the PLS model was based on the original data. The number of PCs considered by the principal component analysis was 13 PCs. These PCs were responsible for explaining 98.4% of the original data variance. When using the PLS model, 11 LVs were considered the most adequate to predict the 95% distillation cut.

Taking into account the performance obtained for both training and validation data sets and test data set, when applying the statistical performance indexes, it was possible to verify that the developed models were not able to predict the 95% distillation of the HVGO samples (Table 6.30). The results, in Table 6.30, showed that the use of PLS model led to slightly better results than the PCR and ANN models, contributing to achieve a good model for the training and validation data sets. Higher NSE value was obtained, for the training and validation data sets, when the PLS model was applied. Smaller RMSE, MAE and RSR results were also achieved when using the PLS model to predict the 95% distillation of the HVGO samples. However, for the test data set, even the PLS model failed. Moreover, all models were classified for presented insignificant coefficient of determination and higher errors for the test data set. Different approaches were tried to improve the obtained results. A more critical selection of the inputs, like an outlier analysis, to select the samples used as inputs of the applied models was made. In addition, different models were performed by changing the samples incorporated in the different data sets. However, the results obtained were even worse than the ones presented here.

**Table 6.30:** Performance indexes achieved when using multivariate modelling during training, validation and test data sets, to predict the 95% distillation cut of HVGO samples.

	PCR		PLS		ANN	
	training and validation	test	training and validation	test	training and validation	test
R <sup>2</sup> /NSE	0.18	-0.22	0.71	0.05	0.51	-0.07
MAE	5.98	7.52	3.74	7.51	4.75	7.84
RMSE	8.38	10.3	4.94	9.08	6.45	9.64
MBE	0.00	1.33	0.00	0.79	-0.14	0.25
PBIAS	0.00	0.24	0.00	0.15	-0.03	0.05
RSR	0.91	1.11	0.53	0.97	0.70	1.03
Model performance	<b>unsatisfactory</b>	<b>unsatisfactory</b>	<b>good</b>	<b>unsatisfactory</b>	<b>satisfactory</b>	<b>unsatisfactory</b>

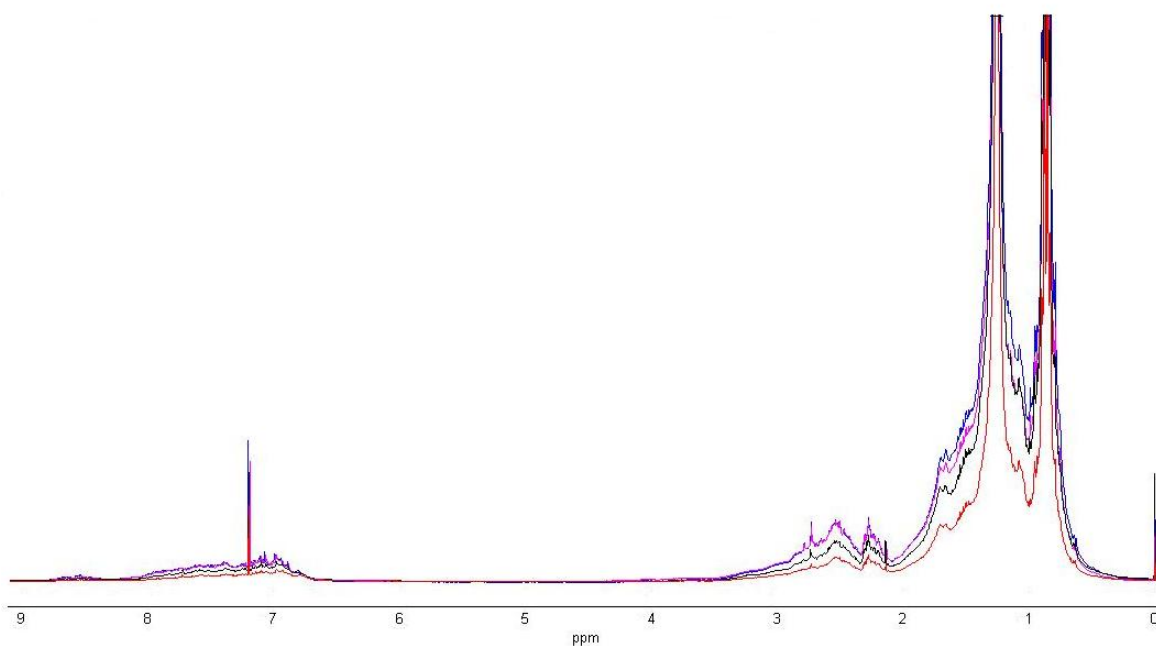
As confirmed in Figure 6.28, that depicts the correlation plot of the measured and predicted results with all models applied and for both data sets, low quality models were achieved independently of the models applied. As illustrated, a weak agreement between the predicted and measured results were obtained. The PCR and ANN models were unable to predict samples with lower or higher 95% distillation results and larger deviations from the correlation line were evidently, as illustrated in Figure 6.28a and Figure 6.28c. On the other hand, slightly better accurate model, for the training and validation data sets, were obtained when the PLS model was used and better agreement between the predicted and measured results was also visualized (Figure 6.28b). Although, larger deviations of the test data set from the correlation line were observed and hence low quality models were achieved, independently of the model applied.



**Figure 6.28:** Correlation between the experimental and predicted 95% distillation cut of the HVGO samples obtained by: a) PCR, b) PLS and c) ANN models.

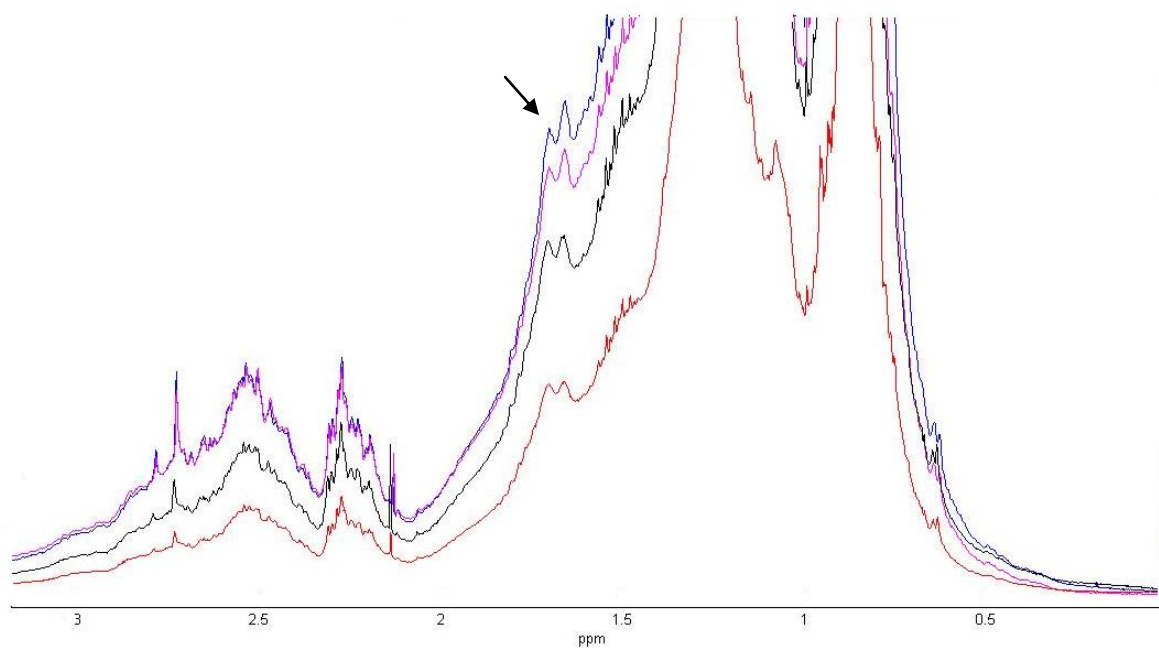
Referring to the use of  $^1\text{H}$  NMR spectroscopy to predict the 95% distillation of the HVGO samples no conclusions can be taken. Good predictions were obtained when using the PLS model for the training and validation data sets but bad predictions were achieved for the test data set. The prediction of the 95% distillation of the HVGO samples was another case where more information should be given to the developed models in an attempt to improve the obtained predictions. It was expected that the developed models should be able to predict the 95% distillation of the HVGO samples. However, unsatisfactory predictions were obtained and such predictions could be related with: i) the distillation of the heaviest fractions, which were difficult to distillate; ii) the standard method used which was characterized as a limited and complicated method where the distillation was performed at a reduced pressure and with just one theoretical plate; and iii) the reproducibility of the standard method. Due to the higher reproducibility of the standard method it was possible that small differences between the chemical composition of the samples were not differentiate with the distillation, as a consequence, different HVGO samples presenting the same distillation results were obtained. As example, Figure 6.29 presents the HVGO samples with the same distillation temperature ( $T=550\text{ }^\circ\text{C}$ ) for the 95% distillation cut. Apparently no differences were identified when comparing the  $^1\text{H}$  NMR spectrum of each of the samples but analysing more carefully it was possible to conclude that samples with differences in the chemical composition were giving the same output value.



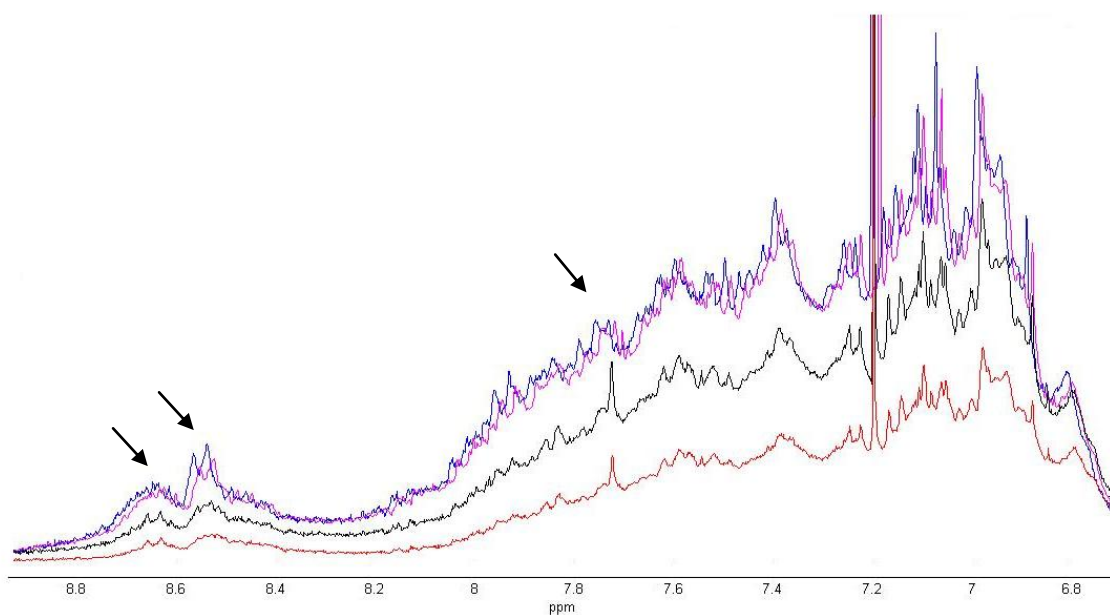


**Figure 6.29:** Comparison between some HVGO samples presenting the same distillation temperature for the 95% distillation cut (the spectrum of each HVGO sample was differentiated with different colours).

In Figure 6.30 it is clearly demonstrated that greatest differences between the HVGO samples occurred around the 1.6 to 1.8 ppm, where some beta CH, CH<sub>2</sub> and hydro-aromatic appears. It will be possible to say that the spectra represented by the blue and magenta colour were similar but different from the other two spectra which were also presenting some similarities between them. At the aromatic region (Figure 6.31) differences between the four HVGO samples were notorious, especially at the heaviest aromatic region. The greatest differences were found between the 8.4 to 8.8 ppm were some tri- and tetra-aromatic rings appears and around 7.6 to 7.8 ppm which was characterized for the presence of diaromatic and most of tri- and tetra-aromatic. Once again it was verified that the samples presented by the blue and magenta spectra were similar but different from the other two. The similarities and differences between the presented spectra could be the reason why weaker predictions were obtained and why the model couldn't be able to predict with higher accuracy the 95% distillation.



**Figure 6.30:** Expansion of the aliphatic region of the  $^1\text{H}$  NMR spectra of Figure 6.29 obtained in the 95% distillation cut (the spectrum of each HVGO sample was differentiate with different colours).



**Figure 6.31:** Expansion of the aromatic region of the  $^1\text{H}$  NMR spectra of Figure 6.29 obtained in the 95% distillation cut (the spectrum of each HVGO sample was differentiate with different colours).

### 6.2.1.3. HVGO – Final Remarks

$^1\text{H}$  NMR spectroscopy in combination with different multivariate models was applied to predict the most important physical-chemical properties used to evaluate and characterize the quality of the HVGO samples. Density, kinematic viscosity, carbon residue and distillation were the properties analysed. As demonstrated, very good predictions were obtained for both data sets, training and validation data sets and test data set, when using  $^1\text{H}$  NMR spectroscopy to predict the density, kinematic viscosity and carbon residue. These results were very important and confirmed that  $^1\text{H}$  NMR spectroscopy can be used alternatively to the traditional standard methods. The results showed (see Table 6.31) that the use of PLS model led to more accurate results than PCR and ANN models when predicting the kinematic viscosity and carbon residue. However, when predicting the density, the application of the ANN model led to better performance indexes.

The prediction of the different distillation cuts were not achieved with high accuracy as occurred for the other properties analysed. However, the first distillation cuts analysed (5, 10 and 50%) were possible to predict by  $^1\text{H}$  NMR spectroscopy, where good predictions were obtained. The use of the PLS model led to more accurate results when predicting the 5 and 10% distillation cuts, whereas better performance indexes were achieved for the 50% distillation cut when using the ANN model. Worse performances were obtained when predicting the 90 and 95% distillation cuts. The models failed when predicting the test data set of both distillation cuts. Good performance indexes were achieved when using the PLS model for the training and validation data sets of both distillation cuts (90 and 95%) whereas, unsatisfactory results were obtained when predicting the test data set of both distillation cuts (90 and 95%). Different approaches were tried in an attempt to improve the final results obtained, such as, the change of the samples that composed the different data sets. However, the results obtained were practically the same. The reason why the HVGO distillation was not so well predicted perhaps can be justify with the chemical composition of the HVGO samples and the distillation results. It was verified, by analysing the  $^1\text{H}$  NMR spectra, that in all HVGO samples collected, there were some different samples giving similarity results at a given distillation cut. Consequently, the models were not able to predict the distillation of the HVGO samples with the expected accuracy. With the results obtained, for the 5, 10 and 50% distillation cuts, it was possible to conclude that  $^1\text{H}$

NMR spectroscopy can also be used to predict the distillation of the HVGO samples. However, the performance indexes achieved, especially for the 90 and 95% distillation cuts, led to conclude that more information should be given to the develop models, suggesting that increasing the number of samples, especially for the test data set, a further improvement in the accuracy could be obtained.

**Table 6.31:** Results obtained for different properties of HVGO analysed when using <sup>1</sup>H NMR in multivariate models.

			number PC	number LV	total variance	R <sup>2</sup> /NSE	MAE	RMSE	MBE	PBIAS	RSR	model performance
<b>Density</b>	PCR	training and validation	14		98.1%	0.64	0.002	0.004	0.0000	0.00	0.60	satisfactory
		test				-0.48	0.003	0.007	0.0008	0.09	1.22	unsatisfactory
	ANN	training and validation				<b>0.76</b>	<b>0.002</b>	<b>0.003</b>	<b>0.0000</b>	<b>0.00</b>	<b>0.49</b>	<b>very good</b>
		test				<b>0.76</b>	<b>0.002</b>	<b>0.003</b>	<b>-0.0003</b>	<b>-0.03</b>	<b>0.49</b>	<b>very good</b>
	PLS	training and validation		11		0.73	0.002	0.003	0.0000	0.00	0.52	good
		test			0.28	0.003	0.005	0.0002	0.02	0.84	unsatisfactory	
<b>Kinematic viscosity</b>	PCR	training and validation	16		98.3%	0.68	0.54	0.78	0.00	0.00	0.56	good
		test				-0.20	0.96	1.60	0.19	1.75	1.10	unsatisfactory
	ANN	training and validation				0.87	0.34	0.50	0.02	0.14	0.36	very good
		test				0.84	0.46	0.58	0.07	0.64	0.40	very good
	PLS	training and validation		20		<b>0.97</b>	<b>0.16</b>	<b>0.22</b>	<b>0.00</b>	<b>0.00</b>	<b>0.16</b>	<b>very good</b>
		test			<b>0.93</b>	<b>0.27</b>	<b>0.39</b>	<b>-0.12</b>	<b>-1.12</b>	<b>0.27</b>	<b>very good</b>	
<b>Carbon residue</b>	PCR	training and validation	15		98.2%	0.55	0.05	0.07	0.000	0.00	0.67	satisfactory
		test				0.48	0.06	0.09	-0.030	-8.01	0.72	unsatisfactory
	ANN	training and validation				0.72	0.04	0.06	-0.001	-0.21	0.53	good
		test				0.71	0.06	0.07	0.001	0.35	0.54	good
	PLS	training and validation		17		<b>0.91</b>	<b>0.02</b>	<b>0.03</b>	<b>0.000</b>	<b>0.00</b>	<b>0.30</b>	<b>very good</b>
		test			<b>0.76</b>	<b>0.05</b>	<b>0.06</b>	<b>-0.003</b>	<b>-0.80</b>	<b>0.50</b>	<b>very good</b>	
<b>Distillation (5%)</b>	PCR	training and validation	13		98.3%	0.18	8.08	10.03	0.00	0.00	0.90	unsatisfactory
		test				0.29	7.86	9.63	-1.97	-0.51	0.84	unsatisfactory
	ANN	training and validation				0.62	4.55	6.85	-0.61	-0.16	0.62	satisfactory
		test				0.55	5.49	7.64	-2.93	-0.76	0.67	satisfactory
	PLS	training and validation		10		<b>0.75</b>	<b>4.06</b>	<b>5.53</b>	<b>0.00</b>	<b>0.00</b>	<b>0.50</b>	<b>very good</b>
		test			<b>0.66</b>	<b>5.54</b>	<b>6.67</b>	<b>-2.80</b>	<b>-0.73</b>	<b>0.58</b>	<b>good</b>	

			number PC	number LV	total variance	R <sup>2</sup> /NSE	MAE	RMSE	MBE	PBIAS	RSR	model performance
<b>Distillation (10%)</b>	PCR	training and validation	12		98.1%	0.24	6.00	7.78	0.00	0.00	0.87	unsatisfactory
		test				0.07	7.21	9.00	-0.51	-0.13	0.96	unsatisfactory
	ANN	training and validation				0.76	2.83	4.38	0.41	0.10	0.49	very good
		test				0.18	6.66	8.47	-1.46	-0.37	0.91	unsatisfactory
PLS	training and validation		10		<b>0.78</b>	<b>3.09</b>	<b>4.21</b>	<b>0.00</b>	<b>0.00</b>	<b>0.47</b>	<b>very good</b>	
	test				<b>0.66</b>	<b>4.65</b>	<b>5.62</b>	<b>-0.20</b>	<b>-0.05</b>	<b>0.60</b>	<b>good</b>	
<b>Distillation (50%)</b>	PCR	training and validation	13		98.2%	0.35	3.52	4.52	0.00	0.00	0.80	unsatisfactory
		test				0.41	3.58	4.40	0.91	0.20	0.77	unsatisfactory
	ANN	training and validation				<b>0.66</b>	<b>2.73</b>	<b>3.44</b>	<b>0.23</b>	<b>0.05</b>	<b>0.61</b>	<b>good</b>
		test				<b>0.72</b>	<b>2.58</b>	<b>3.00</b>	<b>1.00</b>	<b>0.22</b>	<b>0.52</b>	<b>good</b>
PLS	training and validation		13		0.90	1.41	1.82	0.00	0.00	0.32	very good	
	test				0.54	2.95	3.86	1.08	0.24	0.68	satisfactory	
<b>Distillation (90%)</b>	PCR	training and validation	12		98.1%	0.25	4.78	6.70	0.00	0.00	0.86	unsatisfactory
		test				0.20	5.23	7.01	0.81	0.15	0.89	unsatisfactory
	ANN	training and validation				0.58	3.81	5.02	0.25	0.05	0.64	satisfactory
		test				0.50	4.32	5.55	-0.40	-0.08	0.71	unsatisfactory
PLS	training and validation		12		<b>0.75</b>	<b>2.92</b>	<b>3.85</b>	<b>0.00</b>	<b>0.00</b>	<b>0.50</b>	<b>good</b>	
	test				0.48	4.44	5.67	0.25	0.05	0.72	unsatisfactory	
<b>Distillation (95%)</b>	PCR	training and validation	13		98.4%	0.18	5.98	8.38	0.00	0.00	0.91	unsatisfactory
		test				-0.22	7.52	10.3	1.33	0.24	1.11	unsatisfactory
	ANN	training and validation				0.51	4.75	6.45	-0.14	-0.03	0.70	satisfactory
		test				-0.07	7.84	9.64	0.25	0.05	1.03	unsatisfactory
PLS	training and validation		11		<b>0.71</b>	<b>3.74</b>	<b>4.94</b>	<b>0.00</b>	<b>0.00</b>	<b>0.53</b>	<b>good</b>	
	test				0.05	7.51	9.08	0.79	0.15	0.97	unsatisfactory	

### 6.3. LVGO, HVGO and Fuel Oil – Chemical Composition

Although all the potentialities and advantages, previously demonstrated, of  $^1\text{H}$  NMR spectroscopy to determine the properties of the different fractions analysed, LVGO, HVGO and fuel oil, the ability of  $^1\text{H}$  NMR spectroscopy to analyse the chemical composition of such heavy fractions was assessed. The analysis of the molecular composition of these fractions could be considered a challenge once these heavy fractions, produced by conversion of residues, were very complex, containing hundreds of individual hydrocarbons varying in size and molecular structure and functional groups. Moreover, the large amount of very similar molecules present in each fraction could lead to signals overlap. Due to the overlap that characterized the NMR peaks many of the structural detail information that the NMR spectrum might give may be lost and the identification of some molecules impossible. However, some still are visible and NMR spectroscopy have been demonstrated as a paramount technique for the determination of the: i) aromatic and aliphatic content; ii) aliphatic  $\text{CH}_3$ ,  $\text{CH}_2$ ,  $\text{CH}$  and quaternary C; iii) aromatic  $\text{CH}$  and quaternary C; iv) H and C next to aromatic rings and farther away; and v) long aliphatic chains. With this, in order to analyse the chemical composition of such fractions besides the use of  $^1\text{H}$  NMR spectroscopy, that contained the molecular structure of all hydrocarbons contained in a petroleum cut, the quantitative  $^{13}\text{C}$  NMR spectroscopy was also used. The idea of using the quantitative  $^{13}\text{C}$  NMR spectroscopy was to obtain more detail information of the molecular structure. Table 6.32 and Table 6.33 present the assignment of bands in the  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra, respectively. These chemical shift regions were selected taking into account some published works in the study of petroleum and derivatives (Altgelt *et al.*, 1994; Molina V *et al.*, 2007; Behera *et al.*, 2008; Process NMR Associates LLC, 2012a; Process NMR Associates LLC, 2012b) and confirmed by the 2D spectra obtained when analysing a petroleum fraction.

**Table 6.32:** Comparison between LVGO, HVGO and fuel oil by using  $^1\text{H}$  NMR spectroscopy.

	Chemical shift range (ppm)	Type of proton	Quantification (%)		
			LVGO	HVGO	Fuel Oil
Aliphatic region	0.5 – 1.0	H in $\gamma$ -CH <sub>3</sub> or farther from an aromatic ring; CH <sub>3</sub> in paraffins	29.4	19.1	18.8
	1.0 – 1.9	H in $\beta$ -CH <sub>3</sub> , $\beta$ -CH <sub>2</sub> and $\beta$ -CH			
	1.0 – 1.6	H in $\beta$ -CH <sub>2</sub> and $\beta$ -CH <sub>3</sub> and farther from an aromatic ring; CH and CH <sub>2</sub> in paraffins	51.3	55.6	47.6
	1.6 – 1.9	H in $\beta$ -CH <sub>2</sub> and $\beta$ -CH and farther from an aromatic ring	5.6	8.5	6.2
	1.9 – 3.4	$\alpha$ -CH <sub>3</sub> , $\alpha$ -CH <sub>2</sub> and $\alpha$ -CH			
	1.9 – 2.25	H in $\alpha$ -CH <sub>3</sub>	2.7	3.7	4.4
	2.2 – 2.8	H in $\alpha$ -CH <sub>2</sub>	5.6	5.7	9.8
	2.8 – 3.4	H in $\alpha$ -CH <sub>2</sub> and $\alpha$ -CH	0.9	2.8	3.5
	3.4 – 4.5	H in CH <sub>2</sub> and CH $\alpha$ to aromatic rings	0.0	0.0	0.8
Aliphatics	0.5 – 4.5	aliphatic H	95.5	95.4	91.1
Aromatic region	6.5 – 7.3	monoaromatic H	2.7	1.8	3.5
	7.3 – 7.8	diaromatic H	1.8	1.8	3.5
	7.8 – 10.0	triaromatic H	0.0	1.0	2.6
Aromatics	6.5 – 10.0	aromatic H	4.5	4.6	9.6



**Table 6.33:** Comparison between LVGO, HVGO and fuel oil by using  $^{13}\text{C}$  NMR spectroscopy.

	Chemical shift range (ppm)	Type of carbon	Quantification (%)		
			LVGO	HVGO	Fuel Oil
Aliphatic region	11.0 -15.0	$\gamma$ -CH <sub>3</sub> or farther from an aromatic ring	5.9	5.0	3.9
	15.0 – 18.0	$\beta$ -CH <sub>3</sub> to an aromatic ring	1.7	2.5	1.3
	18.0 – 20.5	$\alpha$ -CH <sub>3</sub> shielded by one adjacent ring or group; some CH <sub>3</sub> $\alpha$ hydroaromatic and naphthenic CH <sub>2</sub>	5.9	6.7	4.5
	20.5 – 22.5	CH <sub>3</sub> not shielded by adjacent groups; some CH <sub>3</sub> $\alpha$ hydroaromatic and naphthenic CH <sub>2</sub>	7.6	6.7	5.2
	22.5 – 24.0	$\gamma$ -CH <sub>2</sub> and farther adjacent to terminal CH <sub>3</sub> ; $\beta$ -CH <sub>2</sub> in unsubstituted tetralin structures; some CH <sub>3</sub> $\alpha$ hydroaromatic and naphthenic CH <sub>2</sub>	2.5	2.5	1.3
	24.0 – 27.5	some CH <sub>2</sub> naphthenic; $\alpha$ -CH not shielded; $\beta$ -CH <sub>2</sub> in propyl side chains; $\beta$ -CH <sub>3</sub> in isopropyl side chains	6.8	7.5	5.2
	27.5 – 37.0	CH <sub>2</sub> not adjacent to CH in alkyl groups; CH <sub>2</sub> adjacent to alkyl CH in some $\alpha$ -CH <sub>2</sub> and CH <sub>2</sub> adjacent to terminal CH <sub>3</sub> in alkyl substituents with more than four carbons; some CH <sub>2</sub> naphthenic; some ring joining methylene (32 – 43 ppm)	38.3	36.6	31.2
	37.0 – 60.0	internal (bridgehead) naphthenic C or CH; CH in alkyl side chains (not isopropyl or isobutyl); CH <sub>2</sub> in alkyl side chains adjacent to CH; some ring joining methylene (32 – 43 ppm)	16.1	15.8	12.3
Aliphatics	11.0 – 60.0	aliphatic region	84.8	83.3	64.9
Aromatic region	100.0 – 115.0	some olefinic (other spread through aromatic region)	0.0	0.0	1.9
	115.0 – 129.5	protonated aromatic carbon atoms	9.3	7.5	18.2
	129.5 – 138.0	bridgehead aromatic carbon atoms	4.2	5.0	9.8
	138.0 – 160.0	substituted aromatic carbon atoms; heteroatom (N, O, S) aromatic	1.7	4.2	5.2
Aromatics	100.0 – 160.0	aromatic region	15.2	16.7	35.1

Taking into account the information given in Table 6.32, it was verified that, in general, the results obtained from the analysis of the LVGO and HVGO spectra could not lead to any conclusion relatively to the aliphatic and aromatic hydrogen content, since similar results of the aliphatic (0.5 – 4.5 ppm) and aromatic (6.5 – 10.0 ppm) hydrogens content were obtained. As would be expected, higher differences were notable when comparing fuel oil with LVGO and HVGO samples. The aromatic content of the fuel oil samples were higher, with the fuel oil samples richer in aromatic protons and having higher quantity of protons that were present in mono-, di- and triaromatic structures. It was verified that the strong increase in aromatic hydrogens of the fuel oil samples, related with the increase of the aromatic ring number, was accompanied with an increase in the hydrogen's  $\alpha$  and a decrease in the hydrogen's  $\beta$  and  $\gamma$ , compared with the quantities of these molecules in LVGO and HVGO samples. This behaviour was an indicative that fuel oil samples were richer in aromatic compounds and poor in aliphatic structures when compared with LVGO and HVGO samples. These results, obtained by  $^1\text{H}$  NMR spectroscopy, were confirmed by the analysis of the quantitative  $^{13}\text{C}$  NMR spectra. As visualized in Table 6.33, fuel oil samples were characterized for presenting higher quantity of aromatic carbon content and smaller percentage of aliphatic carbon content, when compared with LVGO and HVGO samples. The other most pronounced differences were related with the presence of some olefinic compounds (100.0 – 115.0 ppm) and the higher percentage at the 138.0 to 160.0 ppm at the  $^{13}\text{C}$  NMR spectra. Olefins are rarely found in crude oils, the reason for which any evidence of these compounds were found in the  $^{13}\text{C}$  NMR spectrum of LVGO and HVGO samples, however they may be present in substantial amounts in cracked refinery streams. Since fuel oil was produced at the visbreaker unit and since during the visbreaking process there are some thermal cracking process occurring it is possible that, depending on the severity conditions, some olefinic molecules were formed, as visualized in the  $^{13}\text{C}$  NMR spectra. Relatively to the higher percentage at the 138.0 to 160.0 ppm, this could be related with the higher quantity of compounds bearing heteroatoms in the heavier fractions, such as fuel oil. The same occurs when comparing this region (138.0 – 160.0 ppm) of the LVGO and HVGO spectra. It was known that HVGO samples presented higher quantity of compounds with heteroatoms when compared with LVGO samples, thus explaining the higher percentage of substituted aromatic carbon atoms at the HVGO samples.

When comparing the  $^1\text{H}$  NMR spectra of LVGO and HVGO samples, even though no conclusions were made relatively to the content of aromatic and aliphatic compounds, it was possible to see that LVGO samples were richer in  $\gamma$ -methyl hydrogen or farther away from aromatic rings, represented by the band at 0.5 to 1.0 ppm. These results were indicative that the LVGO samples were, probably, richer in long alkyl chains suggesting that smaller quantity of aromatic compounds were found. Relatively to the aromatic region (6.5 – 10.0 ppm), the results indicated that LVGO samples were richer in protons attached to single rings (6.5 – 7.3 ppm) while HVGO samples had protons attached to single-ring and multiring aromatics. With this information, it was concluded that HVGO samples were richer in heavier aromatic compounds. The same conclusions were taken from the quantitative  $^{13}\text{C}$  NMR spectra. Moreover, with the quantitative  $^{13}\text{C}$  NMR spectra it was confirmed that LVGO samples presented long chains indicating higher quantity of longer paraffinic structures. The absorption of 11.0 to 15.0 ppm, which indicated  $\gamma$ - $\text{CH}_3$  groups or farther away from an aromatic ring, allowed us to conclude relatively to the length of the alkyl chains. In addition, the region at 27.5 to 37.0 ppm, also suggested longer chains of LVGO samples once this band represents  $\text{CH}_2$  groups from aliphatic compounds. Furthermore, the higher results at 20.5 to 22.5 and 37.0 to 60.0 also indicated that there was a contribution of long chains. At the aromatic region of the  $^{13}\text{C}$  NMR spectra, the higher percentage at 115.0 to 129.5 ppm suggested that LVGO samples presented more protonated aromatics than HVGO samples indicating that the aromatic compounds in the LVGO samples were farther away from each other. Suggesting, once again, that HVGO samples were composed with heavier aromatic compounds. Besides the heavier percentage of aromatic compounds, it was verified that the HVGO samples presented higher quantity of carbons attached to naphthenic compounds than LVGO samples, due to the higher results at 18.0 to 20.5 and 24.0 to 27.5 ppm.

NMR, especially the combination of  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectroscopy, demonstrated, once again, to be a paramount technique in the structural characterization of such complex mixtures as LVGO, HVGO and fuel oil. The structural analysis of these spectra demonstrated to be very useful for the knowledge of the chemical composition with the identification of structural groups, such as, the paraffinic, naphthenic and aromatic compounds. The knowledge of the chemical composition can be very useful for the comprehension of the influence of the nature of these petroleum fractions in the refinery

operations. For example, the nature of HVGO, an intermediate used as feedstock of the hydrocracking unit, influences the course of the reaction occurring during hydrocracking. Moreover, the chemical composition of LVGO, a portion of gas oil that is recovered at the vacuum unit and mixed with gas oil from the atmospheric unit, is also important once the presence of higher quantity of aromatic compounds can affect the quality of gas oil, decreasing its cetane number, and cause environmental problems.

# 7

## CONCLUSIONS AND FUTURE WORK

It is presented in this chapter the final conclusions taken from the developed work as well as some suggestions for future work. In the developed work, very important issues in the field of petroleum refining were presented. These important issues were related to the use of  $^1\text{H}$  NMR spectroscopy combined with multivariate methods for predicting physical-chemical properties of heavy petroleum fractions.

## **7. CONCLUSIONS AND FUTURE WORK**

7.1. General Conclusions.....	289
7.2. Suggestions for Future Work.....	293

## 7.1. General Conclusions

The characterization of light vacuum gas oil, heavy vacuum gas oil and fuel oil by proton nuclear magnetic resonance spectroscopy was very important in the determination of significant physical-chemical properties for the complete assessment of their quality and for process control. With this work, it was demonstrated that combining proton nuclear magnetic resonance spectroscopy with different multivariate methods (principal component regression, partial least squares and artificial neural networks) it was possible to develop multivariate calibration models that had a high potential to predict the properties of such petroleum fractions, even though these fractions were characterized as the heavy ends of crude oils. Kinematic viscosity, density, carbon residue, P-value, sulfur content, distillation and flash point were example of some of the properties analysed and most of them where determined with good precision and accuracy.

Nuclear magnetic resonance spectroscopy has been suggested as one of the possible techniques for industrial analysis and has already been used in some refinery laboratories. As found in the literature, nuclear magnetic resonance spectroscopy can offer precise analysis in real-time and control the properties of the crude from the feed to the final product. The greatest challenge is to verify the capability of a nuclear magnetic resonance spectrometer to monitor some properties and for process control. This is the motivation of this work, to verify if a proton nuclear magnetic resonance spectrometer can be used to predict some heavy petroleum properties and consequently evaluate the possibility of acquiring an online nuclear magnetic resonance spectrometer.

To take advantage from all nuclear magnetic resonance benefits and to exploit the quantitative aspects of nuclear magnetic resonance spectroscopy some pre-processing and data manipulations of the nuclear magnetic resonance spectra were necessary to be performed. Phase and baseline correction, chemical shift alignment, data reduction, binning, normalization and intensity scaling were the main pre-processing steps used. All these different types of data pre-processing were considered important and necessary, due to the increased complexity and dimensionality of the nuclear magnetic resonance data, to: i) exploit the information content of the nuclear magnetic resonance data; ii) increase the prospect of detecting subtle changes in the nuclear magnetic resonance data; and iii) enhance the interpretability of the resulting multivariate models. From other approaches

found in the literature, these data manipulations were considered the most adequate and the ones that mostly contributed to significantly benefit of the analysis of nuclear magnetic resonance data. It was also important to have in attention that for obtaining a good quantitative spectrum some experimental factors need to be controlled and respected. Attention was also given to the sample preparation, data acquisition, pulse sequence and instrumental parameters.

The impact of the pre-processing procedure, used before and after obtaining the spectrum, demonstrated to be very important for further application of the multivariate data methods. Principal component regression, partial least squares and artificial neural networks were the multivariate methods used to predict the physical-chemical properties of the petroleum fractions. The use of these multivariate methods proved to be very advantageous in the characterization of the petroleum fractions under analysis and enable the rapid computation of the physical-chemical properties prediction models.

For the light vacuum gas oil samples, the models were developed to predict the density and the distillation at atmospheric pressure. For the heavy vacuum gas oil samples the same type of models were used to determine the kinematic viscosity, density, carbon residue and the distillation at reduced pressure. Finally, to predict the kinematic viscosity, density, carbon residue, P-value, sulfur content and the flash point of the fuel oil samples more models were also developed. The development of all these models were only possible after the samples underwent a characterization by following some standard methods and by using the proton nuclear magnetic resonance spectroscopy. Hundreds of samples of the different streams were analysed with the standard methods and proton nuclear magnetic resonance spectroscopy and consequently a good data base was created. It was in this data base that the different methods were applied and reliable models obtained.

Comparing the results obtained for the different petroleum fractions analysed, it was notorious that the best results were obtained for all light vacuum gas oil properties due to its lower complexity, when compared with heavy vacuum gas oil and fuel oil samples.

When analysing light vacuum gas oil samples, proton nuclear magnetic resonance spectroscopy proved to have a higher potential to predict the density and the atmospheric distillation, the two most important properties used to evaluate light vacuum gas oil



quality. The results showed that the use of artificial neural networks model led to more accurate results, in almost all developed models. Very good predictions were obtained for both data sets, training and validation data sets and test data set, and for all properties analysed.

For the heavy vacuum gas oil samples, it was possible to conclude that proton nuclear magnetic resonance spectroscopy had high potential to predict heavy vacuum gas oil properties, such as density, kinematic viscosity and carbon residue. The higher problem occurs when predicting the distillation of the heavy vacuum gas oil samples. Good predictions were obtained when analysing the first distillation cuts (5, 10 and 50%), however, worse performances were achieved when predicting the 90 and 95% distillation cuts, especially for the test data set. The distillation of the heaviest fractions and the limitations of the standard method used could contribute to the less accurate models achieved when predicting the 90 and 95% distillation cuts. Although, more information should be given to the developed models, suggesting that increasing the number of samples, especially for the test data set, a further improvement in the accuracy could be obtained.

When analysing fuel oil samples, it was demonstrated that proton nuclear magnetic resonance spectroscopy had high potential to predict almost all properties analysed. Accurate results were obtained when predicting the kinematic viscosity, density, sulfur content, carbon residue and flash point. The most difficult property to predict was the P-value. Besides the good predictions obtained for the training and validation data sets, only satisfactory predictions were achieved for the test data set. However, it is expected that increasing the data set, i.e., the number of samples analysed, a further improvement in accuracy and robustness of the models could be obtained.

When comparing the applied multivariate models, it was demonstrated that, better predictive performance were obtained, for both data sets, when using the artificial neural networks and the partial least squares models. Worst predictions were achieved for all properties analysed when using the principal component regression model. It was with the artificial neural networks model that better predictions were obtained, due to its ability to develop a model around non-linear relationships between the measured properties and the predicted results, and it was with the linear predictive model, principal component regression, that worst performance indexes were achieved. The partial least squares model

also demonstrated to be a good model giving normally better predictions for the training and validation data sets. It was with the artificial neural networks model that most reliable predictions when determining the kinematic viscosity at both temperatures, 100 and 50 °C, the carbon residue, the flash point, the P-value and the density of fuel oil samples were achieved. Moreover, for the light vacuum gas oil samples, the artificial neural networks also demonstrated higher quality models when predicting the density and the 5, 10 and 90% distillation cuts. In addition, the artificial neural networks was more reliable to predict the density and 50% distillation cut of heavy vacuum gas oil samples. On the other hand, the partial least squares model confirmed great capabilities to predict the sulfur content of fuel oil samples, the 50 and 90% distillation cuts of light vacuum gas oil samples and the kinematic viscosity, carbon residue and the 5 and 10% distillation cuts of heavy vacuum gas oil samples.

The results obtained demonstrated that proton nuclear magnetic resonance spectroscopy in combination with multivariate data methods is a very important technique that can be used to predict the properties of light (light vacuum gas oil) or heavy fractions (heavy vacuum gas oil and fuel oil) with high accuracy and precision. In addition, using the proton nuclear magnetic resonance spectroscopy and the developed models will contribute to quantify all properties analysed, almost replacing the standards methods and avoiding the tedious sample preparations used every day.

Proton nuclear magnetic resonance spectroscopy has proven to be a very important technique in the study of petroleum fractions and can become an alternative to the traditional time-consuming laboratory methods. Using a nuclear magnetic resonance spectrometer, many benefits are expected to be obtained. It demonstrated to be a rapid and non-invasive procedure, offering improvement in the analysis, reduction in the number of analysis and unique ability to provide all control properties with just one analysis. This leads us to conclude that it will be advantageous to have an online nuclear magnetic resonance spectrometer to determine such properties.

## 7.2. Suggestions for Future Work

Concerning the adequate results obtained when using  $^1\text{H}$  NMR spectroscopy combined with multivariate data methods to predict some properties of LVGO, HVGO and fuel oil streams, it will be interesting to develop an extensive work with an increase number of samples and of properties to demonstrate the interesting and benefits of nuclear magnetic resonance spectroscopy.

Since the final results depend on the pre-measuring, pre-processing and data manipulations of the NMR spectra it will be a good practice to verify the effect of other approaches in the final developed models. It will be interesting to evaluate the use of other solvents during the preparation of the samples for NMR analysis, such as, the use of an internal or external standard. In addition, it will also be interesting to select which information of the spectrum should be used depending on the property in analysis.

Besides PCR, PLS and ANN models, other models, such those generated by genetic program and based on fuzzy logic should be employed to explore the potentialities of  $^1\text{H}$  NMR spectroscopy in predicting the physical-chemical properties. In addition, other methodologies to divide the information into the three different data sets, training, validation and test data sets, should be also tried. Moreover, other approaches of the models used should be explored, such as, the selection of other inputs, instead of using all selected PCs, when applying the ANN model.

To improve the performance indexes obtained with the developed models it will be advisable to increase the data base with more quantity of samples analysed. This will contribute to increase the precision and accuracy of the developed models to predict all properties under analysis. In addition, it will be required a continuous maintenance and upgrade of these developed models to ensure that the models continues to predict with high precision the required properties. This ensures that even occurring significant changes in the process, the models will be prepared to respond with a good accuracy.

To verify all results obtained it will be useful to implement an online  $^1\text{H}$  NMR spectrometer to characterize such properties of light vacuum gas oil, heavy vacuum gas oil and fuel oil streams and to develop the multivariate models based on this information.



# 8

## **REFERENCES**



- Abdi, H., **2003**. *Partial Least Squares (PLS) Regression*. In: Encyclopedia of Social Sciences Research Methods (eds. M. Lewis-Beck, A. Bryman and T. Futing), 1-7. Thousand Oaks: Sage.
- Abdi, H., **2010**. Partial least squares regression and projection on latent structure regression (PLS Regression). *Wiley Interdisciplinary Reviews: Computational Statistics* 2 (1), 97-106.
- Ahmadjian, M., Brown, C.W., **1976**. Petroleum identification by laser Raman spectroscopy. *Analytical Chemistry* 48 (8), 1257-1259.
- Ahmadloo, F., Asghari, K., Renouf, G., **2010**. Performance Prediction of Waterflooding in Western Canadian Heavy Oil Reservoirs Using Artificial Neural Network. *Energy & Fuels* 24 (4), 2520-2526.
- Alam, T.M., Alam, M.K., **2004**. Chemometric Analysis of NMR Spectroscopy Data: A Review. *Annual Reports on NMR Spectroscopy* 54, 41-80.
- Ali, F.A., Ghaloum, N., Hauser, A., **2005**. Structure Representation of Asphaltene GPC Fractions Derived from Kuwaiti Residual Oils. *Energy & Fuels* 20 (1), 231-238.
- Altgelt, K.H., Boduszynski, M.M., **1994**. *Composition and Analysis of Heavy Petroleum Fractions*, first edition, Marcel Dekker, New York.
- Aske, N., Kallevik, H., Johnsen, E.E., Sjöblom, J., **2002**. Asphaltene Aggregation from Crude Oils and Model Systems Studied by High-Pressure NIR Spectroscopy. *Energy & Fuels* 16 (5), 1287-1295.
- ASTM Standards D93, **2011**. *Standard Test Methods for Flash Point by Pensky-Martens Closed Cup Tester*, ASTM International, West Conshohocken, PA.
- ASTM Standards D445, **2009**. *Standard Test Method for Kinematic Viscosity of Transparent and Opaque Liquids (and Calculation of Dynamic Viscosity)*, ASTM International, West Conshohocken, PA.
- ASTM Standards D1160, **2006**. *Standard Test Method for Distillation of Petroleum at Reduced Pressure*, ASTM International, West Conshohocken, PA.

ASTM Standards D2163, **2007**. *Standard Test Method for Determination of Hydrocarbons in Liquefied Petroleum (LP) Gases and Propane/Propene Mixtures by Gas Chromatography*, ASTM International, West Conshohocken, PA.

ASTM Standards D2427, **2011**. *Standard Test Method for Determination of C<sub>2</sub> through C<sub>5</sub> Hydrocarbons in Gasolines by Gas Chromatography*, ASTM International, West Conshohocken, PA.

ASTM Standards D2887, **2013**. *Standard Test Method for Boiling Range Distribution of Petroleum Fractions by Gas Chromatography*, ASTM International, West Conshohocken, PA.

ASTM Standards D3606, **2010**. *Standard Test Method for Determination of Benzene and Toluene in Finished Motor and Aviation Gasoline by Gas Chromatography*, ASTM International, West Conshohocken, PA.

ASTM Standards D4052, **2011**. *Standard Test Method for Density, Relative Density, and API Gravity of Liquids by Digital Density Meter*, ASTM International, West Conshohocken, PA.

ASTM Standards D4053, **2009**. *Standard Test Method for Benzene in Motor and Aviation Gasoline by Infrared Spectroscopy*, ASTM International, West Conshohocken, PA.

ASTM Standards D4530, **2007**. *Standard Test Method for Determination of Carbon Residue (Micro Method)*, ASTM International, West Conshohocken, PA.

ASTM Standards D5002, **2005**. *Standard Test Method for Density and Relative Density of Crude Oils by Digital Density Analyzer*, ASTM International, West Conshohocken, PA.

ASTM Standards D5292, **1999**. *Standard Test Method for Aromatic Carbon Contents of Hydrocarbon Oils by High Resolution Nuclear Magnetic Resonance Spectroscopy*, ASTM International, West Conshohocken, PA.

ASTM Standards D5307, **2007**. *Standard Test Method for Determination of Boiling Range Distribution of Crude Petroleum by Gas Chromatography*, ASTM International, West Conshohocken, PA.



- ASTM Standards D7171, **2005**. *Standard Test Method for Hydrogen Content of Middle Distillate Petroleum Products by Low-Resolution Pulsed Nuclear Magnetic Resonance Spectroscopy*, ASTM International, West Conshohocken, PA.
- Bakeev, K.A., **2010**. *Process Analytical Technology*, second edition, John Wiley & Sons, United Kingdom.
- Bansal, V., Kapur, G.S., Sarpal, A.S., Kagdiyal, V., Jain, S.K., Srivastava, S.P., Bhatnagar, A.K., **1998**. Estimation of Total Aromatics and Their Distribution as Mono and Global Di-Plus Aromatics in Diesel-Range Products by NMR Spectroscopy. *Energy & Fuels* 12 (6), 1223-1227.
- Barman, B.N., **2005**. Characterization of Feeds, Intermediates, and Products from Heavy Oil Processes by High-Temperature Simulated Distillation and Thin-Layer Chromatography with Flame Ionization Detection. *Energy & Fuels* 19 (5), 1995-2000.
- Becker, E.D., **2000**. *High Resolution NMR. Theory and Chemical Applications*, third, Academic Press, London.
- Beens, J., Brinkman, U.A.T., **2000**. The role of gas chromatography in compositional analyses in the petroleum industry. *Trends in Analytical Chemistry* 19 (4), 260-275.
- Behera, B., Ray, S.S., Singh, I.D., **2008**. Structural characterization of FCC feeds from Indian refineries by NMR spectroscopy. *Fuel* 87 (10-11), 2322-2333.
- Boduszynski, M.M., **1987**. Composition of heavy petroleums. 1. Molecular weight, hydrogen deficiency, and heteroatom concentration as a function of atmospheric equivalent boiling point up to 1400 °F (760 °C). *Energy & Fuels* 1 (1), 2-11.
- Boduszynski, M.M., **1988**. Composition of heavy petroleums. 2. Molecular characterization. *Energy & Fuels* 2 (5), 597-613.
- Butt, J.A., Duckworth, D.F., Perry, S.G., **1986**. *Characterization of spilled oil samples*, John Wiley & Sons, London.

- Calemma, V., Iwanski, P., Nali, M., Scotti, R., Montanari, L., **1995**. Structural Characterization of Asphaltenes of Different Origins. *Energy & Fuels* 9 (2), 225-230.
- Çamdevýren, H., Demýr, N., Kanik, A., Keskýn, S., **2005**. Use of principal component scores in multiple linear regression models for prediction of Chlorophyll-a in reservoirs. *Ecological Modelling* 181 (4), 581-589.
- Chaloulakou, A., Grivas, G., Spyrellis, N., **2003**. Neural Network and Multiple Regression Models for PM10 Prediction in Athens: A Comparative Assessment. *Journal of the Air & Waste Management Association* 53 (10), 1183-1190.
- Christian, G.D., O'Reilly, J.E., **1986**. *Instrumental Analysis*, second edition, Allyn and Bacon, Boston.
- Claridge, T.D.W., **1999**. *High Resolution NMR Techniques in Organic Chemistry*, Elsevier Science, Oxford.
- Cloarec, O., Dumas, M.E., Trygg, J., Craig, A., Barton, R.H., Lindon, J.C., Nicholson, J.K., Holmes, E., **2004**. Evaluation of the orthogonal projection on latent structure model limitations caused by chemical shift variability and improved visualization of biomarker changes in <sup>1</sup>H NMR spectroscopic metabonomic studies. *Analytical Chemistry* 77 (2), 517-526.
- Coleman, S., Greenfield, T., Stewardson, D., Montgomery, D.C., **2008**. *Statistical Practice in Business and Industry*, John Wiley & Sons, England.
- Cookson, D.J., Smith, B.E., **1985**. Determination of structural characteristics of saturates from diesel and kerosine fuels by carbon-13 nuclear magnetic resonance spectrometry. *Analytical Chemistry* 57 (4), 864-871.
- Cookson, D.J., Smith, B.E., **1987**. One and two-dimensional NMR methods for elucidating structural characteristics of aromatic fractions from petroleum and synthetic fuels. *Energy & Fuels* 1 (1), 111-120.
- Cookson, D.J., Smith, B.E., **1990**. Calculation of jet and diesel fuel properties using carbon-13 NMR spectroscopy. *Energy & Fuels* 4 (2), 152-156.

- Cooper, J.B., Wise, K.L., Groves, J., Welch, W.T., **1995**. Determination of Octane Numbers and Reid Vapor Pressure of Commercial Petroleum Fuels Using FT-Raman Spectroscopy and Partial Least-Squares Regression Analysis. *Analytical Chemistry* 67 (22), 4096-4100.
- Craig, A., Cloarec, O., Holmes, E., Nicholson, J.K., Lindon, J.C., **2006**. Scaling and normalization effects in NMR spectroscopic metabonomic data sets. *Analytical Chemistry* 78 (7), 2262-2267.
- Curry, B., Rumelhart, D.E., **1990**. MSnet: A Neural Network which Classifies Mass Spectra. *Tetrahedron Computer Methodology* 3 (3-4), 213-237.
- de Andrade, D.F., Fernandes, D.R., Miranda, J.L., **2010**. Methods for the determination of conjugated dienes in petroleum products: A review. *Fuel* 89 (8), 1796-1805.
- de Peinder, P., Visser, T., Petrauskas, D.D., Salvatori, F., Soulimani, F., Weckhuysen, B.M., **2009**. Partial least squares modeling of combined infrared,  $^1\text{H}$  NMR and  $^{13}\text{C}$  NMR spectra to predict long residue properties of crude oils. *Vibrational Spectroscopy* 51 (2), 205-212.
- Decroocq, D., **1997**. Major Scientific and Technical Challenges about Development of New Processes in Refining and Petrochemistry. *Revue de L'Institut Français Du Pétrole* 52 (5), 469-489.
- Dente, M., Bozzano, G., Bussani, G., **1997**. A comprehensive program for visbreaking simulation: product amounts and their properties prediction. *Computers & Chemical Engineering* 21 (10), 1125-1134.
- Doan, B.T., Gillet, B., Blondel, B., Beloeil, J.C., **1995**. Analysis of polyaromatics in crude gas oil mixtures: a new strategy using  $^1\text{H}$  2D NMR. *Fuel* 74 (12), 1806-1811.
- Douda, J., Alvarez, R., Navarrete Bolaños, J., **2008**. Characterization of Maya Asphaltene and Maltene by Means of Pyrolysis Application. *Energy & Fuels* 22 (4), 2619-2628.
- Durand, E., Clemancey, M., Quoineaud, A.-A., Verstraete, J., Espinat, D., Lancelin, J.-M., **2008**.  $^1\text{H}$  Diffusion-Ordered Spectroscopy (DOSY) Nuclear Magnetic Resonance

(NMR) as a Powerful Tool for the Analysis of Hydrocarbon Mixtures and Asphaltenes. *Energy & Fuels* 22 (4), 2604-2610.

Eriksson, L., Johansson, E., Kettaneh-Wold, N., Trygg, J., Wikstrom, C., Wold, S., **2006**. *Multi- and Megavariate Data Analysis Part I: Basic Principles and Applications*, second, Umetrics AB, Sweden.

European Standard EN 14078, **2009**. *Liquid petroleum products - Determination of fatty acid methyl ester (FAME) content in middle distillates - Infrared spectrometry method*, European Committee for Standardization, Brussels.

Exarchou, V., Troganis, A., Gerothanassis, I.P., Tsimidou, M., Boskou, D., **2000**. Identification and Quantification of Caffeic and Rosmarinic Acid in Complex Plant Extracts by the Use of Variable-Temperature Two-Dimensional Nuclear Magnetic Resonance Spectroscopy. *Journal of Agricultural and Food Chemistry* 49 (1), 2-8.

Fernández Pierna, J.A., Abbas, O., Baeten, V., Dardenne, P., **2009**. A Backward Variable Selection method for PLS regression (BVSPLS). *Analytica Chimica Acta* 642 (1-2), 89-93.

Fish, R.H., Komlenic, J.J., Wines, B.K., **1984**. Characterization and comparison of vanadyl and nickel compounds in heavy crude petroleum and asphaltenes by reverse phase and size-exclusion liquid chromatography/graphite furnace atomic absorption spectrometry. *Analytical Chemistry* 56, 2452-2460.

Flumignan, D.L., Sequinel, R., Hatanaka, R.R., Boralle, N., Oliveira, J.E.d., **2012**. Multivariate calibrations on <sup>1</sup>H NMR profiles for prediction of physicochemical parameters of Brazilian commercial gasoline. *Fuel* 99 (0), 180-187.

Foddi, O., Grosso, M., Tronci, S., Baratti, R., **2009**. *Prediction of Distillate Yields in a Visbreaking Process by Neural Networks*, Scientific Literature Digital Library and Search Engine, United States.

FORD Standards EU-AJ 051-01, **2001**. *Method of Carbon Type Analysis of Lubricating Oils by Infrared Spectroscopy*, Industry Standards & Regulations, Englewood.

- Forshed, J., Torgrip, R.J.O., Åberg, K.M., Karlberg, B., Lindberg, J., Jacobsson, S.P., **2005**. A comparison of methods for alignment of NMR peaks in the context of cluster analysis. *Journal of Pharmaceutical and Biomedical Analysis* 38 (5), 824-832.
- Fraser, L.-A., Mulholland, D.A., Fraser, D.D., **1997**. Classification of limonoids and protolimonoids using neural networks. *Phytochemical Analysis* 8 (6), 301-311.
- Friedel, R.A., **1959**. Absorption Spectra and Magnetic Resonance Spectra of Asphaltene. *Journal of Chemical Physics* 31 (1), 280-281.
- Galp Energia, 2010a. *Refinação & Distribuição*. [accessed on January 2011]; Available from: <http://www.galpenergia.com/PT/agalpenergia/Os-nossos-negocios/Presenca-no-mundo/Portugal/Paginas/ARL-em-Portugal.aspx>.
- Galp Energia, 2010b. *Refinaria de Matosinhos*. [accessed on January 2011]; Available from: <http://www.galpenergia.com/PT/agalpenergia/os-nossos-negocios/Refinacao-Distribuicao/ARL/Refinacao/RefinariaMatosinhos/Paginas/Refinaria-de-Matosinhos.aspx>.
- Galp Energia, **2010c**. *Refinaria de Matosinhos - Data Book de Segurança, Saúde e Ambiente*, Galp Energia, Lisboa.
- Galp Energia, **2011a**. *Fábrica de Combustíveis - Memória Descritiva*, Refinaria de Matosinhos, Matosinhos.
- Galp Energia, **2011b**. *Refinaria de Matosinhos - Data Book de Segurança, Saúde e Ambiente*, Galp Energia, Lisboa.
- Galp Energia, **2011c**. *Refinaria de Matosinhos - Diagrama Geral*, Refinaria de Matosinhos, Matosinhos.
- Gautam, K., Jin, X., Hansen, M., **1998**. Review of Spectrometric Techniques for the Characterization of Crude Oil and Petroleum Products. *Applied Spectroscopy Reviews* 33 (4), 427 - 443.

- Gershenson, C., **2003**. *Artificial Neural Networks for beginners*. Cognitive and computing sciences, University of Sussex.
- Gil, V.M.S., Geraldés, C.F.G.C., **2002**. *Ressonância Magnética Nuclear. Fundamentos, Métodos e Aplicações.*, second edition, Fundação Calouste Gulbenkian, Lisboa.
- Günther, H., **1995**. *NMR spectroscopy: basic principles, concepts, and applications in chemistry*, second edition, John Wiley & Sons, England.
- Guo, L., Sprenger, P., Garland, M., **2008**. A combination of spectral re-alignment and BTEM for the estimation of pure component NMR spectra from multi-component non-reactive and reactive systems. *Analytica Chimica Acta* 608 (1), 48-55.
- Hall, M.J., **2001**. How well does your model fit the data? *Journal of Hydroinformatics* 3 (1), 49-55.
- Hernández-Caraballo, E.A., M. Marcó-Parra, L., **2003**. Direct analysis of blood serum by total reflection X-ray fluorescence spectrometry and application of an artificial neural network approach for cancer diagnosis. *Spectrochimica Acta Part B: Atomic Spectroscopy* 58 (12), 2205-2213.
- Hirasaki, G.J., Lo, S.-W., Zhang, Y., **2003**. NMR properties of petroleum reservoir fluids. *Magnetic Resonance Imaging* 21 (3-4), 269-277.
- Hsu, C.S., Robinson, P.R., **2006**. *Practical Advances in Petroleum Processing*, Volume 1, Spring Science + Business Media, New York.
- IP Standards 336, **2004**. *Petroleum products - Determination of sulfur content - Energy-dispersive-X-ray fluorescence method*, European Committee for Standardization, Brussels.
- IP Standards 392, **1990**. *Determination of aromatic hydrogen and carbon content - High resolution nuclear magnetic resonance spectroscopy method*, Institute of Petroleum, London.

- IP Standards 499, **2011**. *Determination of aromatic carbon content of lubricant mineral base oils and middle distillate petroleum fractions - Carbon-13 nuclear magnetic resonance (NMR) spectroscopy method*, Institute of Petroleum, London.
- ISO Standards 3405, **2011**. *Petroleum products - Determination of distillation characteristics at atmospheric pressure*, International Organization for Standardization, Geneva.
- Jacobsen, N.E., **2007**. *NMR Spectroscopy Explained: Simplified Theory, Applications and Examples for Organic Chemistry and Structural Biology*, John Wiley & Sons, New Jersey.
- Jones, D.S.J., Pujadó, P.R., **2006**. *Handbook of Petroleum Processing*, Springer Science + Business Media, Netherlands.
- Joshi, J.B., Pandit, A.B., Kataria, K.L., Kulkarni, R.P., Sawarkar, A.N., Tandon, D., Ram, Y., Kumar, M.M., **2008**. Petroleum Residue Upgradation via Visbreaking: A Review. *Industrial & Engineering Chemistry Research* 47 (23), 8960-8988.
- Kapur, G.S., Chopra, A., Sarpal, A.S., **2005**. Estimation of Total Aromatic Content of Vacuum Gas Oil (VGO) Fractions (370–560 °C) by  $^1\text{H}$  NMR Spectroscopy. *Energy & Fuels* 19 (3), 1065-1071.
- Kapur, G.S., Findeisen, M., Berger, S., **2000**. Analysis of hydrocarbon mixtures by diffusion-ordered NMR spectroscopy. *Fuel* 79 (11), 1347-1351.
- Kapur, G.S., Sastry, M.I.S., Jaiswal, A.K., Sarpal, A.S., **2004**. Establishing structure–property correlations and classification of base oils using statistical techniques and artificial neural networks. *Analytica Chimica Acta* 506 (1), 57-69.
- Kenkel, J., **1994**. *Analytical Chemistry for Technicians*, second edition, CRC Press, London.
- Keun, H.C., Ebbels, T.M.D., Antti, H., Bollard, M.E., Beckonert, O., Holmes, E., Lindon, J.C., Nicholson, J.K., **2003**. Improved analysis of multivariate data by variable stability scaling: application to NMR-based metabolic profiling. *Analytica Chimica Acta* 490 (1–2), 265-276.

- Kiralj, R., Ferreira, M.M.C., **2006**. The past, present, and future of chemometrics worldwide: some etymological, linguistic, and bibliometric investigations. *Journal of Chemometrics* 20 (6-7), 247-272.
- Kontogianni, V.G., Exarchou, V., Troganis, A., Gerotheranassis, I.P., **2009**. Rapid and novel discrimination and quantification of oleanolic and ursolic acids in complex plant extracts using two-dimensional nuclear magnetic resonance spectroscopy-Comparison with HPLC methods. *Analytica Chimica Acta* 635 (2), 188-195.
- Korres, D.M., Anastopoulos, G., Lois, E., Alexandridis, A., Sarimveis, H., Bafas, G., **2002**. A neural network approach to the prediction of diesel fuel lubricity. *Fuel* 81 (10), 1243-1250.
- Kramer, K.E., Morris, R.E., Rose-Pehrsson, S.L., **2008**. Comparison of two multiplicative signal correction strategies for calibration transfer without standards. *Chemometrics and Intelligent Laboratory Systems* 92 (1), 33-43.
- Kulkarni, R.P., Pandit, A.B., Joshi, J.B., Kataria, K.L., Tandon, D., Kumar, M.M., **2010**. Visbreaking Studies in the Presence of Soaker Internals. *Industrial & Engineering Chemistry Research* 49 (22), 11221-11231.
- Kvalheim, O.M., Aksnes, D.W., Brekke, T., Eide, M.O., Sletten, E., **1985**. Crude oil characterization and correlation by principal component analysis of carbon-13 nuclear magnetic resonance spectra. *Analytical Chemistry* 57 (14), 2858-2864.
- Lee, S.W., Glavincevski, B., **1999**. NMR method for determination of aromatics in middle distillate oils. *Fuel Processing Technology* 60 (1), 81-86.
- Legates, D.R., McCabe, G.J., **1999**. Evaluating the use of “goodness-of-fit” Measures in hydrologic and hydroclimatic model validation. *Water Resources Research* 35 (1), 233-241.
- Lewis, I.A., Schommer, S.C., Hodis, B., Robb, K.A., Tonelli, M., Westler, W.M., Sussman, M.R., Markley, J.L., **2007**. Method for Determining Molar Concentrations of Metabolites in Complex Solutions from Two-Dimensional  $^1\text{H}$ - $^{13}\text{C}$  NMR Spectra. *Analytical Chemistry* 79 (24), 9385-9390.



- Liu, G., Wang, L., Qu, H., Shen, H., Zhang, X., Zhang, S., Mi, Z., **2007**. Artificial neural network approaches on composition–property relationships of jet fuels based on GC–MS. *Fuel* 86 (16), 2551-2559.
- Maier, H.R., Jain, A., Dandy, G.C., Sudheer, K.P., **2010**. Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. *Environmental Modelling & Software* 25 (8), 891-909.
- Marinović, S., Bolanča, T., Ukić, Š., Rukavina, V., Jukić, A., **2012**. Prediction of diesel fuel cold properties using artificial neural networks. *Chemistry and Technology of Fuels and Oils* 48 (1), 67-74.
- Marques, J.S., **1999**. *Reconhecimento de Padrões: métodos estatísticos e neuronais*, IST Press, Lisboa.
- Martens, H., Naes, T., **1989**. *Multivariate Calibration*, John Wiley & Sons, New York.
- Mashrei, M.A., Abdulrazzaq, N., Abdalla, T.Y., Rahman, M.S., **2010**. Neural networks model and adaptive neuro-fuzzy inference system for predicting the moment capacity of ferrocement members. *Engineering Structures* 32 (6), 1723-1734.
- Masili, A., Puligheddu, S., Sassu, L., Scano, P., Lai, A., **2012**. Prediction of physical–chemical properties of crude oils by <sup>1</sup>H NMR analysis of neat samples and chemometrics. *Magnetic Resonance in Chemistry* 50 (11), 729-738.
- McKenzie, J.S., Donarski, J.A., Wilson, J.C., Charlton, A.J., **2011**. Analysis of complex mixtures using high-resolution nuclear magnetic resonance spectroscopy and chemometrics. *Progress in Nuclear Magnetic Resonance Spectroscopy* 59 (4), 336-359.
- Mestrelab Research - Chemistry Software Solutions, 2012. *MestreNova Manual*. [accessed on January 2011]; Available from: <http://mestrelab.com/>.
- Mesures Magazine, 2002. *NMR - Strategic Measurement Operation*. [accessed 18/03/2011]; Available from: <http://www.process->

nmr.com/pdfs/Mesures%20Magazine%20Article%20-%20English%20Translation.pdf.

- Meyers, R.A., **2004**. *Handbook of Petroleum Refining Processes*, third edition, The McGraw-Hill Companies, New York.
- Mi, X., Zou, Y., Wei, W., Ma, K., **2005**. Testing the generalization of artificial neural networks with cross-validation and independent-validation in modelling rice tillering dynamics. *Ecological Modelling* 181 (4), 493-508.
- Miller, J.N., Miller, J.C., **2005**. *Statistics and Chemometrics for Analytical Chemistry*, fifth, Pearson Education Limited, England.
- Molina V, D., Uribe, U.N., Murgich, J., **2007**. Partial Least-Squares (PLS) Correlation between Refined Product Yields and Physicochemical Properties with the <sup>1</sup>H Nuclear Magnetic Resonance (NMR) Spectra of Colombian Crude Oils. *Energy & Fuels* 21 (3), 1674-1680.
- Molina V, D., Uribe, U.N., Murgich, J., **2010**. Correlations between SARA fractions and physicochemical properties with <sup>1</sup>H NMR spectra of vacuum residues from Colombian crude oils. *Fuel* 89 (1), 185-192.
- Molina, V.D., Uribe, U.N., Murgich, J., **2007**. Partial Least-Squares (PLS) Correlation between Refined Product Yields and Physicochemical Properties with the <sup>1</sup>H Nuclear Magnetic Resonance (NMR) Spectra of Colombian Crude Oils. *Energy & Fuels* 21 (3), 1674-1680.
- Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., **2007**. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *American Society of Agricultural and Biological Engineers* 50 (3), 885-900.
- Morrison, R.T., Boyd, R.N., **1996**. *Química Orgânica*, Fundação Calouste Gulbenkian, Lisboa.
- Motlaghi, S., Jalali, F., Ahmadabadi, M.N., **2008**. An expert system design for a crude oil distillation column with the neural networks model and the process optimization

- using genetic algorithm framework. *Expert Systems with Applications* 35 (4), 1540-1545.
- Nash, J.E., Sutcliffe, J.V., **1970**. River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology* 10 (3), 282-290.
- Nielsen, K.E., Dittmer, J., Malmendal, A., Nielsen, N.C., **2008**. Quantitative Analysis of Constituents in Heavy Fuel Oil by <sup>1</sup>H Nuclear Magnetic Resonance (NMR) Spectroscopy and Multivariate Data Analysis. *Energy & Fuels* 22 (6), 4070-4076.
- Panda, S.K., Schrader, W., al-Hajji, A., Andersson, J.T., **2007**. Distribution of Polycyclic Aromatic Sulfur Heterocycles in Three Saudi Arabian Crude Oils as Determined by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Energy & Fuels* 21 (2), 1071-1077.
- Pasadakis, N., Gaganis, V., Varotsis, N., **2001**. Accurate determination of aromatic groups in heavy petroleum fractions using HPLC-UV-DAD. *Fuel* 80 (2), 147-153.
- Pavia, D.L., Lampman, G.M., Kriz, G.S., **1996**. *Introduction to Spectroscopy*, second edition, Saunders College Publishing, London.
- Pires, J.C.M., Martins, F.G., Sousa, S.I.V., Alvim-Ferraz, M.C.M., Pereira, M.C., **2008a**. Selection and validation of parameters in multiple linear and principal component regressions. *Environmental Modelling & Software* 23 (1), 50-55.
- Pires, J.C.M., Martins, F.G., Sousa, S.I.V., Alvim-Ferraz, M.C.M., Pereira, M.C., **2008b**. Prediction of the Daily Mean PM<sub>10</sub> Concentrations using Linear Models. *American Journal of Environmental Sciences* 4 445-453.
- Process Instruments, 2001. *Petroleum*. [accessed on June 2011]; Available from: <http://www.process-instruments-inc.com/pages/petroleum.html>.
- Process NMR Associates LLC, 2012a. *Proton NMR Assignments for Functional Groups of Interest in Petroleum Chemistry*. [accessed on March 2011]; Available from: <http://www.process-nmr.com/Table%201%201H%20Shifts.htm>.

- Process NMR Associates LLC, 2012b. *<sup>13</sup>C NMR Chemical Shift Peak Assignment for Functional Groups of Interest in Petroleum Chemistry*. [accessed on March 2011]; Available from: [http://www.process-nmr.com/Table\\_2\\_13C\\_NMR.htm](http://www.process-nmr.com/Table_2_13C_NMR.htm).
- Ramos, P.F.d.O., de Toledo, I.B., Nogueira, C.M., Novotny, E.H., Vieira, A.J.M., Azeredo, R.B.d.V., **2009**. Low field <sup>1</sup>H NMR relaxometry and multivariate data analysis in crude oil viscosity prediction. *Chemometrics and Intelligent Laboratory Systems* 99 (2), 121-126.
- Rezzi, S., Axelson, D.E., Héberger, K., Reniero, F., Mariani, C., Guillou, C., **2005**. Classification of olive oils using high throughput flow <sup>1</sup>H NMR fingerprinting with principal component analysis, linear discriminant analysis and probabilistic neural networks. *Analytica Chimica Acta* 552 (1–2), 13-24.
- Roggo, Y., Chalus, P., Maurer, L., Lema-Martinez, C., Edmond, A., Jent, N., **2007**. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *Journal of Pharmaceutical and Biomedical Analysis* 44 (3), 683-700.
- Sarpal, A.S., Kapur, G.S., Bansal, V., Jain, S.K., Srivastava, S.P., Bhatnagar, A.K., **1998**. Direct Estimation of Aromatic Carbon (Ca) Content of Base Oils by <sup>1</sup>H-NMR Spectroscopy. *Petroleum Science and Technology* 16 (7), 851 - 868.
- Sarpal, A.S., Kapur, G.S., Chopra, A., Jain, S.K., Srivastava, S.P., Bhatnagar, A.K., **1996**. Hydrocarbon characterization of hydrocracked base stocks by one- and two-dimensional n.m.r. spectroscopy. *Fuel* 75 (4), 483-490.
- Sarpal, A.S., Kapur, G.S., Mukherjee, S., Jain, S.K., **1997**. Characterization by <sup>13</sup>C NMR spectroscopy of base oils produced by different processes. *Fuel* 76 (10), 931-937.
- Sastry, M.I.S., Chopra, A., Sarpal, A.S., Jain, S.K., Srivastava, S.P., Bhatnagar, A.K., **1996**. Carbon type analysis of hydrotreated and conventional lube-oil base stocks by i.r. spectroscopy. *Fuel* 75 (12), 1471-1475.
- Satya, S., Roehner, R.M., Deo, M.D., Hanson, F.V., **2007**. Estimation of Properties of Crude Oil Residual Fractions Using Chemometrics. *Energy & Fuels* 21 (2), 998-1005.

- Savorani, F., Tomasi, G., Engelsen, S.B., **2010**. Icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. *Journal of Magnetic Resonance* 202 (2), 190-202.
- Sergeant, G.D., Stubington, J.F., Barrett, D., Do, P.T.D.H., Raval, K.A., **1995**. A comparative study of products derived from Rundle and Stuart shale oils and petroleum lubricating oil base stocks (based on n.m.r., molecular weight and elemental analysis). *Fuel* 74 (1), 51-56.
- Shell Method Series 1600, **2001**. *State of Peptization of Asphaltenes in Heavy Oil Streams*, Shell Global Solutions International B.V., Netherlands.
- Sheremata, J.M., Gray, M.R., Dettman, H.D., McCaffrey, W.C., **2004**. Quantitative Molecular Representation and Sequential Optimization of Athabasca Asphaltenes. *Energy & Fuels* 18 (5), 1377-1384.
- Silva, S.L., Silva, A.M.S., Ribeiro, J.C., Martins, F.G., Da Silva, F.A., Silva, C.M., **2011**. Chromatographic and spectroscopic analysis of heavy crude oil mixtures with emphasis in nuclear magnetic resonance spectroscopy: A review. *Analytica Chimica Acta* 707 (1-2), 18-37.
- Simanzhenkov, V., Idem, R., **2003**. *Crude Oil Chemistry*, Marcel Dekker, New York.
- Singh, I.D., Kothiyal, V., Ramaswamy, V., Krishna, R., **1990**. Characteristic changes of asphaltenes during visbreaking of North Gujarat short residue. *Fuel* 69 (3), 289-292.
- Skoog, D.A., Holler, F.J., Nieman, T.A., **1997**. *Principles of Instrumental Analysis*, fifth edition, Saunders College Publishing, New York.
- Snape, C.E., **1986**. Application of NMR to fossil fuels: an assessment of recent advances. *Fresenius' Journal of Analytical Chemistry* 324 (8), 781-785.
- Snee, R.D., **1977**. Validation of Regression Models: Methods and Examples. *Technometrics* 19 (4), 415-428.
- Somov, V., Rozental', D., Syroezhko, A., Neronov, D., Passet, A., **1999**. Visbreaking Vacuum Resids. *Chemistry and Technology of Fuels and Oils* 35 (1), 1-3.

- Sousa, S.I.V., Martins, F.G., Alvim-Ferraz, M.C.M., Pereira, M.C., **2007**. Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environmental Modelling & Software* 22 (1), 97-103.
- Speight, J.G., **2001**. *Handbook of Petroleum Analysis*, John Wiley & Sons, New York.
- Speight, J.G., **2002**. *Handbook of Petroleum Product Analysis*, John Wiley & Sons, New Jersey.
- Speight, J.G., **2006**. *The Chemistry and Technology of Petroleum*, fourth, CRC Press, London.
- Stoyanova, R., Nicholson, J.K., Lindon, J.C., Brown, T.R., **2004**. Sample Classification Based on Bayesian Spectral Decomposition of Metabonomic NMR Data Sets. *Analytical Chemistry* 76 (13), 3666-3674.
- The Qualion Company, 2006. *Industrial FT NMR Analysis*. [accessed on March 2011]; Available from: [www.qualion-nmr.com](http://www.qualion-nmr.com).
- University of Illinois at Chicago, 2001. *Nanoscale structures of asphaltene molecule, asphaltene steric-colloid and asphaltene micelles & vesicles*. [accessed on March 2011]; Available from: <http://tigger.uic.edu/~mansoori/Asphaltene.Molecule.html>.
- Verdier, S., Coutinho, J.A.P., Silva, A.M.S., Alkilde, O.F., Hansen, J.A., **2009**. A critical approach to viscosity index. *Fuel* 88 (11), 2199-2206.
- Wang, S., Dong, X., Sun, R., **2010**. Predicting saturates of sour vacuum gas oil using artificial neural networks and genetic algorithms. *Expert Systems with Applications* 37 (7), 4768-4771.
- Warne, K., Prasad, G., Rezvani, S., Maguire, L., **2004**. Statistical and computational intelligence techniques for inferential model development: a comparative evaluation and a novel proposition for fusion. *Engineering Applications of Artificial Intelligence* 17 (8), 871-885.
- Wauquier, J.P., **1995**. *Petroleum Refining 1. Crude Oil, Petroleum Products, Process Flowsheets*, Éditions Technip, Institut Français du Pétrole Publications, Paris.

- Winning, H., Larsen, F.H., Bro, R., Engelsen, S.B., **2008**. Quantitative analysis of NMR spectra with chemometrics. *Journal of Magnetic Resonance* 190 (1), 26-32.
- Wise, B.M., Gallagher, N.B., **1996**. The process chemometrics approach to process monitoring and fault detection. *Journal of Process Control* 6 (6), 329-348.
- Wold, S., Sjöström, M., Eriksson, L., **2001**. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* 58 (2), 109-130.
- Woods, J., Kung, J., Kingston, D., Kotlyar, L., Sparks, B., McCracken, T., **2008**. Pétroles bruts canadiens: étude comparative des fractions SARA à partir d'une technique de séparation HPLC modifiée. *Oil & Gas Science and Technology - Rev. IFP* 63 (1), 151-163.
- Yao, Y., Liu, D., Che, Y., Tang, D., Tang, S., Huang, W., **2010**. Petrophysical characterization of coals by low-field nuclear magnetic resonance (NMR). *Fuel* 89 (7), 1371-1380.
- Yoshida, T., Maekawa, Y., Uchino, H., Yokoyama, S., **1980**. Derivation of structural parameters for coal-derived oil by carbon-13 nuclear magnetic resonance spectrometry. *Analytical Chemistry* 52 (6), 817-820.
- Zerbe, O., 2010. *Lecture Course: NMR Spectroscopy*. [accessed on December 2012]; Available from: <http://www.oci.uzh.ch/group/pages/zerbe/NMR.pdf>.
- Zerzucha, P., Daszykowski, M., Walczak, B., **2012**. Dissimilarity partial least squares applied to non-linear modeling problems. *Chemometrics and Intelligent Laboratory Systems* 110 (1), 156-162.
- Zhang, N., Zhao, S., Sun, X., Xu, Z., Xu, C., **2010**. Storage Stability of the Visbreaking Product from Venezuela Heavy Oil. *Energy & Fuels* 24 (7), 3970-3976.





9

**APPENDIX**



## Appendix A: Economic Evaluation

As referred in Chapter 6, since good results were obtained when predicting fuel oil properties using  $^1\text{H}$  NMR spectroscopy and fuel oil was a final product which properties were required in a real time, an economical evaluation of the implementation of an online NMR spectrometer at fuel oil stream was made. The objective of this economic study was to verify the profitability of the acquisition of an online NMR spectrometer. For this economic evaluation more information should be included, but only the information here presented was the one given by both entities involved, Matosinhos refinery and Qualion Company. Taking into account the information given by Qualion Company (The Qualion Company, 2006), a quotation for the entire project would be about 314000 € (\$ 400905). This investment includes: i) NMR analyzer installation; ii) commissioning and training remote PC for remote diagnostic; iii) assistance sample system; iv) shelter house; and v) project management. It excludes the piping, infrastructure and the humidity and temperature control, which should be prepared and acquired by the refinery. However, these last necessities are low cost when compared with the initial investment. Thereby, the total investment cost was estimated to be about 350000 €.

As already mentioned, the online monitoring of fuel oil properties can contribute to: i) decrease the quantity of cutter stocks used in the production of fuel oil; ii) produce fuel oil with the desired properties; and iii) avoid product loss. Taking into account the quantity of fuel oil produced, the quantity of cutter stocks required and the blending of fuel oil at the collecting tank, it would be possible to save about 5 a 7 \$/t when a good production and optimum blending are obtained. Considering that the production of fuel oil is around 1600 t/day and fuel oil is produced during 333 days in a year (excluding some days when for some reason the production of fuel oil is interrupted) it will be possible to save 2083248 €/year (\$ 2664000).

As the initial investment is around 350000 € (\$ 447405), the payback time of the initial investment will be approximately 3 months. This payback time is a highly attractive investment return.

## **Appendix B: Copyright Permissions**

**FW: NON-RIGHTSLINK**

Permission Requests - UK [permissionsuk@wiley.com]

**Enviado:** quarta-feira, 28 de Agosto de 2013 10:37**Para:** sandralsilva@ua.pt**Importância:**Alta

Dear Sandra Silva,

Thank you for your request.

Permission is granted for you to use the material requested for your thesis/dissertation subject to the usual acknowledgements and on the understanding that you will reapply for permission if you wish to distribute or publish your thesis/dissertation commercially.

Permission is granted solely for use in conjunction with the thesis, and the material may not be posted online separately.

Any third party material is expressly excluded from this permission. If any material appears within the article with credit to another source, authorisation from that source must be obtained.

Kind Regards

Emma Willcox  
Permissions Assistant**WILEY**

---

**From:** sandralsilva@ua.pt [mailto:sandralsilva@ua.pt]**Sent:** Tuesday, August 27, 2013 12:22 PM**To:** Permissions - US**Subject:** NON-RIGHTSLINK**Importance:** High

To whom it may concern,

I requested your permission to use the figure 6.5 "High resolution chromatogram of the <343°C fraction of light Nigerian crude" (page 56) and figure 6.20 "Heavy fuel oil" (page 75) in my PhD thesis. I will give full credit to the original source. If I have permission to reprint the figure I will indicate, in my document, that the figures in question were taken from "Characterization of Spilled Oil Samples. Purpose, sampling, analysis and interpretation".



# RightsLink®

[Home](#)[Account Info](#)[Help](#)

**ACS Publications** Title:  
High quality. High impact.

Characterization of Feeds, Intermediates, and Products from Heavy Oil Processes by High-Temperature Simulated Distillation and Thin-Layer Chromatography with Flame Ionization Detection

Logged in as:  
Sandra Silva  
Account #:  
3000367401

[LOGOUT](#)

**Author:** Bhajendra N. Barman \*  
**Publication:** Energy & Fuels  
**Publisher:** American Chemical Society  
**Date:** Sep 1, 2005  
Copyright © 2005, American Chemical Society

## PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

If credit is given to another source for the material you requested, permission must be obtained from that source.

[BACK](#)[CLOSE WINDOW](#)

Copyright © 2013 [Copyright Clearance Center, Inc.](#) All Rights Reserved. [Privacy statement.](#)  
Comments? We would like to hear from you. E-mail us at [customercare@copyright.com](mailto:customercare@copyright.com)

28/08/13

Re: Permission

## **Re: Permission**

Lee Smith [[lsmith@process-instruments-inc.com](mailto:lsmith@process-instruments-inc.com)]

Enviado: terça-feira, 27 de Agosto de 2013 19:11

Para: [sandralsilva@ua.pt](mailto:sandralsilva@ua.pt)

Sandra,

Yes, you may use that image in your Thesis document. Good luck,

Lee Smith

President

Process Instruments, Inc.

On 8/27/2013 10:35 AM, [sandralsilva@ua.pt](mailto:sandralsilva@ua.pt) wrote:

To whom it may concern,

I would like to ask how I could receive permission to use a figure from Process Instruments, <http://www.process-instruments-inc.com/pages/petroleum.html>, to publish in my PhD thesis.

The figure that I am referring is about a typical on-line Raman spectrum of a gasoline stream – "Raman Spectroscopy Defined".

I look forward in receiving an answer from you.

Best regards,

Sandra Silva

--

Lee Smith  
Process Instruments, Inc.  
825 North, 300 West, Suite NE 220  
Salt Lake City, UT 84103-1459

Phone 801-322-1235 ext. 100

Fax 801-322-5607

[lsmith@process-instruments-inc.com](mailto:lsmith@process-instruments-inc.com)

Website: [www.process-instruments-inc.com](http://www.process-instruments-inc.com)

**RightsLink®**[Home](#)[Account Info](#)[Help](#)**ACS Publications**  
High quality. High impact.**Title:** Distribution of Polycyclic Aromatic Sulfur Heterocycles in Three Saudi Arabian Crude Oils as Determined by Fourier Transform Ion Cyclotron Resonance Mass SpectrometryLogged in as:  
Sandra Silva  
Account #:  
3000367401[LOGOUT](#)**Author:** Saroj K. Panda,<sup>T,‡</sup> Wolfgang Schrader,<sup>§</sup> Adnan al-Hajji,<sup>||</sup> and, and Jan T. Andersson<sup>\*,†</sup>**Publication:** Energy & Fuels**Publisher:** American Chemical Society**Date:** Mar 1, 2007

Copyright © 2007, American Chemical Society

**PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE**

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

If credit is given to another source for the material you requested, permission must be obtained from that source.

[BACK](#)[CLOSE WINDOW](#)

Copyright © 2013 [Copyright Clearance Center, Inc.](#) All Rights Reserved. [Privacy statement.](#)  
Comments? We would like to hear from you. E-mail us at [customercare@copyright.com](mailto:customercare@copyright.com)



**JOHN WILEY AND SONS LICENSE  
TERMS AND CONDITIONS**

Nov 21, 2013

---

This is a License Agreement between Sandra Lopes Silva ("You") and John Wiley and Sons ("John Wiley and Sons") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by John Wiley and Sons, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

License Number	3273540579550
License date	Nov 21, 2013
Licensed content publisher	John Wiley and Sons
Licensed content publication	Wiley Books
Licensed content title	Process Analytical Technology: Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries, 2nd Edition
Book title	
Licensed copyright line	Copyright © 2010, John Wiley and Sons
Licensed content author	Katherine A. Bakeev (Editor)
Licensed content date	Apr 1, 2010
Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	4
Original Wiley figure/table number(s)	Figure 10.1, 10.2, 10.3, 10.4
Will you be translating?	No
Total	0.00 USD
Terms and Conditions	

**TERMS AND CONDITIONS**

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a "Wiley Company") or a society for whom a Wiley Company has exclusive publishing rights in relation to a particular journal (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms

**De:** "Woods, John" <John.Woods2@nrc-cnrc.gc.ca>  
**Assunto:** RE: permission  
**Date:** Wed, 23 Mar 2011 15:53:30 -0700  
**Para:** SANDRA LOPES DA SILVA <sandralsilva@ua.pt>



Good afternoon Sandra,

You have the permission that you seek.  
Thank you very much for referring to the article and for asking permission to use the figure.

Sincerely,

John R. Woods  
National Research Council of Canada

---

From: SANDRA LOPES DA SILVA [sandralsilva@ua.pt]  
Sent: March 23, 2011 6:17 PM  
To: Woods, John  
Subject: Re: permission

Dear Mrs. John Woods,

I am preparing a paper entitled NMR in the analysis of heavy crude oil mixtures: A review to be published by Elsevier in the journal Analytica Chimica Acta. I requested your permission to use the figure 3 "Typical <sup>13</sup>C and <sup>1</sup>H NMR spectra for SARA fractions (results for the heavy oil B sample)" in my article. I will give full credit to the original source. If I have permission to reprint the figure I will indicate, in my article, that the figure in question was taken from "Canadian Crudes: A comparative study of SARA fractions from a modified HPLC separation techniques".

I look forward in receiving an answer from you.

Yours trustfully,  
Sandra

Em Wed, 23 Mar 2011 12:43:11 -0700

"Woods, John" <John.Woods2@nrc-cnrc.gc.ca> escreveu:  
> Good afternoon Sandra,  
>  
> I believe that the company that I work for, National  
> Research Council of Canada, would want to know how you  
> are going to use the figure.  
> I suppose that if there is a reference to where the  
> figure was obtained from, there should be no problem in  
> using it.  
>  
> Best regards,  
>  
> John  
>  
> -----Original Message-----  
> From: SANDRA LOPES DA SILVA [mailto:sandralsilva@ua.pt]  
> Sent: March 22, 2011 1:29 PM  
> To: Woods, John; Kung, Judy; Kingston, David;



# RightsLink®

[Home](#)[Account Info](#)[Help](#)

**ACS Publications** Title:  
High quality. High impact.

1H Diffusion-Ordered Spectroscopy (DOSY) Nuclear Magnetic Resonance (NMR) as a Powerful Tool for the Analysis of Hydrocarbon Mixtures and Asphaltenes

Logged in as:  
Sandra Silva  
Account #:  
3000367401

[LOGOUT](#)

**Author:** Emmanuelle Durand, Martin Clemancey, Anne-Agathe Quoineaud, Jan Verstraete, Didier Espinat, Jean-Marc Lancelin,

**Publication:** Energy & Fuels

**Publisher:** American Chemical Society

**Date:** Jul 1, 2008

Copyright © 2008, American Chemical Society

## PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

If credit is given to another source for the material you requested, permission must be obtained from that source.

[BACK](#)[CLOSE WINDOW](#)

Copyright © 2013 [Copyright Clearance Center, Inc.](#) All Rights Reserved. [Privacy statement.](#)  
Comments? We would like to hear from you. E-mail us at [customercare@copyright.com](mailto:customercare@copyright.com)

**ELSEVIER LICENSE  
TERMS AND CONDITIONS**

Aug 27, 2013

This is a License Agreement between Sandra Lopes Silva ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

Supplier	Elsevier Limited The Boulevard, Langford Lane Kidlington, Oxford, OX5 1GB, UK
Registered Company Number	1982084
Customer name	Sandra Lopes Silva
Customer address	Campus Santiago, Departamento de Quimica Aveiro, other 3810
License number	3217170928012
License date	Aug 27, 2013
Licensed content publisher	Elsevier
Licensed content publication	Fuel
Licensed content title	Hydrocarbon characterization of hydrocracked base stocks by one- and two-dimensional n.m.r. spectroscopy
Licensed content author	Amarjeet Singh Sarpal, Gurpreet Singh Kapur, Anju Chopra, Surendra Kumar Jain, Som Prakash Srivastava, Akhilesh Kumar Bhatnagar
Licensed content date	March 1996
Licensed content volume number	75
Licensed content issue number	4
Number of pages	8
Start Page	483
End Page	490
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	2
Format	both print and electronic

**ELSEVIER LICENSE  
TERMS AND CONDITIONS**

Aug 27, 2013

This is a License Agreement between Sandra Lopes Silva ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

Supplier	Elsevier Limited The Boulevard, Langford Lane Kidlington, Oxford, OX5 1GB, UK
Registered Company Number	1982084
Customer name	Sandra Lopes Silva
Customer address	Campus Santiago, Departamento de Quimica Aveiro, other 3810
License number	3217171240352
License date	Aug 27, 2013
Licensed content publisher	Elsevier
Licensed content publication	Fuel
Licensed content title	Structural characterization of FCC feeds from Indian refineries by NMR spectroscopy
Licensed content author	Babita Behera, Siddharth S. Ray, I.D. Singh
Licensed content date	August 2008
Licensed content volume number	87
Licensed content issue number	10-11
Number of pages	12
Start Page	2322
End Page	2333
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	2
Format	both print and electronic

