MOHAMMED V UNIVERSITY - RABAT
Faculty of Medicine and Pharmacy

Year: 2019 **THESIS N °: 06/19 CSVS**

DOCTORAL COLLEGE: LIFE AND HEALTH SCIENCES

Speciality: Drug Sciences

Option: Analytical Chemistry- Chemometrics

DOCTORAL THESIS

Presented by

Issam BARRA

Theme:

THE APPLICATION OF THE FTIR SPECTROSCOPYAND CHEMOMETRIC TOOLS IN THE QUALITY CONTROL OF PETROLEUM PRODUCTS MARKETED IN MOROCCO

Keywords: Chemometrics, FTIR, Petroleum Products, Quality Control.

Publicly defended by the 19/07/2019 in the presence of the jury:

Prof. Yahia CHERRAH	Mohammed V University in Rabat-FMPR	**President**
Prof. Mustapha ABOUATIA	Mohammed V University in Rabat-FMPR	**Reporter**
Prof. Sanaa CHALA	Mohammed V University in Rabat-FMPR	**Reporter**
Prof. Fouad FETHI	Mohammed First University in Oujda-FS	**Reporter**
Prof. Mohamadine EL M'RABET	The Hassan II Agronomic and Veterinary Institute in Rabat	**Examinator**
Prof. Yvan VANDER HEYDEN	Vrije Universiteit Brussel-VUB-Belgium	**Examinator**
Prof. Abdelaziz BOUKLOUZE	Mohammed V University in Rabat-FMPR	**Supervisor**

À ma chère épouse **Farida**, *en témoignage de mon amour et mon affection.*

À mon petit ange **Ziyad**, *Je te souhaite une vie pleine de joie et de succès*

*À mes chers parents (*__Cherkaoui Barra__ *&* __Hakima Ouriki__*), Pour ces longues années de soutien inconditionnel, pour leur confiance permanente. Pour tous les sacrifices et les encouragements qu'ils ont consentis pour nous, pour que nous puissions poursuivre nos études. J'en suis conscient et très reconnaissant.*

À mes frères **Abdelilah** *et sa femme* **Hasna**, **Ayoub**, **Zakaria** *et* **Ali** *et mes sœurs* **Amina** *et* **Nihad**, *En témoignage de mon affection fraternelle, de ma profonde tendresse et reconnaissance, je vous souhaite une vie pleine de bonheur et de succès. Que vos rêves soient en tous exaucés.*

*À mon oncle (*__Salah__ *et ma deuxième mère* __Bahija__*) et mon oncle* **Ahmed** *et tous les membres de sa petite famille (*__Fatouma__, __Mohamed__ *et* __Hafida__, __Jalal__ *et* __Ilham__*).*

À mes grands-mères **Zohra** *et* **Mahjouba**, *à ma belle-famille, ma belle-mère* **Saadia**, **Said**, **Naima** *et* **Laila** *et mes tantes* **Bahija**, **Hafida** *et* **Malika** *Qui m'ont accompagné par leurs prières, douceur, que Dieu, le tout puissant, vous protège et vous garde.*

À la mémoire de ma tente **Souad Barra** *et mes grands-parents* **Bouaaza Barra** *&* **Salh Ouriki**, *Vous demeurez toujours présents à nos esprits. Tellement vous nous manquez ce jour-là. Aucune dédicace ne pourra traduire notre profond amour et gratitude pour tout l'amour et l'affection que vous nous avez offerts. Que le paradis soit votre céleste demeure.*

À la grande famille **Barra**,

À la grande famille **Ouriki**,

À la grande famille **Bougrami**,

À mes ami(e)s, particulièrement My Ahmed Bellimam, Nourdine Tijani, Marouane Balouki, Tarik Addajou, Abdelfettah Boukhaled, Othmane Elorfi, abdelhadi Ait Nouiss, M'hamed Elouark, Hicham Sahbani, Mourad Kharbach, Mohammed Bousrabat, Mohammed Alaoui Mansouri, et à tous ceux qui me sont chers, À toutes les personnes qui ont participé à l'élaboration de ce travail à tous ceux que j'ai omis de citer.

Issam BARRA

"Data do not yield information except with the intervention of the mind. Information does not yield meaning except with the intervention of imagination"

Theodore Levitt

AKNOWLEDGEMENTS

It is a great pleasure to get this opportunity to express my deepest appreciation and gratitude to all those that were directly or indirectly involved in the completion of this doctoral thesis.

My deepest gratitude should go to my supervisor Prof. **Abdelaziz Bouklouze** from the Biopharmaceutical-Toxicological research team and the responsible of the unit of Instrumental analysis and data mining for giving me such a great opportunity to work on this promising field. It has been a real pleasure to work with him. He has always been available to advise me and demand a high quality of work in all my endeavors. He encouraged and expected me to think more independently about our experiments and results. I am very grateful for his patience, motivation, and determination.

I would like to express my cordial thanks and gratitude to Prof. **Mostafa Qannari**, Prof. **Mohamed Hanafi** and all members of Statistics, chemometrics and Sensometrics team for giving me the opportunity to pursue my research stay with their team, as well as for their valuable recommendations, their fruitful discussions, their unlimited support and for the excellent working facilities at ONIRIS INRA, Nantes.

I would like to express my thanks and gratitude to Prof. **Rasmus Bro** and Prof. **Jose Manuel Amigo Rubio** for giving me the opportunity to participate to the Chemometric School of Copenhagen and for the excellent courses delivered at the department of Food of the Faculty of Science of Copenhagen.

I would like to deeply thank Prof. **Yahia Cherrah** Director of the Laboratory of Pharmacology and Toxicology at the Faculty of Medicine and Pharmacy Rabat for his guidance, his continuous encouragement, and his kind advices during my thesis.

My gratitude and all my regards to Prof. **Jamal Taoufik** (Director of the Doctoral College "Sciences de la Vie et de la Santé" and Vice Doyen of Pharmacy at the Faculty of Medicine and Pharmacy) for carrying out the success of our educational program. My

sincere gratitude and thankfulness to the professors of ongoing seminars at the doctoral college “Cedoc SVS”; for their coaching, assiduity and their sense of responsibility.

I would like to extend my thanks to my colleagues from the same research team at the faculty of medicine and pharmacy of Rabat and from the Institute of Forensic Science of the Royal Gendarmerie, namely, Dr **Mourad Kharbach**, Dr **Mohammed Bousrabat**, Dr **Mohammed Alaoui Mansouri**, Dr **Mly Ahmed Bellimam** and Mr **Abdelfettah Boukhaled** for their advices, guidance and help.

LIST OF ARTICLES

Publications related to the thesis

1. **Issam Barra**, Mohammed Alaoui Mansouri, Yahia Cherrah, Mourad Kharbach, Abdelaziz Bouklouze, FTIR fingerprints associated to a PLS-DA model for rapid detection of smuggled non-compliant diesel marketed in Morocco, Journal of Vibrational Spectroscopy, Volume 101, March 2019, Pages 40-45.
2. **Issam Barra**, Mohammed Alaoui Mansouri, Mourad Kharbach, Mohammed Bousrabat, Yahia Cherrah, Abdelaziz Bouklouze, Discrimination and quantification of Moroccan gasoline adulteration with diesel using Fourier Transform Infrared Spectroscopy and chemometric tools, Journal of AOAC INTERNATIONAL, Volume 102, N° 3, 2019, Pages 966-970.
3. **Issam Barra**, Mohammed Alaoui Mansouri, Mourad Kharbach, Yahia Cherrah, Mohamed Hanafi, El Mostafa Qannari, Abdelaziz Bouklouze; Discrimination and classification of diesel fuels using FTIR and GC-MS analysis and chemometrics methods, Talanta, statut "under review".

Additional publications

4. Mourad Kharbach, Rabie Kamal, Mohammed Bousrabat, Mohammed Alaoui Mansouri, **Issam Barra**, Katim Alaoui, Yahia Cherrah, Yvan Vander Heyden, Abdelaziz Bouklouze ; 'Fatty-acid profiling vs UV-Visible fingerprints for geographical classification of Moroccan Argan oils', Food Control, 2019.
5. Mourad Kharbach, Rabie Kamal, Mohammed Bousrabat, Mohammed Alaoui Mansouri, **Issam Barra**, Katim Alaoui, Yahia Cherrah, Yvan Vander Heyden, Abdelaziz Bouklouze ; 'Characterization and classification of PGI Moroccan Argan oils based on their FTIR fingerprints and chemical composition', Chemometrics and Intelligent Laboratory Systems, 2017.

LIST OF COMMUNICATIONS

ORAL COMMUNICATIONS:

1. **Issam Barra**, Mohammed Alaoui Mansouri, Mourad Kharbach, Yahia Cherrah, Mohamed Hanafi, El Mostafa Qannari, Abdelaziz Bouklouze, Prédiction rapide de l'indice de cetane dans les carburants diesel par le couplage FTIR-PLSR, Journées Pratiques Francophones des Sciences Analytiques 2019, April 25-26 2019, Marrakech, Morocco.
2. **Issam Barra**, Mohammed Alaoui Mansouri, Mourad Kharbach, Yahia Cherrah, Abdelaziz Bouklouze, Identification of gasoline adulteration with diesel using the GC/MS Technique, 1st International Conference on Materials and Environmental Science, December 1-3 2016, Oujda, Morocco.

POSTER COMMUNICATIONS

1. **Issam Barra**, Mohammed Alaoui Mansouri, Mourad Kharbach, Yahia Cherrah, Abdelaziz Bouklouze, Le couplage de la spectroscopie infrarouge FTIR avec les outils chimiométrique une solution pour le contrôle de qualité des produits pétroliers, 7th International Meeting on Chemometric and Quality, October 23-25 2018, Fez, Morocco.
2. **Issam Barra**, Mohammed Alaoui Mansouri, Yahia Cherrah, Abdelaziz Bouklouze, Rapid detection of smuggled non-compliant diesel in Moroccan market using mid infrared spectroscopy coupled to a PLS-DA model, The 17th Chemometrics in Analytical Chemistry Conference, June 25-29 2018, Halifax, Canada.
3. **Issam Barra**, Mohammed Alaoui Mansouri, Yahia Cherrah, Abdelaziz Bouklouze, The use of MIR-FTIR coupled to chemometric tools to differentiate between diesel fuels from different suppliers, The 17th Chemometrics in Analytical Chemistry Conference, June 25-29 2018, Halifax, Canada.

4. **Issam Barra**, Mohammed Alaoui Mansouri, Mourad kharbach, Yahia Cherrah, Abdelaziz Bouklouze, Classification of diesel samples by supplier using MIR-FTIR coupled to chemometric tools, Congrès Annuel CHIMIOMETRIE XIX, January 30-31 2018, Paris, France.
5. **Issam Barra**, Mohammed Alaoui Mansouri, Mourad kharbach, Yahia Cherrah, Abdelaziz Bouklouze, Determination and quantification of Moroccan gasoline adulteration with diesel using Infrared Spectroscopy and Chemometric Tools, Congrès Annuel CHIMIOMETRIE XVII, January 30-31 2017, Paris, France.

Table of content

List of figures 13

List of tables 16

List of Abbreviations 17

Abstract 18

Résumé 19

ملخص 20

GENERAL INTRODUCTION 21

LITERATURE REVIEW 24

CHAPTER I. REVIEW OF PETROLEUM PRODUCTS 25

1. Introduction 26

2. Chemical composition of petroleum products 26

2.1. Hydrocarbons 27

2.2. Sulfur compounds 28

2.3. Nitrogen compounds 29

2.4. Oxygen compounds 30

2.5. Metal compounds 30

3. Oil refining 31

3.1. The distillation process 31

3.2. The refining 32

3.3. Petroleum products 33

4. Petroleum products: Moroccan context 35

4.1. Oil refining in Morocco 35

4.2. The oil market after SAMIR 37

5. Key findings 39
CHAPTER II. QUALITY CONTROL OF PETROLEUM PRODUCTS 41
1. Introduction 42
2. Sampling 42
3. Physic-chemical analysis 43
3.1. Color of Petroleum Products 43
3.2. Density 43
3.3. Viscosity 44
3.4. Pour point 45
3.5. Cetane number 46
3.6. Distillation at Atmospheric Pressure 46
3.7. Cold Filter Plugging Point 47
3.8. Sulfur content 48
4. Key findings 48
CHAPTER III. INFRARED SPECTROSCOPY 50
1. Introduction 51
2. Discovery of infrared radiation 51
3. Fundamentals of infrared spectroscopy 52
3.1. Electromagnetic radiation 52
3.2. Molecule's Vibrational energy level 52
4. Techniques of samples examination in spectroscopy 59
4.1. Absorbed light 59
4.2. Transmittance 61
4.3. Reflective processes 62

5. Spectroscopic instruments 63
5.1. Conventional infrared spectrometers: dispersive spectrometers 63
5.2. Fourier Transform Infrared Spectroscopy 64
6. Attenuated Total Reflectance (ATR) 65
7. Qualitative interpretation of the infrared spectrum 67
8. Quantification with infrared spectroscopy 69
9. Key findings 69
CHAPTER IV: CHEMOMETRICS 70
1. Introduction 71
1.1. Definitions 72
1.2. Mathematical preprocessing applied to the spectral data 72
2. Chemometric methods 77
2.1. Unsupervised method: the Principal Component Analysis (PCA) 77
2.2. Supervised methods 81
2.3. Statistical criteria for assessing the quality of models 89
3. A review of Chemometric tools applied to the field of petroleum products 90
3.1. Classification studies 90
3.2. Quantification studies 91
4. Key findings 92
EXPERIMENTAL PART: 93
THE APPLICATION OF FTIR SPECTROSCOPY AND CHEMOMETRIC TOOLS IN THE QUALITY CONTROL OF PETROLEUM PRODUCTS MARKETED IN MOROCCO 93
I. Introduction 94
II. Materials and methods 96

1. GC-MS analysis 96
2. FTIR spectra acquisition 96
3. Sample collection 97
3.1. Rapid detection of smuggled non compliant Diesel products 97
3.2. Classification of diesel samples by suppliers 97
3.3. Discrimination and quantification of gasoline adulteration with diesel 97
3.4. Prediction of diesel quality parameters 97
4. Software and data processing 97
5. The quality control test 98
III. Results and discussion 99
1. FTIR fingerprints associated to a PLS-DA model for rapid detection of smuggled non-compliant diesel marketed in Morocco 99
2. Discrimination and classification of diesel fuels using FTIR and GC-MS analysis and Chemometrics methods 109
3. Discrimination and quantification of Moroccan gasoline adulteration with diesel using Fourier Transform Infrared Spectroscopy and Chemometric tools 125
4. The association of the FTIR spectroscopy with PLS2 and PLS1 algorithms for the prediction of ten quality parameters of diesel fuels 130
CONCLUSION AND PERSPECTIVES 142
References 144
Annex 1. Published Articles 156

List of figures

Figure 1: Crude oil composition and the relationship between boiling point / molar mass / structure[15] 27
Figure 2: Atmospheric and vacuum distillation of crude oil and examples of associated fractions 32
Figure 3: crude oil imports by Moroccan SAMIR in 2014 [20] 36
Figure 4: Production of SAMIR in petroleum products in 2014 [20] 36
Figure 5: Diesel fuels importations by Companies 38
Figure 6: Gasoline importations by Companies 38
Figure 7: Hydrometer used for density determination[26] 44
Figure 8: Apparatus assembly for the pour point test [28] 45
Figure 9: Apparatus assembly for the determination of distillation points of petroleum products at atmospheric pressure using a gas burner [31] 47
Figure 10: Apparatus assembly for the determination of the CFPP [32] 48
Figure 11: Spectral domains of electromagnetic radiation 51
Figure 12: Oscillator Model 53
Figure 13: Harmonic potential - Representation of potential binding energy as a function of interatomic distance 54
Figure 14: Potential energy curve versus displacement under the anharmonic model 57
Figure 15: Molecular vibration modes: symmetric (purple) and asymmetric elongation (orange) 58
Figure 16: Modes of molecular vibration- deformation in the plane of the molecule 58
Figure 17: Modes of molecular vibration- deformation out of the plane of the molecule 59
Figure 18: Interaction of radiations with the matter 59
Figure 19: An incident light on a surface A of a material of thickness dx and the concentration of the molecules C 60
Figure 20: Transmittance of electromagnetic radiation 62
Figure 21: A schematic diagram of the classical dispersive IR spectrophotometer 64
Figure 22: Schematic illustration of a modern FTIR Spectrophotometer 65
Figure 23: Principle of the ATR system 66
Figure 24: After mean-centering all variables will have equal “length” and a mean value=zero 73
Figure 25: the effect of 1st derivative on FTIR spectra 73

Figure 26: the effect of smoothing preprocessing on FTIR spectra 74
Figure 27: the effect of the normalization preprocessing on FTIR spectra 75
Figure 28: the effect of the application of the SNV on a group of spectra.....76
Figure 29: the effect of the application of the Multiplicative Scatter Correction on a group of spectra 77
Figure 30: (a) initial representation of the data, (b) representation in the new PC1 vs PC2 space 78
Figure 31: Matrix decomposition of the PCA 79
Figure 32: Illustration of how classification is accomplished in PLS-DA (estimated class values) 83
Figure 33: Data of a PLSR can be arranged as two tables, matrices, X and Y[76].....85
Figure 34: The geometric representation of PLSR [76] 88
Figure 35: Examples of chromatograms obtained by GC-MS of authentic diesel (A) and smuggled diesel (B).....101
Figure 36: FTIR spectra of authentic diesel samples and smuggled samples before pretreatment (A) and after pretreatments (SNV and mean centering) (B) 102
Figure 37: Scores plot of PCA model with two principal components showing smuggled diesel samples in red and authentic samples in green.....103
Figure 38: Plots of loadings of the PCA model.....104
Figure 39: PLS-DA Scores plot with two latent variables (the first and the second) presenting two separate groups of samples. In red smuggled products and authentic samples in green 105
Figure 40: Plots of loadings of the PLS-DA model 106
Figure 41: Estimated class values for calibration and validation sets obtained by the PLS-DA model.....107
Figure 42: FTIR spectra of diesel samples.....110
Figure 43: Chromatograms obtained by the GC-MS technique for Diesel fuels 111
Figure 44: PCA Scores plot, PC1vs PC2, on the FTIR spectra.....112
Figure 45: PC1 to PC4 loading plots of the PCA model on FTIR spectra 113
Figure 46: 3D PLS-DA scores plots of the FTIR model 115
Figure 47: LV1 to LV4 loading plots of PLS-DA-FTIR model.....116
Figure 48: Estimated Class Values of training and test sets obtained by PLS-DA model models established on FTIR spectra 117
Figure 49: PCA Scores plot, PC1vs PC3, on the GC-MS data 118
Figure 50: PC1 to PC4 loading plots of the PCA model on GC-MS data.....119
Figure 51: LV1 to LV4 loading plots of the PLS-DA-GCMS model 120

Figure 52: 3D PLS-DA scores plots of the GC-MS model 121
Figure 53: Estimated Class Values of training and test sets obtained by PLS-DA model models established on GC-MS data 122
Figure 54: Fourier transform infrared spectra of three types of samples: gasoline, diesel, and adulterated gasoline samples with diesel 126
Figure 55: Scores graph for the first two principal components obtained for the discrimination of three classes of samples: Gasoline, Diesel and Adulterated samples 127
Figure 56: Predicted versus real adulterant contents graph obtained by the PLS model for calibration set (black) and test set (Red). 128
Figure 57: FTIR spectra of diesel samples before preprocessing (A) and after applying the 1st derivative and the mean-centering pretreatments (B) 132
Figure 58:PCA scores plot (PC1 Vs PC2) of the ten responses 134
Figure 59: PCA scores plot (PC1 Vs PC2) based on FTIR spectra of the diesel samples 135
Figure 60: Predicted versus real values of: (A): Conductivity, (B): Pour Point, (C) Viscosity, (D): CFPP, (E): Flash Point, (F): Color, (G): Cetane Number, ((H): Density, (I): Water contain, (J): 40% Distillation 137
Figure 61: Predicted versus real Cetane number values graph obtained by the PLS-FTIR model 140

List of tables

Table 1: Structure of hydrocarbon compounds in the petroleum products....28
Table 2: Structure of sulfur compounds in petroleum products 29
Table 3: Structure of Nitrogen Compounds in Petroleum Products 29
Table 4: Structure of Nitrogen Compounds in Petroleum Products 30
Table 5: Structure of metal compounds in petroleum products....30
Table 6: Examples of global properties of different crude oil fractions (Middle East) 33
Table 7: example of diesel property information [89]....95
Table 9: Testing methods for diesel quality parameters determination....98
Table 10: List of main compounds of diesel with their retention time and monitored ions 100
Table 11: Figures of merit of PLS-DA model 107
Table 12: Discrimination parameters of the FTIR PLS-DA models 114
Table 13: Discrimination parameters of the GC-MS PLS-DA models 123
Table 14: Performance of PLS model for determination of adulterant levels in gasoline samples....129
Table 15: PLS model results for prediction of cetane number in diesel samples 139
Table 16: Overview of the models developed for the prediction of diesel quality parameters using spectroscopic techniques 130
Table 17: Results obtained by the PLS2 model....138

List of Abbreviations

ASTM: American Society for Testing and Materials

ATR: Attenuated Total Reflectance

CN: Cetane Number

FTIR: Fourier Transform Infrared Spectroscopy

GC-MS: Gas Chromatography coupled to Mass Spectrometry

IP: Institute of Petroleum

ISO: International Standardization Organization

LPG: Liquefied Petroleum Gas

LV: Latent Variable

MIR: Medium Infrared

MLR: Multi-Linear Regression

MSC: Multiplicative Scatter

NIR: Near Infrared

PC: Principal Component

PCA: Principal Component Analysis

PCR: Principal Component Regression

PLSR: Partial Least Squares Regression

PLS-DA: Partial Least Square Discriminant Analysis

R^2: Coefficient of correlation

RMSEC: Root Mean Squared Error of Calibration

RMSECV: Root Mean Squared Error of Cross Validation

RMSEP: Root Mean Squared Error of Prediction

SIMCA: Soft Independent Modeling Class Analogy

SNV: Standard Normal Variate

SEP: Standard Prediction Error

SVD: Singular Value Decomposition

VD: Vacuum Distillate

VIP: Variable Importance in Projection

VR: Vacuum Residue

Abstract

Today, petroleum products represent one of the main human needs. Consumers are increasingly demanding higher quality products that meet the international specific requirements.

Certifying the quality of oil derivatives, detecting frauds in these products and meeting the regulatory requirements involves the completion of a dozens of physic-chemical tests that require specific devices known for their expensiveness, sophistication and their destructive nature.

The interest in Fourier Transformed Infrared Spectroscopy (FTIR) has been growing thanks to improvements in instrumentation, the development of the ATR accessories that simplifies the spectra recording. The computing advances and the development of the chemometric tools that use mathematical and statistical tools to visualize, extract and process the information contained in the infrared spectra. The coupling of the FTIR spectroscopy and the chemometric tools has shown great abilities to provide powerful models for authenticating and predicting the quality parameters of petroleum products.

That one may to meet the analytical needs whether the frauds detection or to make easier and speed up the quality parameters control of petroleum derivatives marketed in Morocco, the coupling of the FTIR spectroscopy and the chemometric methods was selected.

On one hand two classification models (PLS-DA) were established, firstly for the detection of smuggled diesels, further, the classification of diesel fuels by suppliers which allow tracing the origin of the fuel. Both models demonstrated a great selectivity and the classes were predicted without ambiguity. On the other hand, the established regression models generate, in all cases, good prediction results. They were characterized by very significant R^2 values, and low RMSEP values.

The combining of the FTIR spectroscopy and the chemometric tools yielded good classification and prediction models. The FTIR spectroscopy approach thus may be highly recommended as a fast alternative for the quality control and frauds detection of the petroleum products.

Keywords: Chemometrics, FTIR, petroleum products, quality control.

Résumé

De nos jours, les produits pétroliers représentent l'un des principaux besoins humain. Les consommateurs exigent de plus en plus des produits de qualité supérieurs qui répondent a leurs besoins d'utilisations spécifiques.

Garantir la qualité des dérivées pétrolières et répondre aux exigences réglementaires nécessite la réalisation d'une vingtaine de tests physico-chimiques qui demandent des appareils spécifiques connus par leur prix cher, sophistication et leur caractère destructif.

L'intérêt porté à la spectroscopie infrarouge à Transformé de Fourrier (FTIR) a été croissant grâce aux améliorations de l'instrumentation, au développement des accessoires ATR permettant de simplifier l'enregistrement des spectres et aux progrès de l'informatique.

Les outils chimiométriques qui utilisent des méthodes mathématiques et informatiques permettant de visualiser, d'extraire et de traiter l'information contenu dans les spectres infrarouge ont démontré de grandes capacité à fournir des modèles performants permettant de classifier et de prédire les paramètres de qualité des produits pétroliers.

Le but de cette thèse été de produire des modèle chimiométriques de classification et prédiction basé le couplage de la spectroscopie infrarouge à transformé de fourrier et des outils chimiométrique.

D'une part les modèles de classification PLS-DA mis en place, le premier pour la détection des produits diesel de la contrebande et le deuxième pour remonter à l'origine des échantillons de diesel de quatre fournisseurs, ont démontrés une bonne sélectivité et les classes ont été prédites sans erreur.

D'autre part les modèles de prédiction calculés génèrent, dans tous les cas, de bons résultats de régression. Caractérisés par des valeurs R2 très significatives, et de faibles erreurs de prédiction RMSEP.

Mots clés: Chimiométrie, Chimie analytique, produits pétroliers, contrôle de qualité.

ملخص

تعد المنتجات البترولية اليوم واحدة من أهم الاحتياجات البشرية. حيت يطالب المستهلكون بشكل متزايد بمنتجات عالية الجودة تلبي احتياجاتهم الخاصة.

يتطلب ضمان جودة المشتقات النفطية القيام بحوالي عشرين اختبارًا فيزيوكيميائيًا تتطلب أجهزة خاصة معروفة بسعرها الباهظ وتطورها وطبيعتها المدمرة.

ازداد الاهتمام باستخدام التحليل الطيفي للأشعة تحت الحمراء بفضل التطور في الأجهزة، وتطوير ملحقاتها التي ساهمت في تبسيط تسجيل الطيف وكذلك التقدم في مجال المعلوميات.

أظهرت أدوات الكيمياء القياسية التي تستخدم طرقا رياضية والقائمة على الكمبيوتر لتصوير المعلومات في أطياف الأشعة تحت الحمراء واستخراجها ومعالجتها قدرة كبيرة على توفير نماذج قوية لتصنيف والتنبؤ بمعلومات جودة المحروقات.

الهدف من هذه الأطروحة هو إنتاج نماذج كيمياء قياسية قادرة على تصنيف والتنبؤ، مبنية على مزاوجة التحليل الطيفي للأشعة تحت الحمراء و أدوات الكيمياء القياسية.

من ناحية نماذج التصنيف، الأول للكشف عن منتجات الديزل المهربة والثاني لتتبع عينات الديزل من أربعة موردين، أظهرت انتقائية جيدة وتم التنبؤ بالأصناف دون خطأ.

من ناحية أخرى، أنتجت نماذج التنبؤ المحسوبة، في جميع الحالات، نتائج انحدار R^2 مقبولة، وقيم أخطاء في التنبؤ RMSEP منخفضة .

كلمات البحث :

الكيمياء القياسية، الكيمياء التحليلية، المنتجات البترولية، مراقبة الجودة.

GENERAL INTRODUCTION

Since the beginning of the industrial era, energy occupies a preponderant place in the life of the human being. The technological progress has enabled man to discover new energy resources, namely all the energy-related products of which we quote: oil and natural gas, which are the only ones capable of responding to the increase in energy needs. Indeed, a wide part of the energy of tomorrow is yet to be discovered. It must then be produced and consumed with a minimum of impact on the environment. This task must be carried out in a world where energy is often at the heart of economic, political and social issues. In addition, for the purpose that energy remains affordable, producers and their partners must constantly look for ways to reduce costs.

Petroleum refining is a heavy industry that converts the crude oil into two classes of products, energy products such as diesel and gasoline, and non-energy products such as petrochemical feedstock, lubricants, paraffin and bitumen. These products are distributed to consumers either directly or through a distribution network including warehouses and service stations(1). The conversion of crude oils is done in refineries using highly automated plants including a number of manufacturing units, such as separation, conversion and purification process(2).The complexity of the process depends on the range of products manufactured, the quality of the crude oil and consumer requirements(3,4).

In Morocco, after the closure of the only national refining unit “SAMIR”, the government authorized the various companies acting in this sector to import the various petroleum derivatives directly from abroad in order to satisfy the internal demand on these essential products. With this opening of the market many deviant practices have appeared, whose main purpose is illicit gain that it is by the adulteration of a more expensive product by another less expensive, the commercialization of non-compliant smuggled products(5) which come essentially from Algeria or the trading of low quality products as a high quality one.

The detection of such frauds requires well-equipped laboratories, and the tests must be performed in accordance with international standards, such as the American Society for Testing

and Materials (ASTM) and the Energy Institute in the United Kingdom (formerly the Institute of Petroleum). The instrumental technology development has allowed the oil products-control field a new generation of specific, reliable and automated analyzers. However, these equipments are complex in use, require sample preparation (destructive), time consuming, demand laboratory labor and well-trained personnel which make the control of the petroleum samples very costly.

In order to meet the analytical needs, whether the frauds detection or to make easier and speed up the control of quality parameters of petroleum derivatives marketed in Morocco, the objectives of this thesis were the application of the FTIR spectroscopy technique and the chemometric methods for enhancing the quality control and the frauds detection problems in the field of petroleum products. According to the bibliographical research, the association of spectroscopy and chemometrics have been shown to be a powerful tools for the quality control in several domains, namely: foods and natural oils control, drugs formulations and fuels(6–10).

In one hand, Fourier Transform Infrared Spectroscopy in the mid-infrared range, the technique of choice in this thesis, is a spectroscopic technique used to record an infrared spectrum of absorption or emission of a solid, liquid or gas sample(11,12). The vibration bands observed in this area are very much related to the hydrocarbons composition in the petroleum samples. The acquisition of the spectrum generally requires a little volume of the sample without any preparation, and finally, the fastness in spectrum recording. On the other hand, the use of the multivariate statistical analysis makes it possible to simultaneously classify and predict the properties of interest of diesel and gasoline samples based only on their FTIR spectra [9–13].

In order to reach the objectives cited above, this manuscript is divided into two principal parts. The former part is devoted to a literature review, contains four chapters. In the first one, we have focus on the crude oil, and discussed their chemical composition. Also we have drawn the petroleum refining process and the various resulting products.

In the second chapter we have described the proposed tests by the international regulations for the quality control and certification of petroleum derivatives. While the third chapter will presents the theoretical aspects of infrared spectroscopy and the spectroscopic techniques used to examine samples.

The last chapter of the bibliographic part is dedicated to the chemometric methods. We started by providing some definitions and specifying the purpose and types of the mathematical preprocessing applied to the spectral data. Afterwards, Detailed explanations about the chemometric methods used to accomplish the set aims (PCA, PLS-DA and PLS regression), as well as, the statistical criteria for assessing the quality of a multivariate model will be provided. The end of this chapter was marked by a bibliographic review of already published work that combined chemometric tools with analytical techniques in the field of petroleum.

The practical part of this manuscript has been written following the IMRAD criteria. We have presented the results obtained as part of this thesis and their discussion that revolve around the feasibility of coupling of the Fourier Transform Infrared Spectroscopy and chemometric tools for the characterization and the prediction of the quality parameters of the petroleum products marketed in Morocco, as follow:

- The rapid detection of smuggled non-compliant diesel samples by the establishment of a PLS-DA model that classified two groups of diesel samples (authentic products and smuggled ones) based on their FTIR spectra.

- Further, the comparison of the performances of two PLS-DA models built on the FTIR and GC-MS analysis to ensure the traceability of diesel product by the classification of diesel samples from four suppliers.

- On the regressions, the first study was the quantification of gasoline adulteration by diesel based on a PLSR model set up on the FTIR fingerprints.

- Afterwards, the prediction of ten diesel quality parameters including the color, viscosity, conductivity at 20°, water content, flash point, filterability limit temperature, 40% of distillation and the pour point was performed based on a PLS2 regression model built by FTIR spectra. Moreover, the prediction of the cetane number which is a very important parameter in quality certification of diesels was improved by the use of a PLS1 model built on the same spectral data.

LITERATURE REVIEW

1- **REVIEW OF PETROLEUM PRODUCTS**
2- **QUALITY CONTROL OF PETROLEUM PRODUCTS**
3- **INFRARED SPECTROSCOPY**
4- **CHEMOMETRICS**

CHAPTER I. REVIEW OF PETROLEUM PRODUCTS

1. Introduction

Oil is a complex mixture of molecules, which the majority is heterogeneous hydrocarbons. Impurities, such as heteroelements and metals, are also present. We will present in detail the chemical composition of petroleum products.

The distillation allows splitting the crude oil into petroleum fractions (gasoline, kerosene, diesel ...) depending on the boiling temperature of different molecules. The lightest oils are the most desired by refiners, as they yield directly a lot of precious light cuts (diesel, gasoline, naphtha). Conversely, heavy oils give more bitumen and residual fuel oil, which must be sold as they are at low prices or to be converted to lighter cuts, particularly by hydrocracking (addition of hydrogen).

2. Chemical composition of petroleum products

Petroleum is a complex mixture of many hydrocarbons resulting from the decomposition of marine organisms living millions of years ago. The chemical composition of the oil depends on the geographical origin: there are currently a hundred different crudes in the market. It is mainly consisting of hydrocarbons (93 to 99% (m / m)[1]) including alkanes or paraffins, cycloalkanes or naphthenes and aromatic compounds. Other elements are often present including organic sulfur compounds (0.01 to 6% (m / m)), nitrogenous (0.05 to 0, 5% (w / w)), oxygenated (0.1 to 0.5% (w / w)) and some metals (0.005 to 0.15% (w / w)), such as nickel and vanadium. .

Figure 1.1 shows the composition of crude oils as a function of boiling temperature and molar mass. The molar mass is estimated from the number of carbon atoms based on the formula of paraffins (alkanes) C_nH_{2n+2}.

[1] % (m / m): mass percentage

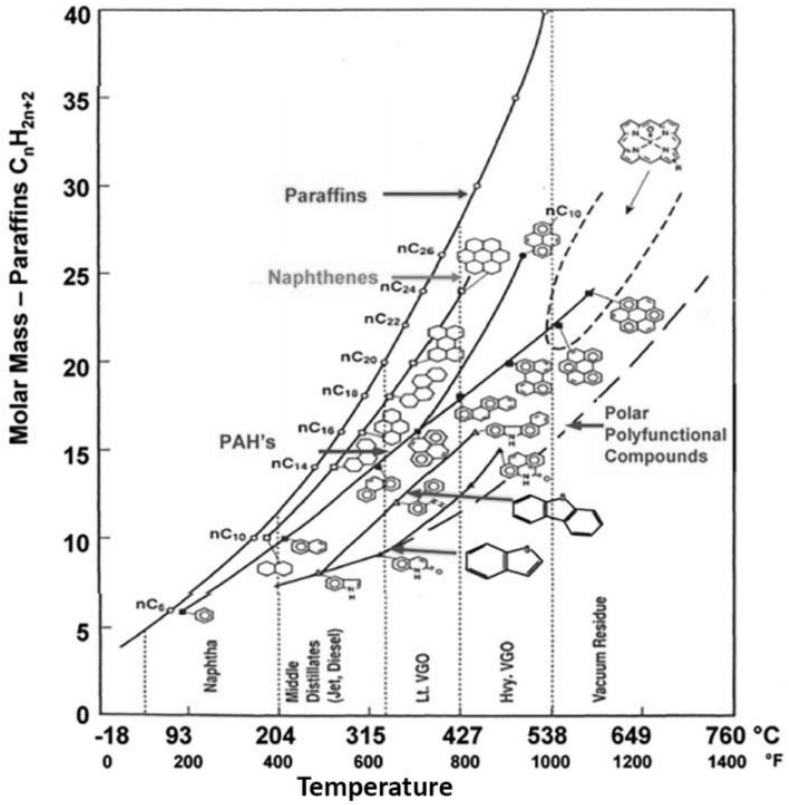


Figure 1: Crude oil composition and the relationship between boiling point / molar mass / structure(15)

This figure illustrates the fact that the oil matrices are very complex and poly-dispersed. Indeed, the number of isomers increases exponentially with the number of carbon atoms. Consequently, when the values of boiling temperatures and molar masses are higher, the more the molecules present are heterogeneous in terms of structures. If, for example, the boiling point is 427 ° C., the number of carbon atoms can vary from 10 to more than 25, and the structures of the corresponding molecules cover both paraffins (alkanes) and polycyclic compounds(3).

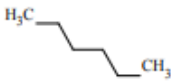
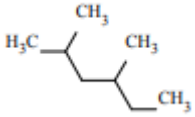
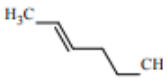
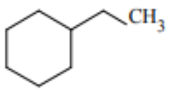
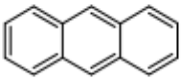
2.1. Hydrocarbons

The hydrocarbons can be distinguished by chemical families and according to the unsaturations degree of the molecular structure(3) (Table 1):

- **Saturated aliphatic hydrocarbons**: normal paraffins called N-paraffins (linear alkanes) or iso-paraffins (branched alkanes).

- **Unsaturated aliphatic hydrocarbons**: olefins (alkenes) which are rarely present in the crude oil because of their reactivity. However, olefins can be produced during the refining processes and especially during conversion processes of heavy cuts.
- **Saturated cyclic aliphatic hydrocarbons**: naphthenes which are cyclic carbon of 5 or 6 atoms that may comprise one or more cycles and branched chains.
- **Aromatic hydrocarbons**: polyunsaturated cyclic compounds present in high amounts in the heavier cuts. They may contain one or more aromatic and / or naphthenic cycles and / or branched chains.

Table 1: Structure of hydrocarbon compounds in the petroleum products

Families	N-paraffines	Iso-paraffines	Olefines	Naphtenes	Aromatiques
Formula	C_nH_{2n+2}	C_nH_{2n+2}	C_nH_{2n}	C_nH_{2n}	C_nH_{2n-8k}
Example					

n: number of carbon atoms

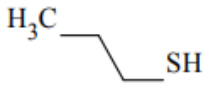
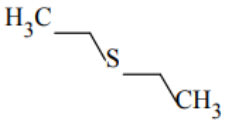
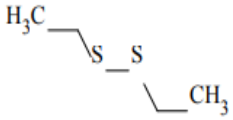
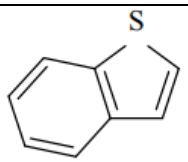
k: number of unsaturations

2.2. Sulfur compounds

Sulfur is the most common hetero-element in oil. The density of the matrix is strongly correlated with the content of this element. The sulfur compounds present in petroleum products belong to different chemical families (Table 2):

- **Thiols or mercaptans**: acidic and corrosive compounds especially present in light cuts, of formula R—S—H.
- **Sulphides**: slightly corrosive and odorless due to their low volatility, of the formula R—S—R' (or polysulfides of formula R—S—…—S—R').
- **Thiophene compounds**: which have an aromatic character.

Table 2: Structure of sulfur compounds in petroleum products

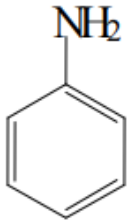
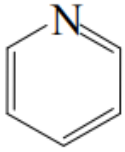
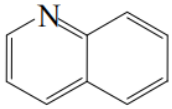
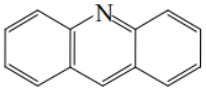
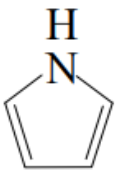
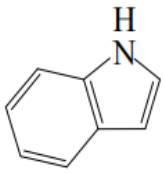
families	Mercaptans	Sulfides	Disulfides	Benzothiophenes
Examples				

The sulfur compounds are at the origin of the atmospheric pollution (SO2 and SO3) and the deactivation of certain catalysts used particularly in the refining processes or in catalytic converters. The European specifications governing the maximum sulfur content in fuels are also regularly lowered. In 2010, the maximum sulfur content in gasoline and diesel was set at 10 ppm (EN228 and EN590)(16).

2.3. Nitrogen compounds

Nitrogen compounds are mainly present in heavy fractions, in smaller quantities than sulfur compounds. They differ essentially in their neutral or basic character (Table 3).Basic nitrogen compounds in addition to neutral nitrogen compounds are known to damage acid catalysts. They then constitute an obstacle to the refining of heavy cuts.

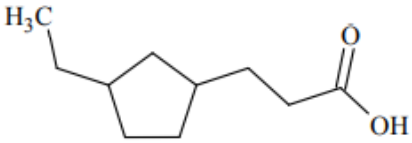
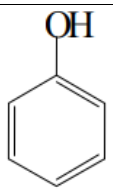
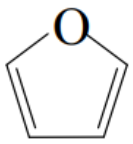
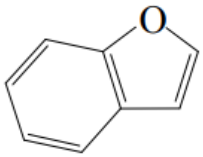
Table 3: Structure of Nitrogen Compounds in Petroleum Products

families	**Basic derivatives**				**Neutral derivatives**	
	Aniline	Pyridine	Quinoline	Acridine	Pyrrole	Indole
Examples						

2.4. Oxygen compounds

Among oxygen compounds present in petroleum products there are naphthenic carboxylic acids, esters, phenols, furans and benzofurans (Table 4). Although present in low levels, mainly in heavy cuts, oxygen compounds have an acidic character which is responsible for the overall acidity of crude oils and causes corrosion problems.

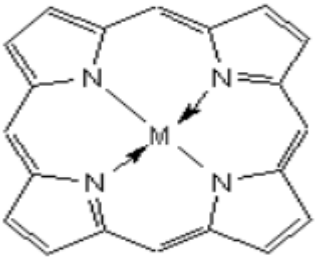
Table 4: Structure of Nitrogen Compounds in Petroleum Products

families	Naphthenic acid	Phenol	Furane	Benzofuran
Examples				

2.5. Metal compounds

Nickel and vanadium are the most common metals in petroleum products and are mainly found in heavy products. They can be found in macro-molecule whose structures are poorly known. They are also present in some smaller compounds of the porphyrin family. In this structure, the pattern is constituted by a set of four pyrrolic cycles, the metal being at the center of this set in the form Ni^{2+} and VO^{+} (Table 5). Although they are present at very low levels, they are also considered like poisons for the catalysts used for the conversion of heavy cuts.

Table 5: Structure of metal compounds in petroleum products

family	Porphyrines
Example	

M=Ni^{2+} or VO^{+}

3. Oil refining

Oil refining means all the treatments and transformations to produce maximum products with high commercial values from crude oil, such as fuels (gasoline, Diesel, kerosene and the fuel oil), intermediates for petrochemicals, plastics. . .

3.1. The distillation process

The distillation is a preliminary step in the refining of petroleum products. It allows the split of crude oil in order to obtain different oil cuts depending on the boiling point (**Figure 2**). Firstly, the atmospheric distillation allows to separate the gas fraction (<35°C), gasoline (35-175°C), kerosene (175-235°C) and diesel (235-350°C). The part of the petroleum product that was not distilled during this operation is called the atmospheric residue (AR) composed of molecules whose boiling point is greater than 350◦C. The atmospheric residue can then be separated by the vacuum distillation into two other petroleum fractions: the vacuum distillate VD (350-550 ° C.) and the vacuum residue VR (> 550 ° C). This operation is carried out under vacuum in order to avoid the cracking of molecules that occurs above 400◦C. Cracking is the phenomenon that corresponds to the breakage of a complex (big) molecule into smaller elements.

The oil fractions are therefore characterized by a range of boiling point and a number of carbon atoms. Their proportion in crude oil and their composition depend on the geographical origin of the oil and the concerned fraction (Table 6). Indeed, it is noted that the evolution of the density and the impurities content (hetero-elements and metals) is a function of the oil fraction. Then when the crude oil is rich in light cuts and contain less impurity, it will be more requested, because it will require fewer refining steps.

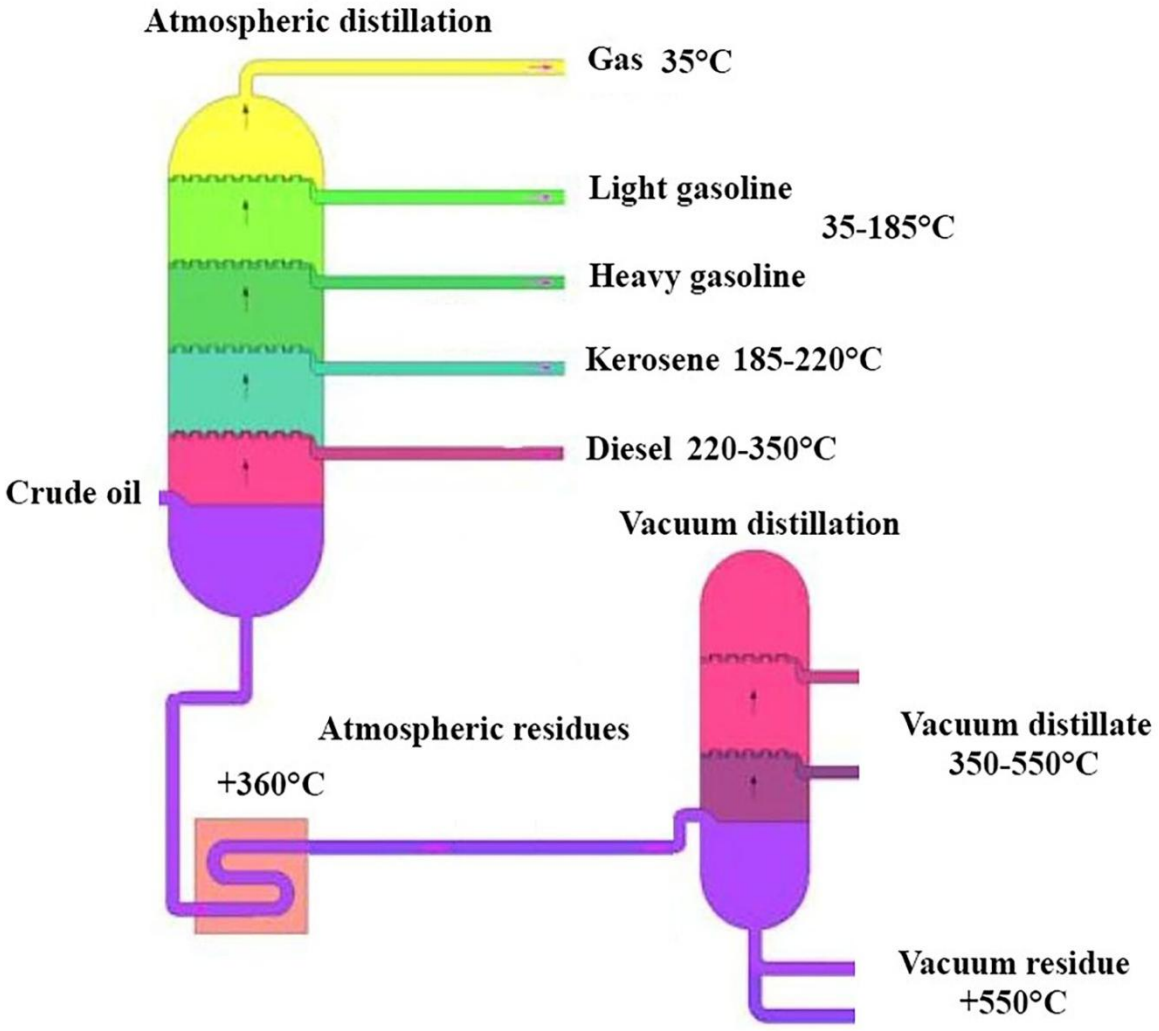


Figure 2: Atmospheric and vacuum distillation of crude oil and examples of associated fractions

3.2. The refining

Commercial products are governed by specifications that guarantee their performance and limit their impact on the environment (sulfur content, polycyclic aromatics, etc.). The various refining operations aim to transform these products in order to obtain the largest possible quantity of recoverable fuel products that meet the specifications. Thus, the gasoline will be transformed to improve their octane number. Diesels will undergo operations to improve their cetane number and reduce their sulfur content. Kerosene will be used for air transport after improvement, in particular, of their cold behavior. Finally, heavy products also produce commercial derivatives

such as heavy fuel oils used as fuel for thermal power stations or for ships, lubricating oils that are dewaxed VDs and bitumen produced from VR.

Table 6: Examples of global properties of different crude oil fractions (Middle East)

Fractions	Gas	Gasoline	Kerosene	Diesel	VD	VR
number of carbon atoms	C_1—C_4	C_4—C_{10}	C_{10}—C_{14}	C_{14}—C_{25}	C_{25}—C_{55}	+C_{55}
Boiling points interval (◦C)	<0	0-180	180-230	230-375	375-600	+600
Yield on crude (% (m / m))	1,37	17,72	6,74	24,37	23,50	26,30
Density (d_4^{15})	0,654	0,742	0,793	0,851	0,935	1,037
Sulfur (% (m / m))	0,003	0,035	0,150	1,400	3,800	5,000
Nitrogen (ppm)	-	-	-	-	≈ 1000	≈ 2000

3.3. Petroleum products

3.3.1. Liquefied petroleum gas (LPG)

LPG Consists of a mixture of paraffinic and olefinic hydrocarbons such as propane and butane. Itis produced for use as a fuel, and is stored and handled as liquids under pressure. LPG has boiling points ranging from about –74 °C to +35 °C, it is colorless, and the vapors are heavier than air and extremely flammable. The important qualities from an occupational health and safety perspective of LPGs are vapor pressure and control of contaminants(17).

3.3.2. Gasoline

The most important refinery product is motor gasoline, a blend of relatively low-boiling hydrocarbon fractions, including reformate, alkylate, aliphatic naphtha (light straight-run naphtha), aromatic naphtha (thermal and catalytic cracked naphtha) and additives. Gasoline

blending stocks have boiling points which range from ambient temperatures to about 185 °C, and a flash point below –40 °C. Additives are used to enhance gasoline performance and provide protection against oxidation and rust formation. Aviation gasoline is a high-octane product, specially blended to perform well at high altitudes.

3.3.3. Kerosene (Jet fuel)

Kerosene is a mixture of paraffins and naphthenes with usually less than 20% aromatics. It has a flashpoint above 38 °C and a boiling range of 185 °C to 220 °C. To improve the Jet fuels properties for flight operation, especially at high altitudes, and to prevent damage to the turbines, certain additives are used (including protecting against frost and corrosion and preventing static charging). This addition of additives differentiates jet fuel from lighter fractions of kerosene (petroleum). By far the world's most commonly used aviation fuel made from kerosene is Jet A1, also called jet fuel.

3.3.4. Diesel fuels

Diesel is a light-colored mixture of paraffins, naphthenes and aromatics, and may contains moderate quantities of olefins. It has a flashpoint above 60 °C and boiling range of about 220 °C to 350 °C, and is often hydro-desulphurized to eliminate sulfur which is toxic to human health, the environment and the engines. Many special additives could be incorporated to improve its quality, it is also blended with bio-fuels such as biodiesel. Currently, up to 7 vol. % biodiesel can be included in diesel fuel in accordance with DIN EN 14214.

3.3.5. Residual products

Many ships and commercial and industrial facilities use residual fuels or combinations of residual and distillate fuels, for power, heat and processing. Residual fuels are dark-colored, highly viscous liquid mixtures of large hydrocarbon molecules, with flashpoints above 121 °C and high boiling points. The critical specifications for residual products are viscosity and low sulfur content (for environmental control)(18).

4. Petroleum products: Moroccan context

Morocco is the country of North Africa where the first oil research gave rise to the hope of discovering important reserves. This hope was reinforced at a recent date by the success of the surveys undertaken on certain points of the zone of the Atlantic plains, but has not been confirmed until the moment(19).

4.1. Oil refining in Morocco

SAMIR or the Moroccan Company of the Refining Industry is a Moroccan group specialized in the refining of petroleum products, located in the city of Mohammedia. It was the only refinery in the Moroccan kingdom. Since its controversial privatization in 1997, it was controlled by the Saudi-Saudi group Corral Petroleum Holding owned by Saudi billionaire Mohammed Al-Amoudi(20)(21).

The refinery is located near the oil port of Mohammedia, in the largest fuel industrial zone in Morocco. It was designed to process 10 million tons per year of crude oil and a storage capacity of 2 million m^3. It was connected to the port by a modern pipeline.

In 2014 SAMIR treated 6 million tons of crude oil, imported mainly from four major oil countries. After the Gulf War, Morocco has diversified its sources of substitution to increase its security stocks(20).

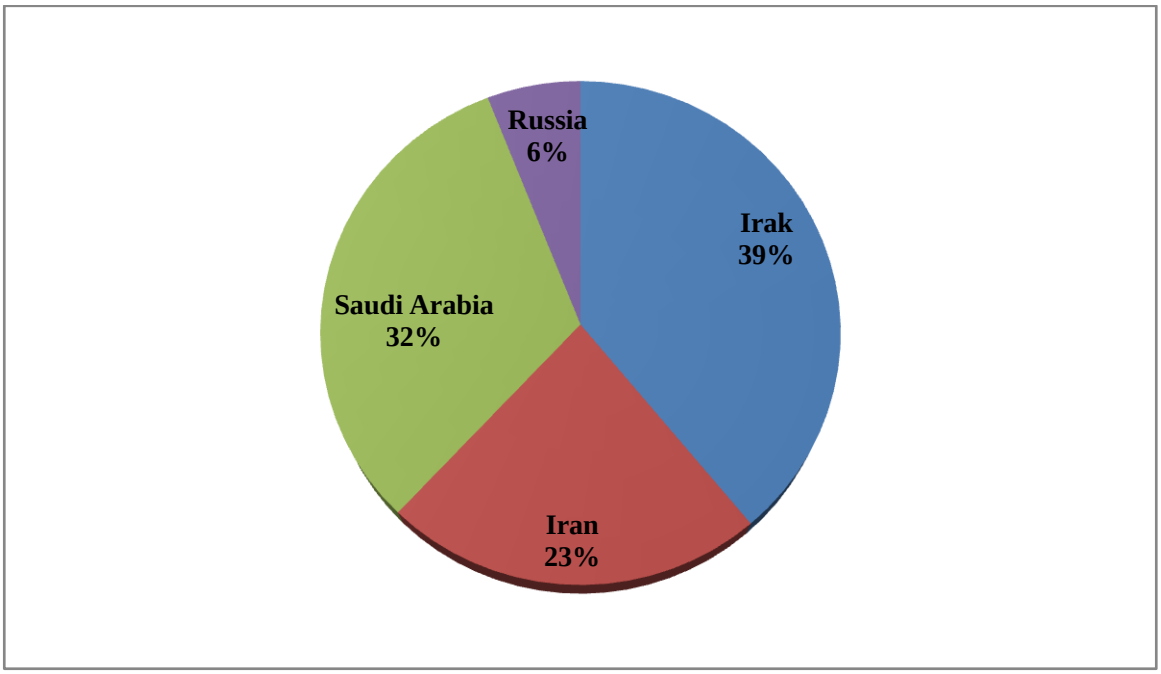


Figure 3: crude oil imports by Moroccan SAMIR in 2014(20)

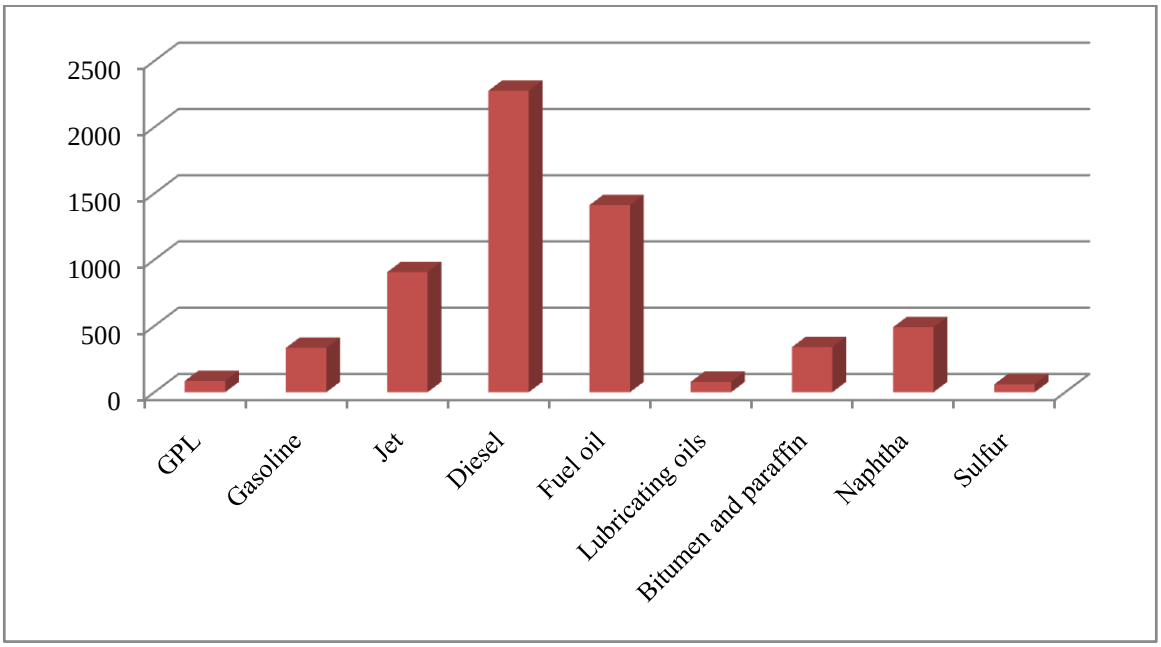


Figure 4: Production of SAMIR in petroleum products in 2014(20)

SAMIR covered about 60% of the national need especially of fuel oils (diesel, gasoline and the fuel) before its activity had to be stopped In August 2015, the refinery had to end its operations because of financial difficulties.

4.2. The oil market after SAMIR

After the closure of the only oil refining unit in Morocco, the government gave the authorization to the various companies marketing these products to import petroleum derivatives from abroad, but in parallel, it implemented a new law elaborated by the Ministry of Energy with the objective of strengthening the quality control of liquid petroleum products and ensuring the availability of refined hydrocarbons or natural gas fuel at service stations or filling stations. Likewise, the new text aims to ensure the supply of petroleum products and natural gas fuel to the national market.

Liquid petroleum product distribution companies in Morocco are classified into two categories:

- Multinationals working through subsidiaries under Moroccan law: Vivo Energy Morocco (Shell), Total Morocco and Libya Oil Morocco.
- Moroccan companies, the most important in the market, are: Afriquia SMDC, Petrom, Ziz, Winxo, Petromin ...

At the end of 2017, Afriquia owned 543 stations or 21.92% of the total. 348 stations for Vivo Energy, 310 for Total, 212 for Petrom, 216 for Ziz, 203 for Winxo and 185 for Libya Oil(22).

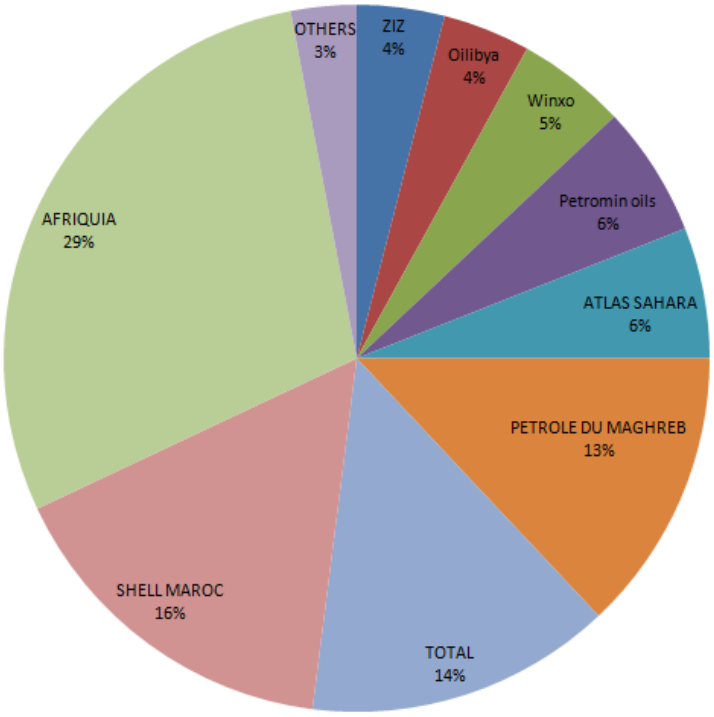


Figure 5: Diesel fuels importations by Companies

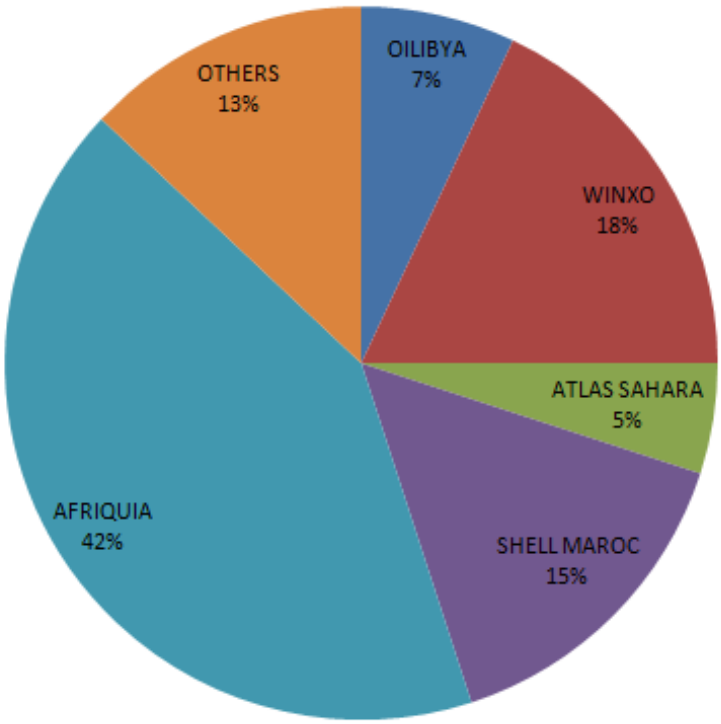


Figure 6: Gasoline importations by Companies

This opening of the oil market has allowed the appearance of a lot of frauds including adulteration (of a more expensive product with another cheaper or even by other chemicals) as the case of the adulteration of unleaded gasoline with diesel or kerosene. Another type of cheating that has been emerged is selling the products of a cheaper brand as those of a more expensive one, or even selling non-compliant smuggled products that are characterized by very poor quality. Smuggled diesel that cross Moroccan borders illicitly especially in the eastern and southern zones (the Moroccan-Algerian border zone) are resulting from an incomplete refining process, and a non-compliant storage and transport conditions.

To avoid these problems, the new law n°1-72-255 aims to enable agents in charge of the quality control of petroleum products to intervene at all stages of refining, importing, storage, transport and sales outlets and to repress fraud through the establishment of adequate sanctions against offenders including, in addition to substantial financial fines, the possibility of suspending temporarily or definitively the approval of operators putting consumer non-compliant products(23).

5. Key findings

Petroleum refining is an industrial process that converts the crude oil into various finished products such as gasoline, Diesel oil or naphtha. Heterogeneous mixture of various hydrocarbons (molecules composed of carbon atoms and hydrogen), the crude oil is unusable as it is (there are several types of crude oil which are distinguished among others by their viscosity and their sulfur content). Its components must be separated in order to obtain the final products that can be exploited directly.

In general there are two main types:

Energy products: such as gasoline, diesel or fuel oil;

Non-energy products: such as lubricants, bitumen and naphthas used in petrochemicals.

Refining is no longer limited to the separation of different hydrocarbons. Complex chemical processes are also implemented to optimize the final products. The various oil cuts can thus undergo transformations, improvements and blends to obtain marketable products and meet the new environmental standards.

The market for petroleum products in Morocco has undergone a great transformation after the halt of the only refining unit SAMIR. The government has encouraged the importers, distributers and marketing company by giving them more facilities. Some people benefit this Indulgence to commit cheats that reduce the quality and threaten consumers and the environment. To put an end to these frauds the government has updated the laws by the requirement of rigorous controls to ensure the good quality of petroleum derivatives in the Moroccan market. Even-though these controls are expensive, time consuming and requires qualified personnel.

CHAPTER II. QUALITY CONTROL OF PETROLEUM PRODUCTS

1. Introduction

Oil is considered a principal chemical and energy resource that has accompanied and catalyzed the technological development of the humanity. It is also one of the main causes of the major ecological disaster (the global warming). However, this technological development generates strong continuing demand of energy sources, which has led to impose strict laws requiring rigorous control of these products.

In order to certify the fuel quality, the control of its parameters is very important in terms of production chain and commercialization. Therefore, it is necessary to apply various laboratory methods to the petroleum and the petroleum products to determine if the material is suitable for processing and (in the case of the products) for sale with a designated use in mind. In Morocco, the Ministry of Energy and Mining have fixed the quality criteria of petroleum products by requiring the determination of several quality parameters mainly based on physic-chemical methods.

This chapter presents some of the methods that are generally applied to control the quality of petroleum products in terms of chemical structures as well as physic-chemical properties but starting of course by the essential step which is the sampling techniques applied in the field of petroleum. There are, of course, many analytical methods that can be applied to the analysis of petroleum and petroleum products, but they vary with sample condition and composition.

2. Sampling

The complexity of the petroleum products brings to the question of the value of sampling: is the part sampled representative of the whole? The various chemical families of a petroleum fraction may have very different physical characteristics and the homogeneity of the whole is often due to a fragile equilibrium between its components.

Before sampling, it is necessary to mix and possibly heat the sample making sure to stay below a temperature that could cause the evaporation of the lighter components.

In the case where agitation and heating are not possible (in case of large volumes) several samples should be taken at different levels in order to reconstitute an average sample.

The procedures are described in ASTM methods (D 270, D 4057) or ISO 3170(3,24).

3. Physic-chemical analysis

Standard test methods, such as those from the American Society for Testing and Materials (ASTM), the Institute of Petroleum (IP), or other agencies, are traditionally used to determine petroleum product properties. These methods are reliable, accurate, and widely accepted but they also have some disadvantages. These methods require a large amount of sample, consume time, and may involve the use of toxic or environmentally dangerous reagents.

3.1. Color of Petroleum Products

The determination of the color of petroleum products is mostly used for manufacturing control purposes and is an important quality characteristic, since color is readily observed by the user of the product. In some cases, the color may serve as an indication of the degree of refinement of the material. When the color range of a particular product is known, a variation outside the established range may indicate possible contamination with another product. However, color is not always a reliable guide to the quality of the product.

According to the ASTM D1500 testing method this test is done by a colorimetric method using a standard light source; a liquid sample is placed in the test container and compared with colored glass disks ranging in value from 0.5 to 8.0. When an exact match is not found and the sample color falls between two standard colors, the higher of the two colors is reported(25).

3.2. Density

The density is an important quality indicator for automotive, aviation and marine fuels, where it affects storage, handling and combustion. Density is temperature-dependent; it is most often reported for oils in the units of g/ml or kg/liter, and less often in the units of kg/m^3.

Petroleum products (unless it is a residual fuel oil or asphalt) will float on water if the density of the petroleum product is less than that of the water.

According to the ASTM D1298, the measurement is done by a hydrometer. The sample is brought to a specified temperature and a test portion is transferred to a hydrometer cylinder that has been brought to approximately the same temperature. The appropriate hydrometer and thermometer, also at a similar temperature, are lowered into the test portion and allowed to settle. After temperature equilibrium has been reached, the hydrometer scale is read, and also the

temperature is taken. The observed hydrometer reading is corrected for the meniscus effect, the thermal glass expansion effect.

The imposed value by the Moroccan regulations is between 0,720 and 0,775 kg/liter at 15°C(26).

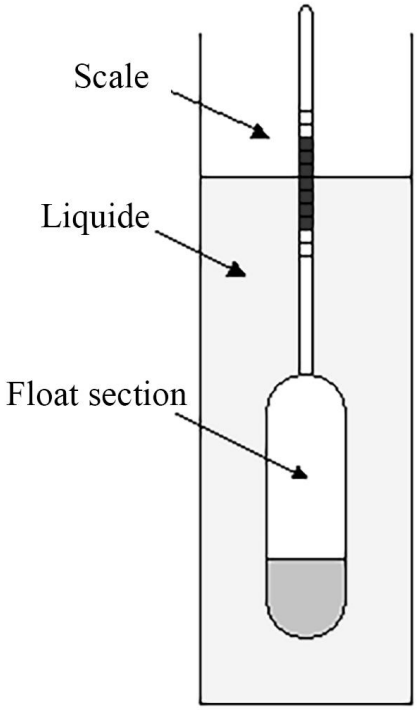


Figure 7: Hydrometer used for density determination(26)

3.3. Viscosity

The viscosity is a measure of a fluid's resistance to flow; the lower the viscosity of a fluid, the more easily it flows. Like the density, viscosity is affected by the temperature. As temperature decreases, viscosity increases.

The viscosity of petroleum fuels is important for the estimation of optimum storage, handling, and operational conditions. Thus, the accurate determination of viscosity is essential to many product specifications.

There are standard test methods for measuring the viscosity of oils, such as ASTM D445 and ASTM D4486, which demand the use of glass capillary kinematic viscometers that will produces absolute measurements in the units of centistokes (cSt)(27).

3.4. Pour point

The pour point of a product is the minimum temperature at which it still flows or the lowest temperature of its utility for certain applications under standard test conditions (ASTM D97).

To measure Pour point first heat the product to be analyzed, and then, the sample is cooled at a specified rate and examined at intervals of 3°C for flow characteristics. The lowest temperature at which movement of the specimen is observed is recorded as the pour point(28).

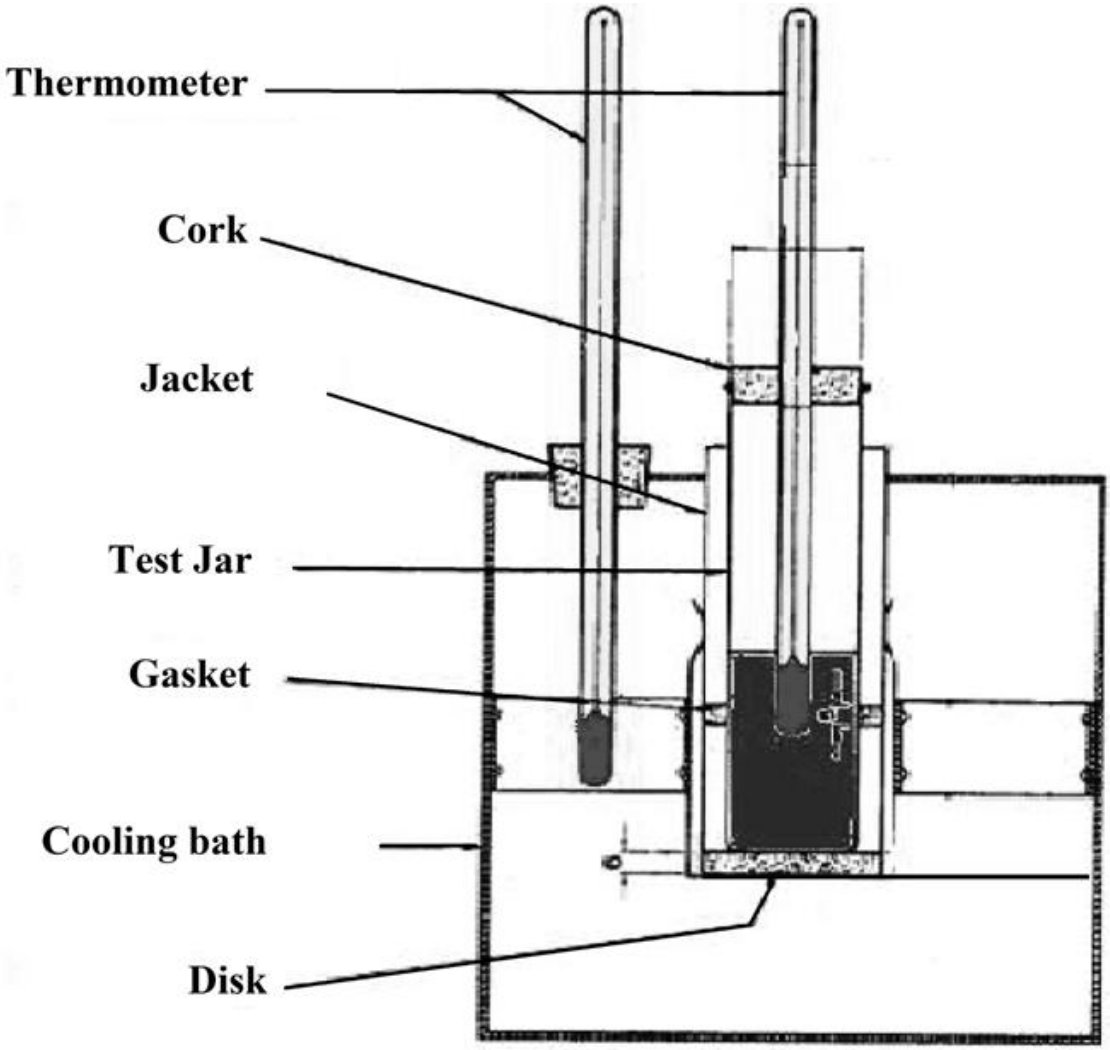


Figure 8: Apparatus assembly for the pour point test(28)

3.5. Cetane number

The cetane number evaluates the ignition ability of a fuel on a scale of 0 to 100. It is particularly important for diesel engines where the fuel must self-ignite under the effect of compression. A fuel with a high cetane number is characterized by its ability to self-ignite. This test can be performed by engine manufacturers, petroleum refiners and in commerce..., it is considered as a primary specification measurement related to the matching of fuels and engines.

The cetane number of a fuel oil is determined according to the ASTM D613 and ASTM D976 is carried out on an experimental diesel engine with variable compression ratio, called Cooperative Fuel Research (CFR) engine. By comparing the fuels combustion characteristics with those for blends of reference fuels of known cetane number under standard operating conditions(29,30).

3.6. Distillation at Atmospheric Pressure

The distillation of a petroleum product corresponds to its vaporization at atmospheric pressure. The distillation characteristics of hydrocarbons have an important effect on their safety and performance, especially in the case of fuels and solvents. The boiling range gives information on the composition, the properties, and the behavior of the fuel during storage and use. The volatility is the major determinant of the tendency of a hydrocarbon mixture to produce potentially explosive vapors.

The ASTM D86 apparatus contains a distillation flask which can contain 100 or 200 cm^3 of product which is heated and distilled at a predetermined speed. The vapors formed are condensed in a copper tube bathed in a mixture of water and crushed ice, then collected in a graduated cylinder. The operator notes the temperature of appearance of the first drop of condensate at the outlet of the tube; this is the initial point of distillation. Then, the temperature is regularly noted when 5, 10, 20... 90 and 95% of the product are distilled and collected in the test tube(31).

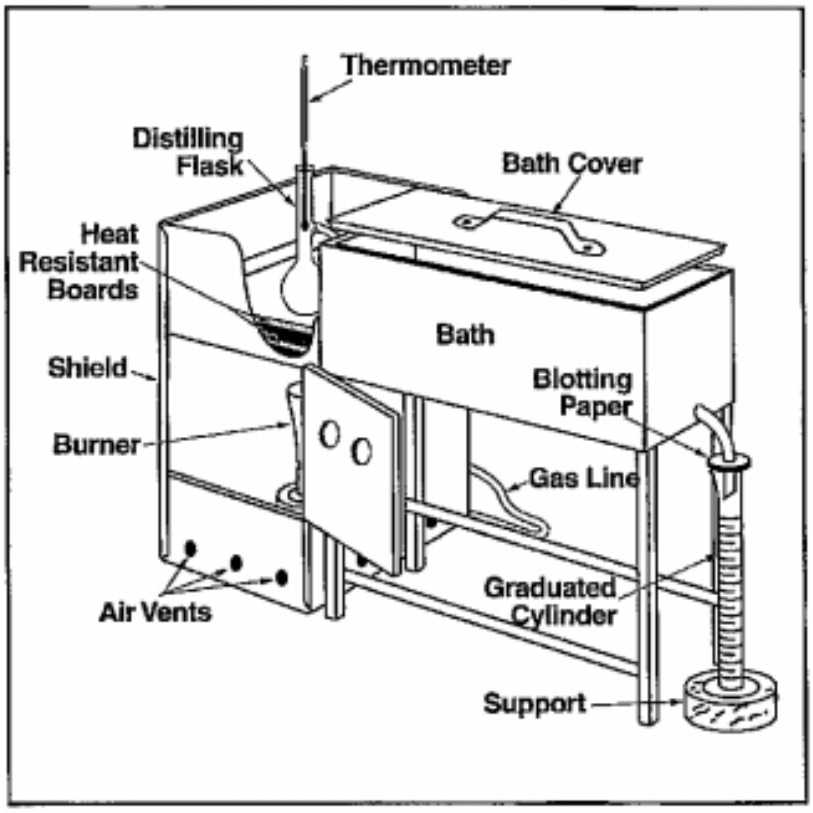


Figure 9: Apparatus assembly for the determination of distillation points of petroleum products at atmospheric pressure using a gas burner(31)

3.7. Cold Filter Plugging Point

CFPP is determined according to the ASTM D6371 methods. It represents the temperature at which a given volume of a liquid ceases to pass through, in a limited amount, a standard filtration apparatus upon cooling under standard conditions (it is supposed to represent reality in a vehicle)(32).

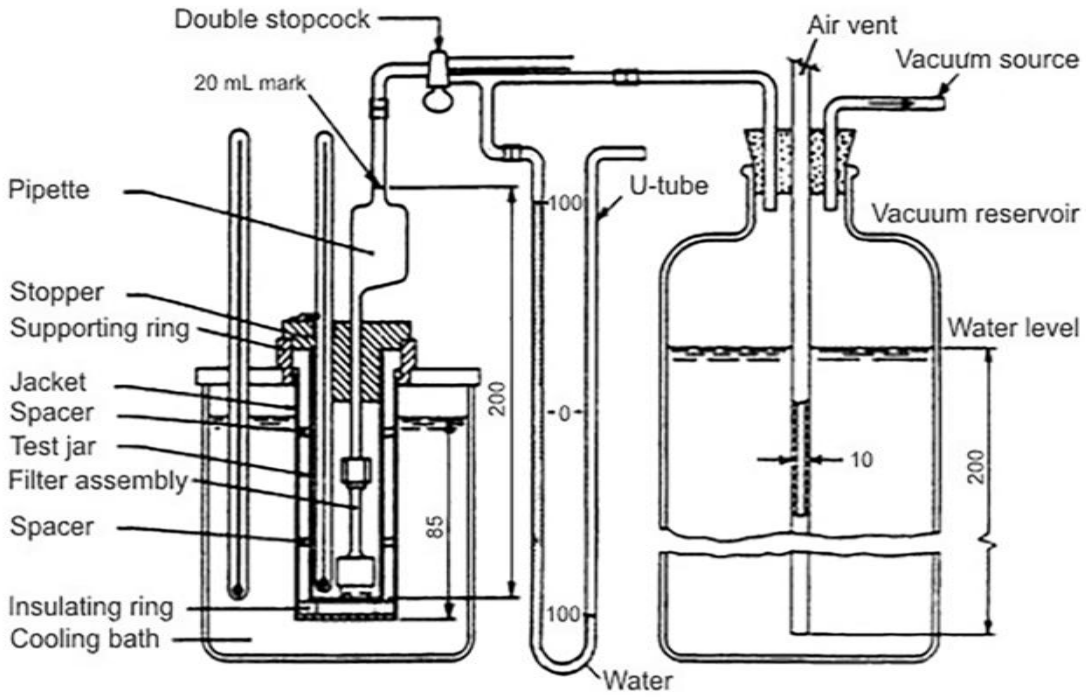


Figure 10: Apparatus assembly for the determination of the CFPP(32)

3.8. Sulfur content

Sulfur is a compound found naturally in the crude oil; therefore, it goes into refined products, such as transportation fuels, when crude oil is processed at the refinery. The SO_2 emitted by engines menaces human health, environment and engines (corrosion problem). Environmental damage to forests, crops and water resources can also result from long-term high sulfur emissions, which contribute to acid rain.

The determination of the sulfur content is done by X-ray fluorescence according to the ASTM D4294 method. The sample is placed in the beam emitted from an X-ray tube. The resultant excited characteristic X radiation is measured, and the accumulated count is compared with counts from previously prepared calibration samples (calibration curve) to obtain the sulfur concentration in mass % and/or mg/kg(33).

4. Key findings

Failure to maintain fuel quality can have disastrous consequences on a large scale, from poor performance and reduced engine life expectancy to costly fuel consumption, as well as a

detrimental impact on the environment. In the petroleum industry, as in other chemical industries, the importance of feedstock and product analyses continues to grow. Instrumental and automated methods are replacing chemical and physical methods in the laboratories.

In this chapter we have proposed the main parameters to control the quality certification of a petroleum derivative as are imposed by the international standards (ISO EN 590 and ISO EN 228).

CHAPTER III. INFRARED SPECTROSCOPY

1. Introduction

Spectroscopic techniques are considered as physical methods of characterization. The vibrational spectroscopy can be defined as the study of the interaction of electromagnetic waves on the ultraviolet, visible and infrared wavelengths and the matter.

The electromagnetic spectrum is generally divided as shown in Figure 11 in various regions as a function of the wavelength of radiation: thus, we find the gamma rays which are the most energetic, the X-rays, the ultraviolet, the visible, infrared, microwave and radio frequency waves. For the IR domain, the observed energy transitions are of the vibrational type. This domain is subdivided into three categories according to the frequency: the domain of the Near Infrared is between 13000 and 4000 cm^{-1}, between 4000 and 400 cm^{-1} is the area of the Mid IR and the far IR between 400 and 10 cm^{-1}.

The objectives of this chapter are to describe the fundamental principles of infrared spectroscopy and to study the modalities of FTIR analysis.

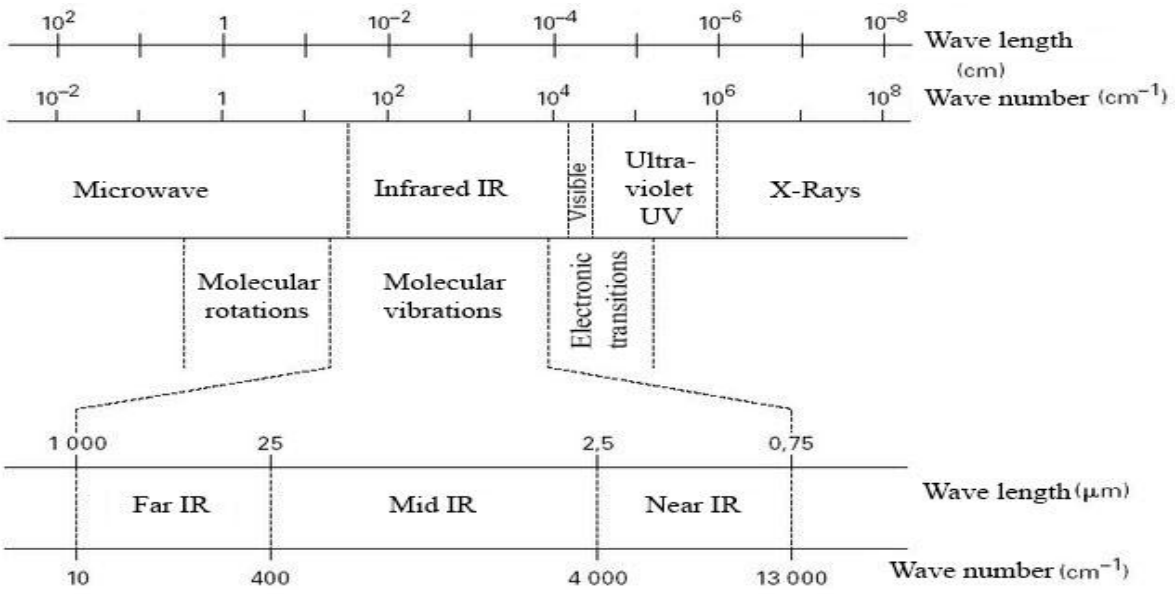


Figure 11: Spectral domains of electromagnetic radiation

2. Discovery of infrared radiation

Infrared radiation was discovered by Sir William Herschel(34). In 1800, the British astronomer Frederic William Herschel placed the thermometer in the solar spectrum to measure the

temperature of different colors. By slowly moving the thermometer from one color to the other, he noticed that the temperature increased from blue to red of the spectrum. Herschel then decided to measure the temperature beyond the red part where the eye sees nothing, we imagine that the increase in temperature would stop outside the visible spectrum. But the surprise was great: he found that the temperature was even higher. He had just discovered the first invisible light. He called these rays "invisible". These were later called "infrared rays" or infrared light(35).

3. Fundamentals of infrared spectroscopy

3.1. Electromagnetic radiation

Electromagnetic radiation has a dual nature adulatory and quantum. The main wave characteristic of the radiation is its vibration frequency ν, expressed in Hertz. The wavelength λ is the distance traveled during a complete cycle. It is related to the frequency by Equation 1:

$$\lambda = \frac{C}{\nu} \qquad \text{Eq.1}$$

With c the celerity of light ($3*10^{8}$ ms^{-1})

The quantum approach allows the description of energy interactions with matter at the molecular level. A luminous radiation behaves as if it was composed of corpuscles called photons. These photons have the property of carrying a finite amount of energy, related to the frequency of radiation by Planck's relation.

$$E = h \times \nu = h \times \frac{c}{\lambda} \qquad \text{Eq.2}$$

With h the Planck constant ($h = 6{,}626176*10^{-34}$ J.s)

3.2. Molecule’s Vibrational energy level

The atoms of a molecule do not remain in fixed position; they vibrate around an average position. In the study of infrared spectra, only vibratory movements are considered.

3.2.1. Diatomic molecule

- ***Harmonic oscillator model***

The classical approach of vibrational theory consists first of all in studying the case of the harmonic oscillator for a diatomic molecule (Figure 12). In a diatomic molecule, atoms undergo two contradictory phenomena. Indeed, there is repulsion between the negative electronic clouds. In addition, the electrons of one atom and the nucleus of the other attract each other. This results in a continuous vibration of the bond.

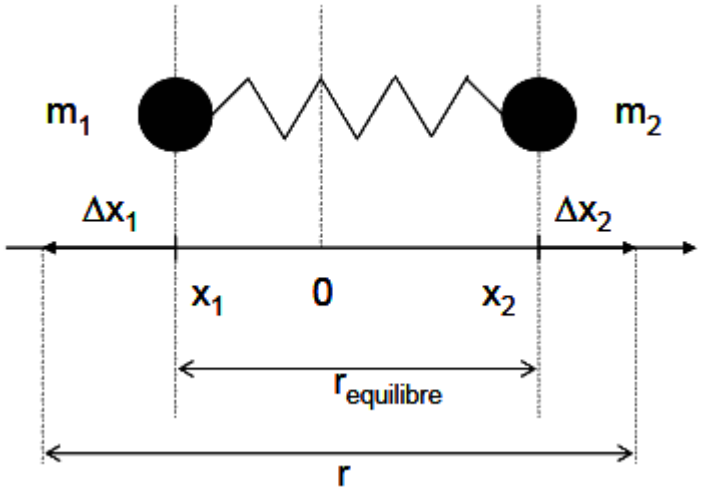


Figure 12: Oscillator Model

The resulting force of these actions goes through a minimum at the equilibrium distance (Figure 13). The potential energy (V) of this system is described by the following equation:

$$V = \frac{1}{2} k \left(r - r_{equilibrium} \right)^2 = \frac{1}{2} k x^2 \qquad \text{Eq.3}$$

With r = internuclear distance, $r_{equilibrium}$ = internuclear distance at equilibrium, $x = r - r_{equilibrium}$ = displacement, k = constant of equilibrium force (in Nm^{-1}).

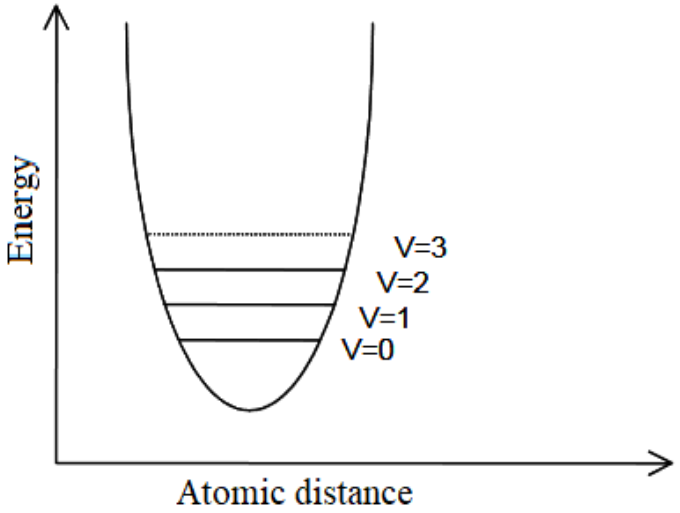


Figure 13: Harmonic potential - Representation of potential binding energy as a function of interatomic distance

Without detailing the various stages of calculation, largely explained(36), we obtain the following results. In classical mechanics, the frequency of the harmonic oscillator is given by:

$$\upsilon = \frac{1}{2\pi}\sqrt{\frac{K}{\mu}} \qquad \text{Eq.4}$$

With $\mu = \frac{m_1 \times m_2}{m_1 + m_2}$ the reduced mass, k is the force constant and m_1, m_2 the two atoms masses.

When a molecule interacts with an electromagnetic wave of frequency ν the increase in the amount of energy contained in the diatomic system results in a light absorption that can be recorded.

The quantum mechanics show that not all values of the total energy are possible at the molecular level. In 1900, Plank proposed the hypothesis according to which the energetic changes of a system can be made only by discontinuous leaps. Molecular energy takes only discrete values. The quantum equations of Schrödinger make it possible to calculate the energies of the possible vibrations for a molecule. For the simple harmonic oscillator, the solution of these equations is:

$$E_V = (v + \frac{1}{2}) \times h \times \nu \qquad \text{Eq.5}$$

With v the quantum vibrational number.

The energy level expressed in cm^{-1} is:

$$G(v) = \frac{E_v}{h \times c} = (v + \frac{1}{2}) \times \frac{1}{\lambda} \quad \text{Eq.6}$$

With $\frac{1}{\lambda}$ the wave number in cm^{-1}.

Thus only certain energies are allowed and even for v = 0, corresponding to the lowest energy, the vibration energy is not zero. Atoms cannot be completely immobile.

The Schrödinger equation also gives the conditions of variation of energy. The quantum leap can only correspond to one energy value. It is possible by absorption or emission of an amount E to go from one energy level to another. The variation of energy that moves from one state to another is:

$$\Delta E = E_{v+1} - E_v = h \times \nu \quad \text{Eq.7}$$

The allowed transition from v = 0 to v = 1 is called the fundamental transition. The other allowed transitions (passage from v = 1 to v = 2 and v = 2 to v = 3 ...) correspond to lower energy bands. They are called hot bands because the temperature increases their intensity. In the case of the harmonic oscillator, these "hot bands" have the same frequency as the fundamental transition.

The harmonic model is simple and its application is limited to the very small displacements of the atoms. Admittedly, it can give an idea of the position of the fundamental bands or the strength of the bonds, but it is not representative of the real molecules. In this model, it is accepted that the links are perfectly elastic. However, these links can break when the amplitude of the vibrations becomes important. The shape of the energy curve is therefore more complex than a simple parabola.

- ***Anharmonic model***

In order to remedy the uncertainties generated by the harmonic model and to obtain a better approximation of the vibrational states of the real molecules, the model of the anharmonic oscillator has been put in place.

Two observations show that molecules are not ideal oscillators. First, the hot bands do not have the same frequency as the fundamental bands. Secondly, harmonic transitions are allowed from v = 0 to v = 2, 3, 4 ... This deviation of the harmonic behavior can be explained by two effects:

The first is mechanical anharmonicity due to cubic effects and higher terms in the expression of potential energy

$$\mathrm{V} = \frac{1}{2} k \times x^2 + \mathrm{k}' \times \mathrm{x}^3 + \cdots \qquad \text{Eq.8}$$

With k'<<k

The second effect is called electrical anharmonicity. It is responsible for the harmonics that correspond to transitions involving 2v or 3v. As shown in the previous figure, the absorption frequencies of harmonics are not exactly multiples of the fundamental frequency.

Generally, we obtain a potential energy curve as a function of the displacement as shown in Figure 14. In quantum mechanics, the application of the Morse model in the Schrödinger equations gives the equation below.

$$\mathrm{E_v} = \left(\mathrm{v} + \frac{1}{2}\right) \times h \times \nu_a - (v + \frac{1}{2})^2 \times h \times \nu_a \times X_e \qquad \text{Eq.9}$$

With ν_a the oscillation frequency and the X_e anharmonic constant

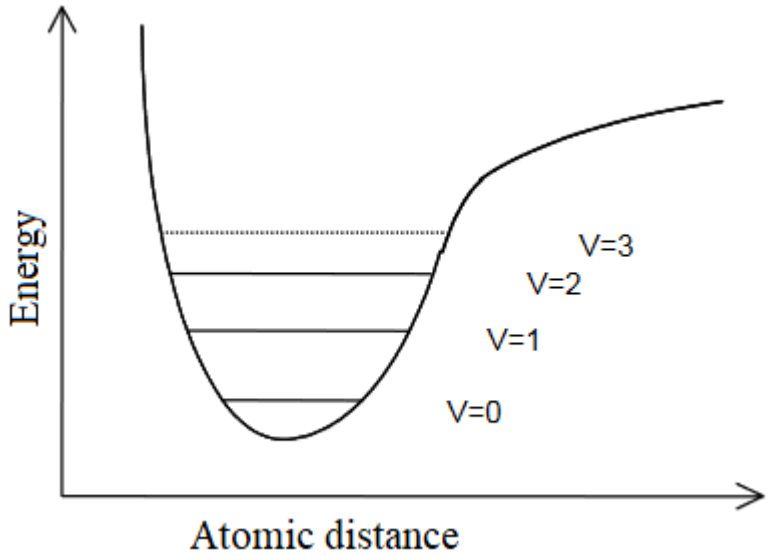


Figure 14: Potential energy curve versus displacement under the anharmonic model

In the hypothesis of a thermodynamic equilibrium, Boltzmann's law, which calculates the number of molecules at a given energy level, shows that most molecules are at the lowest energy level corresponding to v = 0. The three transitions observable under the usual conditions involve variations Δv = +1, +2 or +3. The constancy of anharmonicity being very low, the three spectral bands are at frequencies close to ν, 2ν, 3ν. The band ν is called fundamental, the bands 2ν and 3ν are called first and second harmonics.

3.2.2. Polyatomic molecules

These properties can be extended to polyatomic molecules. In this case, the molecule is considered as a set of oscillators that can interact. A molecule containing N atoms will have 3N-6 degrees of vibration freedom (3N-5 for a linear molecule). This number of degrees of freedom gives the number of fundamental vibration frequencies. A normal mode of vibration corresponds to an interatomic movement where all the atoms vibrate at the same frequency but with different amplitudes.

3.2.3. Types of vibrations

The absorption of an infrared radiation by a molecule will have the effect of vibrating it by modifying the interatomic distances or the angles of bonds. The lifetime of the excited states is

very short (of the order of 10-13s) and the whole returns to its ground state by restoring the energy absorbed in the form of heat. There are two types of vibration:

- **Elongation**: it is a vibration between two given atoms during which the interatomic distance varies according to the axis of the connection.

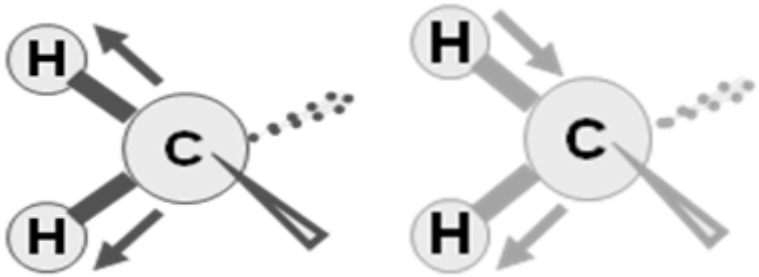


Figure 15: Molecular vibration modes: symmetric (purple) and asymmetric elongation (orange)

- **Deformation**: it is a vibration that changes the shape of the molecule by causing a modification of the binding angles.
  - Deformation in the molecular plane.

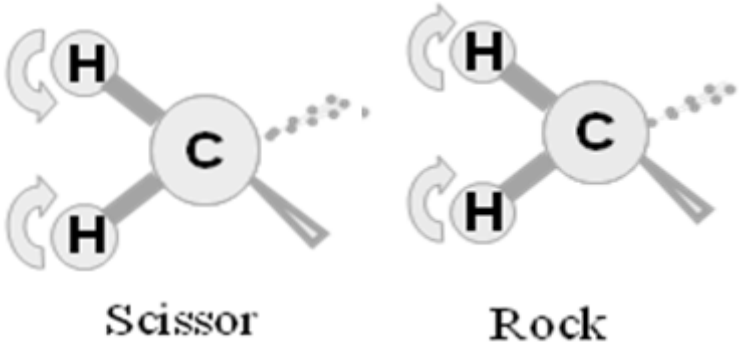


Figure 16: Modes of molecular vibration- deformation in the plane of the molecule

  - Deformations out of the molecular plane.

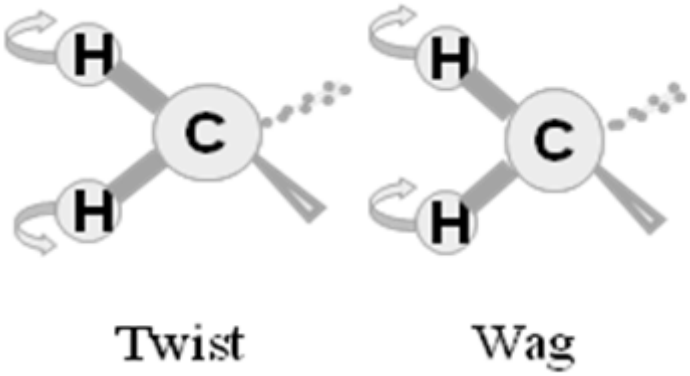


Figure 17: Modes of molecular vibration- deformation out of the plane of the molecule

Both types of vibration elongation and deformation can be symmetrical or asymmetrical. Generally, elongations are located in a range of wave numbers greater than that of deformations(37).

4. Techniques of samples examination in spectroscopy

If the material is exposed to an electromagnetic radiation, for example, infrared light, the radiation may be absorbed, transmitted, reflected or diffused.

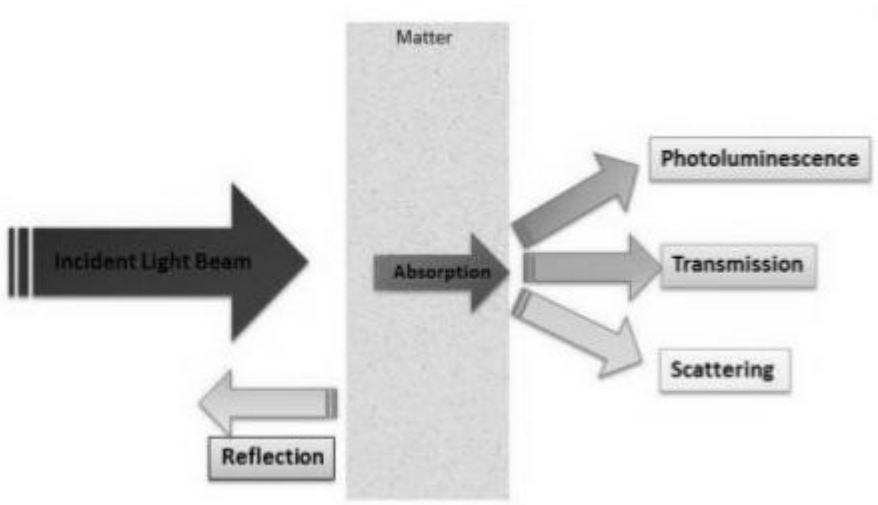


Figure 18: Interaction of radiations with the matter

4.1. Absorbed light

The absorption is a probabilistic phenomenon that depends on the number of incident photons (the intensity of the incident ray) and the number of absorbers (the molecules in the sample).

Therefore, if there are no absorbent species, no photon will be absorbed. On the other hand, if the concentration of the absorbing species becomes infinite, no more photon will arrive at the detector and the environment will be opaque. For this reason the absorptions (or transmissions) are expressed in percentage, the first case corresponding to an absorption of 0% (or a 100% transmission), the second case corresponds to a 100% absorption (or a 0% transmission). For intermediate cases, the absorption will vary between 0% and 100% depending on the concentration, but also according to the wavelength of the photons. The absorption and scattering of light take place at the atomic and molecular level of the sample. In order for the energy transmitted by light to be absorbed by the sample, it must correspond to the energy available between the states of the atoms or molecules. The probability that the light of the incident beam interacts with the material is related to the cross section σ. It is considered as an effective surface of the atom or the molecule, it takes into account the absorption and diffusion of the light.

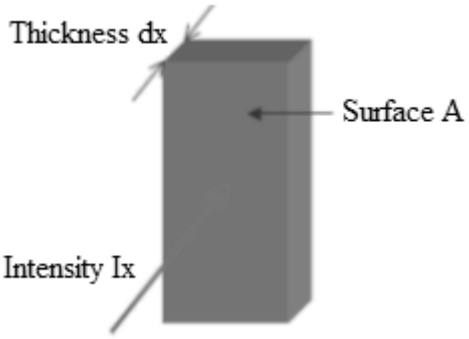


Figure 19: An incident light on a surface A of a material of thickness dx and the concentration of the molecules C

The number of molecules illuminated by the incident light Ix is CAdx. The total area "effective" whose molecules are present is σCAdx.

The probability that the light is absorbed or diffused in the thickness dx is:

$$-\frac{dI_x}{I_x} = \frac{\sigma\, CA}{A} d_x = \sigma\, C d_x \qquad \text{Eq.10}$$

Where dI_x is the change of intensity through dx.

Now let's integrate both sides

$$-\frac{\mathrm{dI_x}}{\mathrm{I_x}} = \frac{\sigma\,\mathrm{CA}}{A}\mathrm{d_x} = \sigma\,\mathrm{Cd_x}$$

$$-\int_{I_0}^{I}\frac{\mathrm{dI_x}}{\mathrm{I_x}} = -\int_{I_0}^{I}\sigma\,\mathrm{Cd_x}$$

We obtain:

$$ln\,I - ln\,I_0 = ln\frac{I}{I_0}$$

From we have: $I = I_0 e^{-\sigma\ CX} = I_0 e^{-\mu X}$

The coefficient $\mu = \sigma\,C$ is the coefficient of the linear attenuation.

It can be seen that the intensity of the light decreases exponentially with the depth in the material. The unit of the linear attenuation coefficient μ is generally expressed in cm^{-1}.

μ is a wavelength function μ = μ (λ). So the law of Beer-Lambert is also a function of λ.

$$I(\lambda) = I_0(\lambda)e^{-\mu(\lambda)X} \qquad \text{Eq.11}$$

4.2. Transmittance

If $I_0(\lambda)$ represents the intensity of the incident ray and $I(\lambda)$ the intensity of the light transmitted (Figure 20), the transmittance T (or the transmission percentage) of a solution is defined as follows:

$$T = \frac{I(\lambda)}{I_0(\lambda)} = e^{-\mu(\lambda)X} \qquad \text{Eq.12}$$

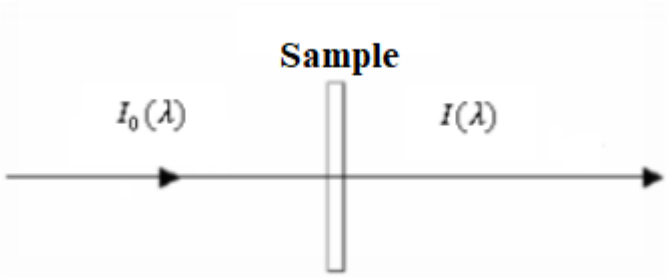


Figure 20: Transmittance of electromagnetic radiation

Note that the quantity called absorbance A is defined as follows:

$$A = \log\frac{1}{T}$$

$$A = \log\frac{I_0(\lambda)}{I(\lambda)} = \log\frac{I_0(\lambda)}{I_0(\lambda)e^{-\mu(\lambda)X}} = \log e^{\mu X}$$

$$A = \mu X \log e = 0{,}4343\mu X$$

Let the extinction coefficient ε be such that: ε = 0.4343 σ

$$A = \varepsilon.\mathrm{C}.\mathrm{X} \qquad \text{Eq.13}$$

Note that when radiation is reflected in the interaction with the material, the Beer-Lambert law is not applicable(38).

4.3. Reflective processes

As discussed in the preceding paragraphs, an infrared spectrum is generally obtained by passing infrared radiation through a sample and determining the energy fraction of the incident radiation absorbed. So the infrared beam passes through the sample and the energy is measured.

Reflection IR spectroscopy was developed using the combination of IR spectroscopy with the reflection theories. In reflection spectroscopy techniques, the absorption properties of a sample can be extracted from the reflected light. These reflection techniques can be used for samples that are difficult to be analyzed by the conventional transmission method(39). The reflection depends on the nature of the surfaces, polished, smooth as a mirror, rough and that tend to

reflect light in all directions. When electromagnetic radiation arrives at the interface of a second material, it interacts with the surface of the sample, depending on the characteristics of the surface and its environment (the refractive index). Generally, reflection techniques can be divided into two categories: internal reflection and external reflection(39). In the internal reflection process, the interaction of the electromagnetic radiation at the interface between the sample and the material with a higher refractive index is studied, while the external reflection techniques come from the reflected radiation by the surface of the sample. The External reflection covers two different types of reflection: specular reflection and diffuse reflection. In practice, all three types of reflection can occur at the same time but with different contributions.

5. Spectroscopic instruments

There are two types of IR spectrophotometers:

- The classical infrared spectrophotometer.
- The Fourier Transform Infrared Spectrometer.

Their differences are mainly in the wavelength selector system.

5.1. Conventional infrared spectrometers: dispersive spectrometers

The main elements of the standard IR instrumentation consist of 4 pieces(40)

- A light source.
- A dispersive element, diffraction by a grid or a prism.
- A detector.
- An optical system of mirrors.

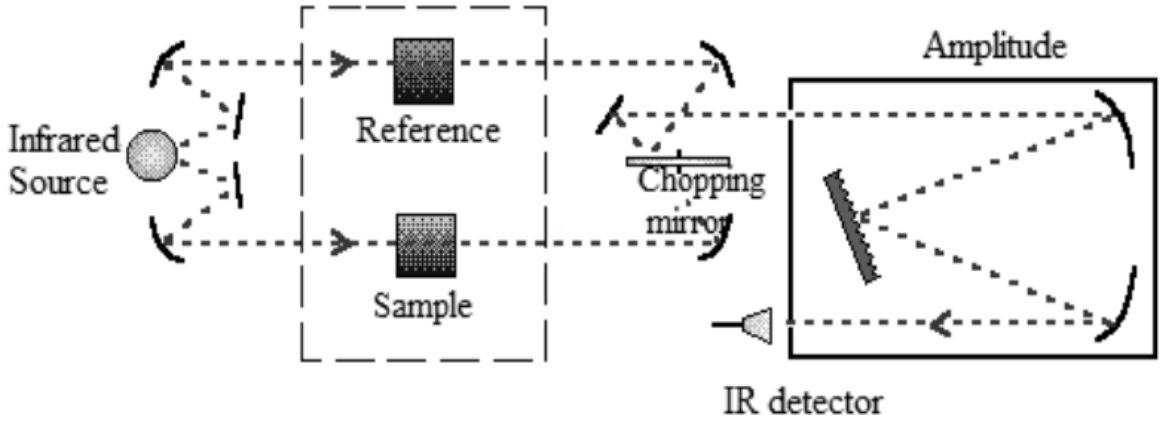


Figure 21: A schematic diagram of the classical dispersive IR spectrophotometer

The infrared radiation from the source is separated by reflection on mirrors. The radiation passes through a monochromator both by a reference sample and then through the sample. The beams are reflected on a rotating mirror, which alternates the passage of beams that come from both the sample and the reference sample to the dispersion element and finally to the detector that measures the amount of energy for each frequency that passes through the sample, in order to obtain the spectrum. The disadvantages of these devices are:

- The slowness of measures.
- The mechanical complexity.

5.2. Fourier Transform Infrared Spectroscopy

Dispersive infrared spectrometer has many limitations because it examines component frequencies individually, resulting in slow speed and low sensitivity. Fourier Transform Infrared (FTIR) spectrometer is preferred over dispersive spectrometer, since it is capable of handling all frequencies simultaneously with high throughput, reducing the time required for analysis(41).

Every FTIR analyzer is based on an interferometer. With regard to the Michelson interferometer, the incident infrared radiation is sent on a semi-transparent mirror (the separator) inclined at 45°, reflecting a part of the beam and passing the other. The reflected beam is then sent to a fixed mirror while the transmitted radiation is sent to a moving mirror. The two mirrors are placed in a perpendicular position to the beams they receive. The two beams are then returned in the incident direction and are recombined. During the recombination, there is an

interference resulting from the phase shift directly dependent on the position of the moving mirror. The finally obtained signal is an interferogram containing all the contributions of the wavelengths of the incident beam, making it possible to make an overall analysis.

The modulated infrared radiation then passes through the sample to be analyzed and is then sensed by a detector that delivers a signal proportional to the intensity received. Since the purpose of a FTIR analysis is to know which IR frequencies have been absorbed by the sample, after the signal has been collected by the detector, the Fourier transform of the interferogram reproduces the entire absorption spectrum (Figure 22).

In order to select only the characteristic information of the sample, it is essential to record before the sample the single beam spectrum of the source also called reference spectrum or background. The latter represents the absorptions caused by all substances other than the sample (carbon dioxide, water vapor ...).

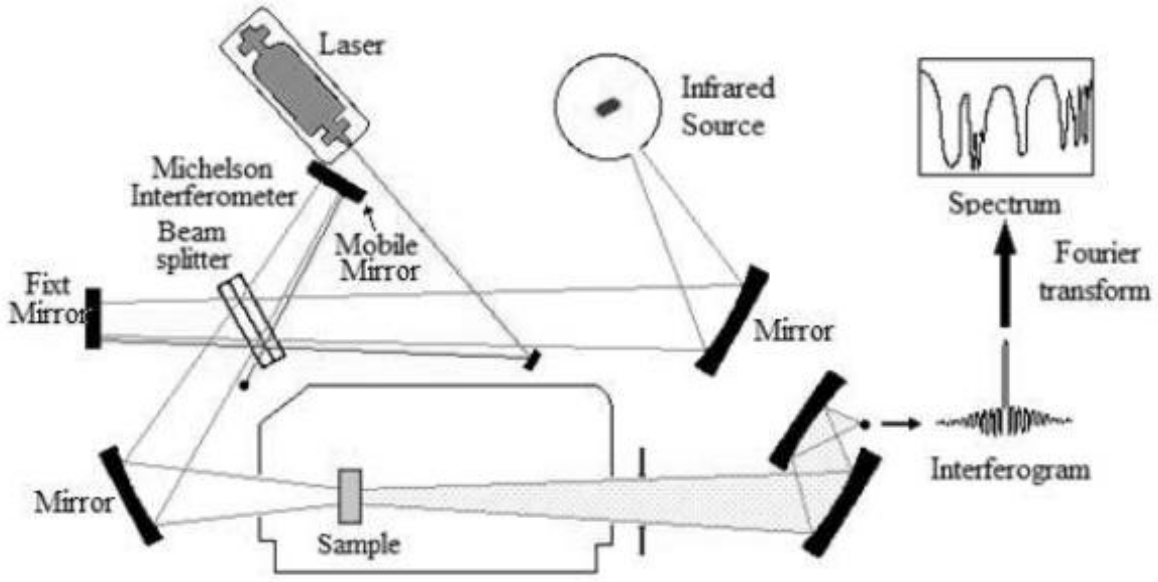


Figure 22: Schematic illustration of a modern FTIR Spectrophotometer

6. Attenuated Total Reflectance (ATR)

An attenuated total reflection accessory operates by measuring the changes that occur in a totally internally reflected infrared beam when the beam comes into contact with a sample.

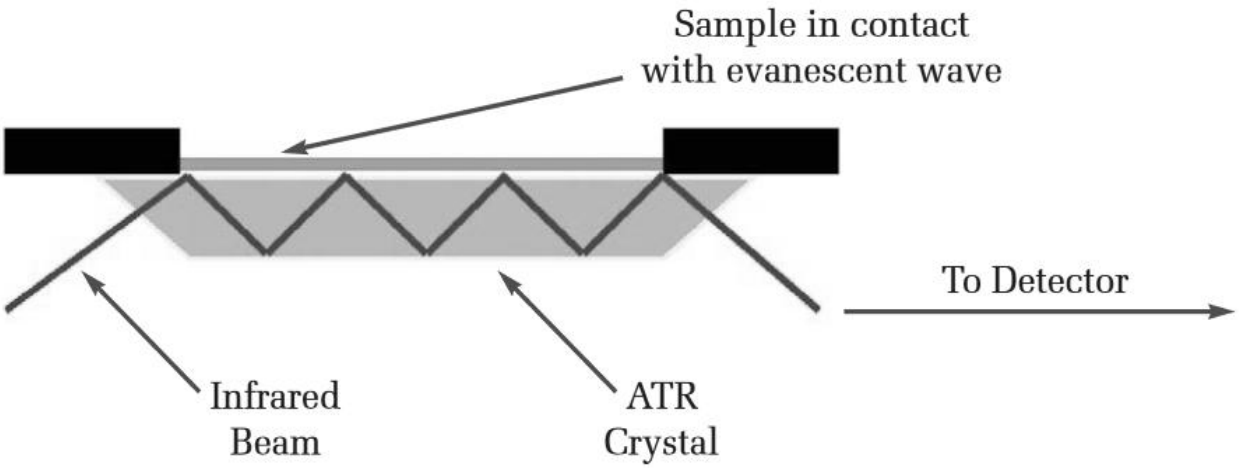


Figure 23: Principle of the ATR system

An infrared beam is directed onto an optically dense crystal with a high refractive index at a certain angle. This internal reflectance creates an evanescent wave that extends beyond the surface of the crystal into the sample held in contact with the crystal. It can be easier to think of this evanescent wave as a bubble of infrared that sits on the surface of the crystal. This evanescent wave protrudes only a few microns (0.5 μ - 5 μ) beyond the crystal surface and into the sample. Consequently, there must be good contact between the sample and the crystal surface. In regions of the infrared spectrum where the sample absorbs energy, the evanescent wave will be attenuated or altered. The attenuated energy from each evanescent wave is passed back to the IR beam, which then exits the opposite end of the crystal and is passed to the detector in the IR spectrometer. The system then generates an infrared spectrum. For the technique to be successful, the following two requirements must be met:

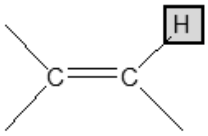
- ✓ The sample must be in direct contact with the ATR crystal, because the evanescent wave or bubble only extends beyond the crystal 0.5 μ - 5 μ.
- ✓ The refractive index of the crystal must be significantly greater than that of the sample or else internal reflectance will not occur the light will be transmitted rather than internally reflected in the crystal. Typically, ATR crystals have refractive index values between 2.38 and 4.01 at 2000 cm^{-1}. It is safe to assume that the majority of solids and liquids have much lower refractive indices.

7. Qualitative interpretation of the infrared spectrum

The principle of infrared spectroscopy is based on the theory that different molecules with different combination of atoms produce their unique spectra. Since, all groups have their characteristic vibrational frequencies; infrared spectroscopy can be used to qualitatively identify substances(42).

Therefore it is also possible to correlate the spectrum to the chemical structure as shown in table 7. This table is a summary; it shows only the most important bands of the MIR domain.

Table 7: Characteristic wavelengths of the main chemical groups in MIR

Bond	Type of Compound	Frequency Range, cm^{-1}	Intensity
C-H	Alkanes	2850-2970	Strong
C-H	Alkenes 	3010-3095 675-995	Medium strong
C-H	Alkynes 	3300	Strong
C-H	Aromatic rings	3010-3100 690-900	Medium strong
O-H	Monomeric alcohols, phenols Hydrogen-bonded alchohols, phenols Monomeric carboxylic acids Hydrogen-bonded carboxylic acids	3590-3650 3200-3600 3500-3650 2500-2700	Variable Variable, sometimes broad Medium broad
N-H	Amines, amides	3300-3500	medium
C=C	Alkenes	1610-1680	Variable
C=C	Aromatic rings	1500-1600	Variable
C≡C	Alkynes	2100-2260	Variable
C-N	Amines, amides	1180-1360	Strong
C≡N	Nitriles	2210-2280	Strong
C-O	Alcohols, ethers, carboxylic acids, esters	1050-1300	Strong
C=O	Aldehydes, ketones, carboxylic acids, esters	1690-1760	Strong

8. Quantification with infrared spectroscopy

The quantification by infrared is made based on the Beer's Law as mentioned in the equation 13:

$$A = \varepsilon . C . X \qquad \text{Eq.13}$$

With A is the Absorbance, ε is the molar extinction coefficient or the absorptivity, l is the Path length and c is the Concentration of the sample.

The height or area of a peak in an absorbance spectrum is proportional to concentration, which is why Beer's Law can be used to determine the concentrations of molecules in samples.

The y-axis of an infrared spectrum can also be plotted in units called percent transmittance (%T), which measures the percentage of light transmitted by a sample. %T spectra are calculated as follows:

$$\%T = 100 \times \left(\frac{I_0}{I}\right) \qquad \text{Eq.14}$$

Where %T is the Percent Transmittance, I_0 is the Intensity in the background spectrum and I is the Intensity in the sample spectrum.

9. Key findings

Infrared spectroscopy is a means of diagnosis to determine the nature of the chemical bonds present in a molecule. Thus, infrared spectroscopy is a very powerful means of characterization to identify molecular groups and obtain a lot of microscopic information about their conformation and their possible interactions.

The Infrared spectra contain a lot of useful chemical information, not only the interpretable bands, but also fingerprints that can be extracted by other tools such as chemometric tools that we will see in the next chapter.

CHAPTER IV: CHEMOMETRICS

1. Introduction

Chemometrics, born around 1970, refers to all the mathematical and statistical techniques applied to chemistry(43). Chemometric techniques bring together all the computer, statistical and mathematical methods adapted to the specific needs of chemists and which allow maximum exploitation of the data collected.. In contrast to the "height" or "surface" methods of the bands, the algorithms used in multivariate analysis, based on linear global methods of data condensations, offer the means of reducing the number of experimental data available (too much) to a restricted digital data by keeping from the mass of information contained in a spectrum only the most significant combinations for the analysis. This new concept of working in a factor space rather than spectral data performs filtering and attenuates uncorrelated component noise.

Chemometric tools, based on their objectives and in relation to the data on which they will need to be applied, can be classified into two principal classes.

Unsupervised methods or data exploration tools use only spectral data, for example. Only measurements, no information about the state of the system. The goal is to better understand how the data is structured.

Supervised methods are methods that use in addition to the spectral data at least one quantitative or qualitative information: thus information related to the state of the system. They can be divided into two categories:

- Discrimination and classification techniques.
- Regression and prediction techniques.

The purpose is to build a model that predicts the state from measurements. For this, a first set or a calibration set is necessary. This model will then be applied to predict the state of new samples contained in a second set, or a test set.

The aim of this chapter is to present the mathematical tools used for infrared spectra preprocessing, data visualization, samples classification and the quantification of some parameters from spectra.

1.1. Definitions

1.1.1. Univariate data

Univariate data are the data formed by a single measure or a single variable that defines the property of the sample, their processing is simple, for example (pH, density, viscosity...).

1.1.2. Multivariate data

The multivariate data is data formed by several measurements or several variables simultaneously that define the property of the sample, their processing is not simple, for example the infrared spectrum.

1.2. Mathematical preprocessing applied to the spectral data

Several phenomena can affect the infrared spectrum including the particle size of the sample(44) and by variations of the optical path(45). This is why it is necessary to have a well-defined sample preparation and analysis protocol(46). To eliminate or reduce these interferences, mathematical pre-treatments are applied to the spectra. The most commonly used treatments are described in this section.

1.2.1. Mean Centering

One of the most common preprocessing methods, mean-centering calculates the mean of each column and subtracts this from the column. Another way of interpreting mean-centered data is that, after mean-centering, each row of the mean-centered data includes only how that row differs from the average sample in the original data matrix.

Mean centering has the effect of including an adjustable intercept in multivariate models. For example, mean-centering both the X and Y blocks in a regression model effectively allows for a non-zero intercept of the regression line. This is critical in many inferential regression problems where the intercept is not necessarily at a Y of zero when X goes to zero(47).

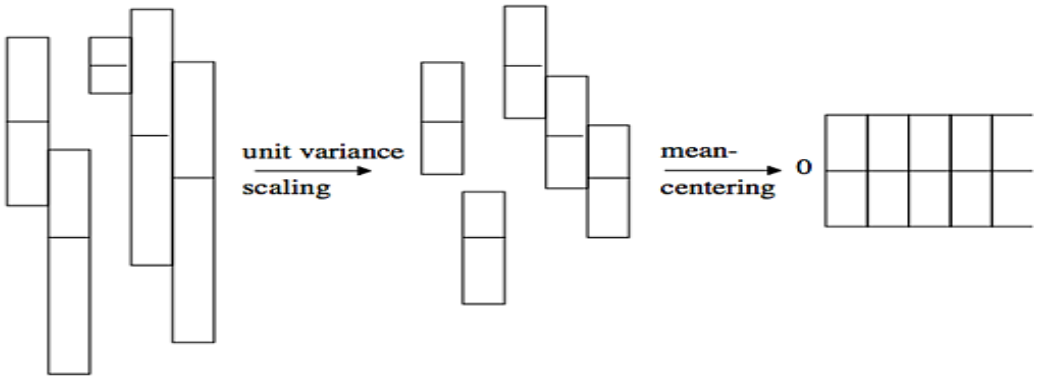


Figure 24: After mean-centering all variables will have equal "length" and a mean value=zero

1.2.2. Derivative

The derivative has historically been the first preprocessing used. It reduces the drift of the baseline(48),to separate more clearly the absorption bands(49) and to highlight some parts of the spectral information(50).

The simplest form of derivative proposed by Savitzky-Golay is a point-difference first derivative, in which each variable (point) in a sample is subtracted from its immediate neighboring variable (point). This subtraction removes the signal which is the same between the two variables and leaves only the part of the signal which is different. When performed on an entire sample, a first derivative effectively removes any offset from the sample and de-emphasizes lower-frequency signals. A second derivative would be calculated by repeating the process, which will further accentuate higher-frequency features(51).

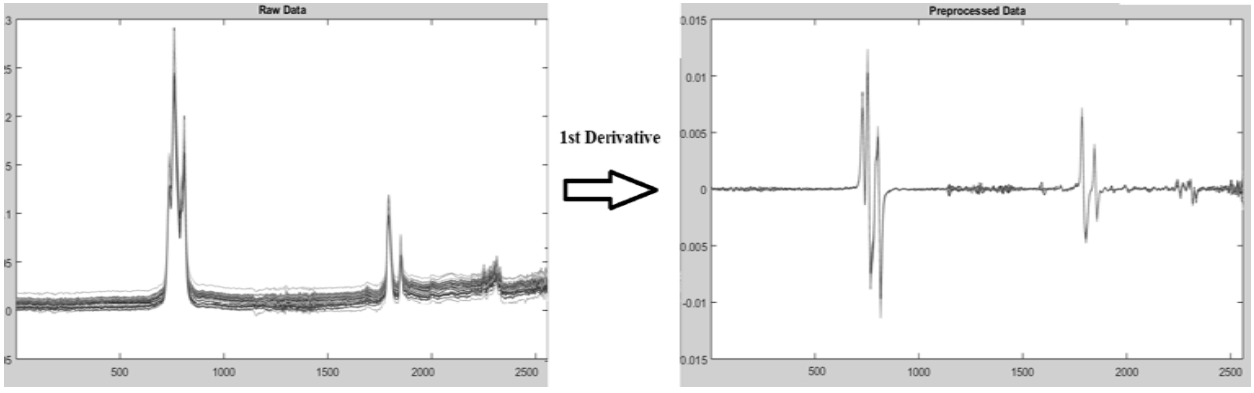


Figure 25: the effect of 1st derivative on FTIR spectra

1.2.3. Smoothing

Smoothing is a low-pass filter used for removing high-frequency noise from samples. Often used on spectra, this operation is done separately on each row of the data matrix and acts on adjacent variables. Smoothing assumes that variables which are near to each other in the data matrix (i.e., adjacent columns) are related to each other and contain similar information which can be averaged together (mobile average) to reduce noise without significant loss of the signal of interest(51).

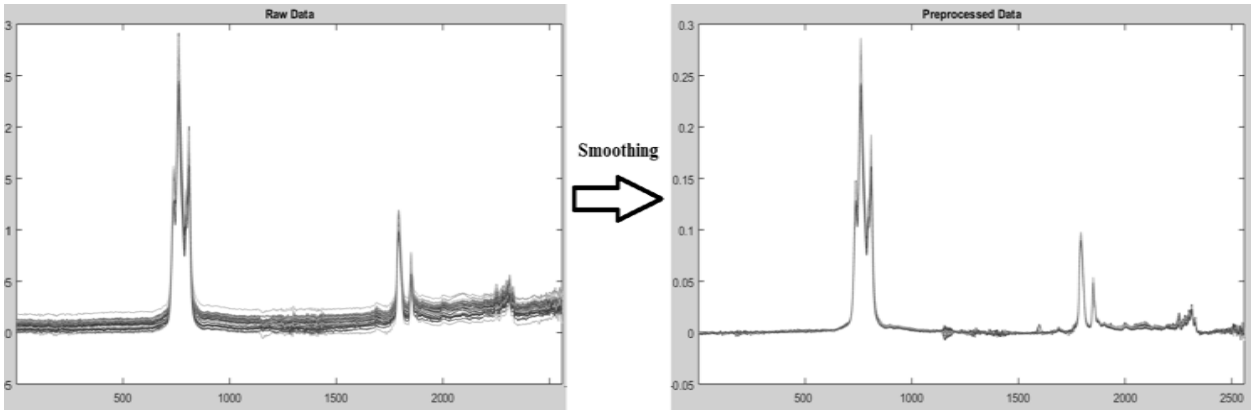


Figure 26: the effect of smoothing preprocessing on FTIR spectra

1.2.4. Normalization

The normalization of a spectrum consists in dividing each spectral value x_λ of a spectrum x by a representative number a(52).

$$xnorm_\lambda = \frac{x_\lambda}{a} \qquad \lambda = 1.....\lambda_{max} \qquad \text{Eq.15}$$

"a" can be calculated as follows:

- The maximum value:

$$a = max\ (x_\lambda) \qquad \lambda = 1.....\lambda_{max} \qquad \text{Eq.16}$$

- The sum of all variables from the spectrum x

$$a = \sum_{\lambda=1}^{\lambda_{max}} (x_\lambda) \qquad \text{Eq.17}$$

This normalization is also called "surface unit".

- The sum of squares of all variables from the spectrum x

$$a = \sqrt{\sum_{\lambda=1}^{\lambda_{max}} (x_\lambda)^2} \quad \text{Eq.18}$$

This normalization is also called length unit.

Note that when there is an unknown variation (offset) between the spectra it is useful to subtract the average value x from spectra of each spectral variable xλ.

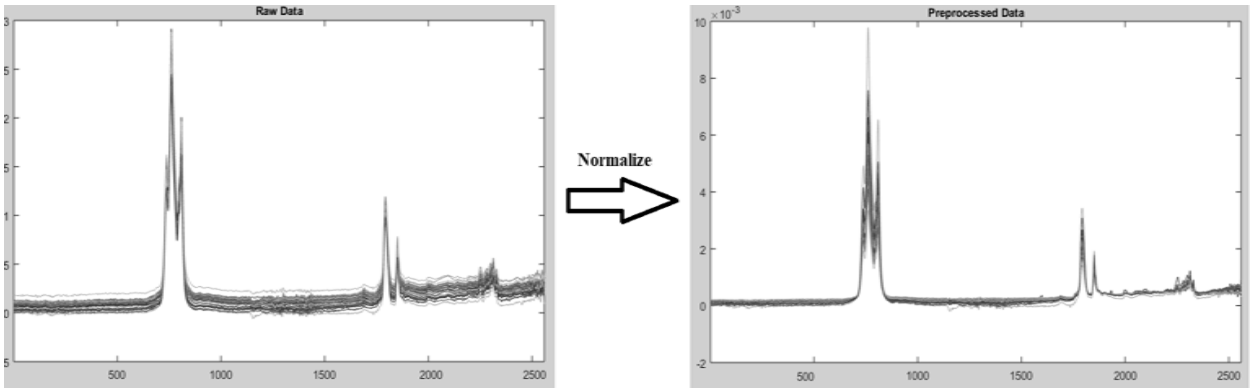


Figure 27: the effect of the normalization preprocessing on FTIR spectra

1.2.5. Standard Normal Variate

In many situations, the spectra exhibit uncontrolled variations in overall intensity depending mainly on the length of the optical path through the measurement vessel or within the sample. When this journey is very short, it is difficult to have perfectly reproducible acquisition conditions. The SNV pretreatment reduces strongly the variations of general intensity of the spectra. It allows to establish a baseline to the collection of spectra until it is inexistent(53). The SNV consists in dividing the value of each variable of a spectrum by a representative number of the general intensity of this spectrum. The SNV is most often applied where each spectrum is also scaled by its own standard deviation.

Spectral data are centered and reduced according to the following equation:

$$x_{SNV;i} = \frac{x_i - \bar{x}}{\sqrt{\frac{\sum_{i=1}^{i=p}(x_i - \bar{x})^2}{p-1}}} \quad \text{Eq.19}$$

With:

- x_i: Signal intensity at the wavelength i

  P: The number of wavelengths

- $x_{SNV;i}$: The corrected value at wavelength i.

- $\bar{x}$: The average value of the signal with:

$$\bar{x} = \sum_{\lambda=1}^{\lambda_{max}} (x_\lambda) \qquad \text{Eq.20}$$

The main advantage of the SNV method is due to its application to each taken spectrum separately, without reference to the samples of the calibration set. Several studies have shown that the use of SNV preprocessing allows the improvement of the results of quantitative analyzes(54).

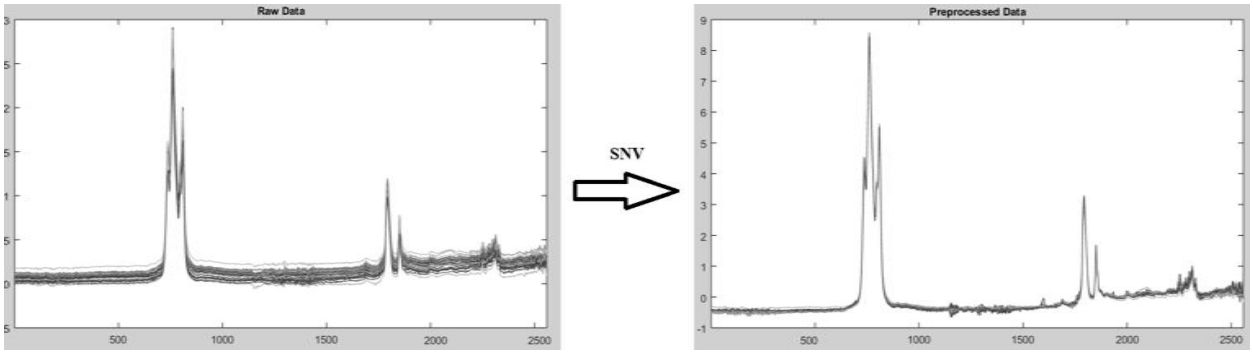


Figure 28: the effect of the application of the SNV on a group of spectra

1.2.6. Multiplicative Scatter Correction

The MSC was developed to eliminate multiplier and additive effects due to light scattering in the sample. It allows improving the linearity of the relationship between absorbance and concentration of variables in samples(55).To perform the MSC correction, it is necessary to have a reference spectrum. The average of the spectra of the calibration batch is considered as a good approximation it is used by default(56).

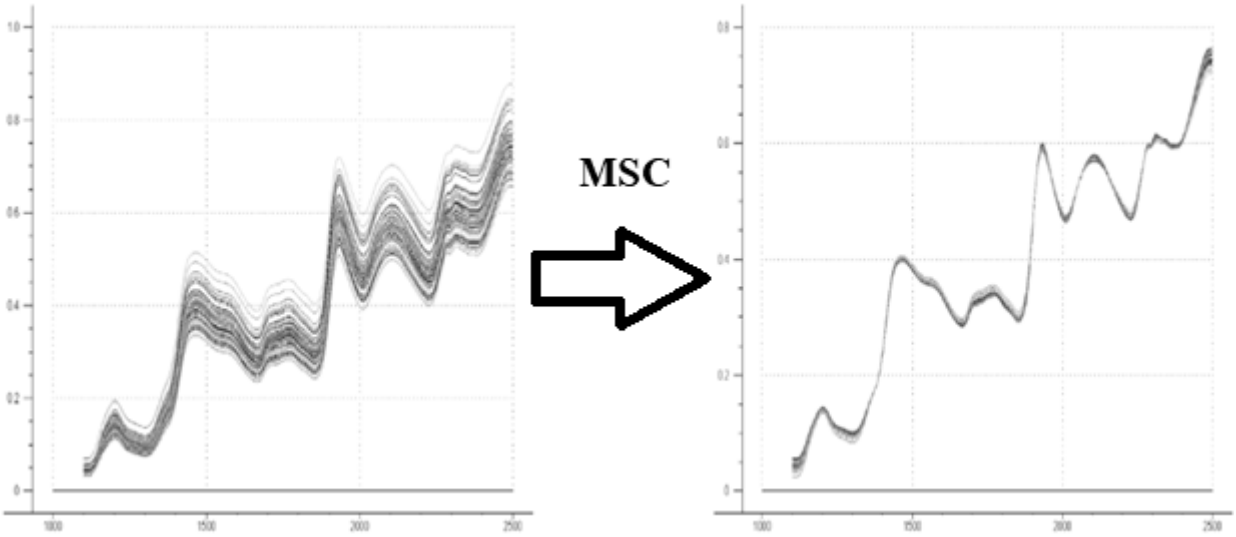


Figure 29: the effect of the application of the Multiplicative Scatter Correction on a group of spectra

2. Chemometric methods

2.1. Unsupervised method: the Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is an unsupervised multivariate analysis technique used to reduce the size of data dimensions without losing information about the samples. This allows the data to be analyzed and examined in a more conceptual way(57,58).

2.1.1. History of PCA

Principal Components Analysis was introduced for the first time in the statistics by Pearson, who formulated the analysis to find the lines and the adjustment planes nearest a system of points in space. PCA has been mentioned by Fisher and Mackenzie(59)as a variance analysis for modeling response data, they also emphasized the algorithm NIPALS, rediscovered later by Wold(60) It is Hotelling(61) who developed the PCA at its current stage. Since then, its usefulness has been rediscovered in many scientific fields. It is known by many names: singular value decomposition (SVD) is used in numerical analysis(62). The expansion of Karhunen-Loéve is used in electrical engineering(63), vector analysis (Eigenvector) is often used in the physical sciences and the Hotelling transformation used in the analysis of the image.

In chemistry, the PCA was introduced by Malinowski about 1960 under the name of the Principal Factorial Analysis, and after 1970 a large number of chemical applications were published(64).

2.1.2. Principle of the ACP

The PCA is based on the transformation of a group of original quantitative variables into a new variable space formed of the linear combinations of the initial variables thus reducing the space. A spectrum of 1000 wavelengths can be represented by a point in the new space. These combinations are called principal components PCs, That must be interpreted independently of each other, because they contain a part of the variance that must not be expressed in any other principal component(64).

From a geometric point of view, the PCA is considered as a method of data rotation so that the observer is the best placed to understand the relationships between the individuals, these individuals are represented in a new system of axis formed by the principal components. The first axis is chosen in the direction of the greater dispersion (greater variability) of the cloud of points. The second axis is orthogonal to the first to avoid having the same information in the second principal component. See Figure30. Each newly created axis defines a direction that describes some of the global information. Suppose that the first principal component represents 65% of the total variance, and that the second principal component represents 20%, the projection in the plane of the first two principal components will thus make it possible to visualize 85% of the total information.

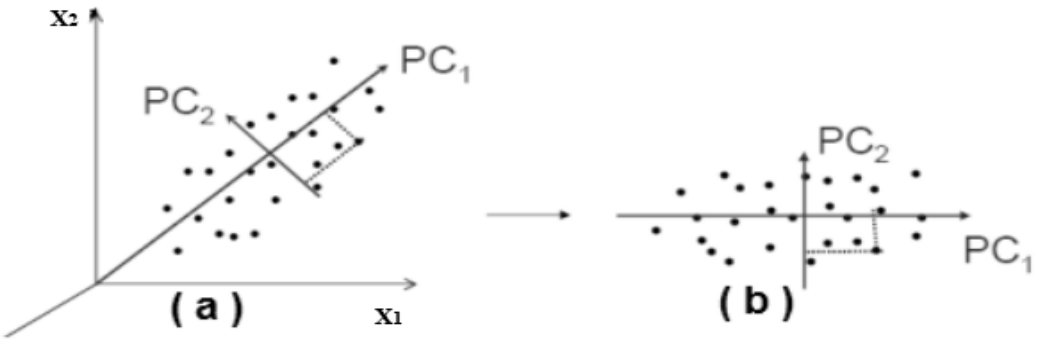


Figure 30: (a) initial representation of the data, (b) representation in the new PC1 vs PC2 space

2.1.3. Theory of the PCA model

The method consists in decomposing an X data matrix into a portion of the structure and part of the noise according to the equation.

$$X = T.P^T + E \qquad \text{Eq.21}$$

With X (n, p), the original data matrix consists of n rows (objects) and p columns (variables). T (n, k) is the matrix of scores with n rows and q columns (number of principal components), it describes the samples in the space of principal components. P (p, k) is the matrix of loadings with p rows and q columns, it describes the original variables in the PCs space. P^T (k, p) is the transposed matrix of P. E (n, p) is the matrix of residuals between the original data and their projections(65,66). So X is the sum of two matrices: the $T.P^T$ matrix called structure has a high percentage of the variance and the residual matrix E that collects the noise(67–69). Thus the matrix decomposition of the PCA allows to obtain matrices of the factorial coordinates T called scores and factorial contributions P called loadings and the gaps E which contain a low value of the residual variability in order to describe the behavior of the samples see the figure31.

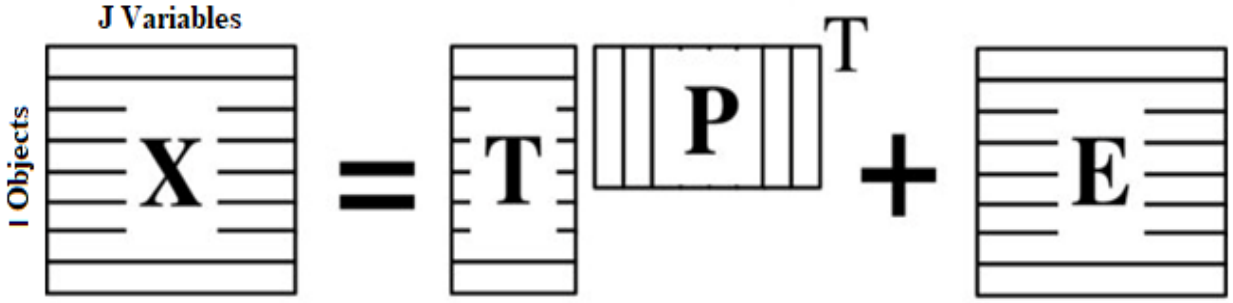


Figure 31: Matrix decomposition of the PCA

The coordinates T and the factorial contributions P respectively contain all the important information in relation with the objects and the variables. The E deviations contain the residual variability.

2.1.4. Mathematical approach

- **Centering the initial matrix**

The most commonly used centering is to subtract the average of the variables according to the equation:

$$X_{\text{corrected }\,,i,j} = X_{i,j} - \bar{x}_{.,i} \qquad \text{Eq.22}$$

With $X_{corrected,i,j}$ the absorbance transformed for the sample j at the wavelength i and $\bar{x}_{.,i}$ the average absorbance at the wavelength i.

In the case where the variables are heterogeneous (different units), the data is centered and reduced. Thus, all the variables have the same weight in the calculation of the principal components. Subsequently, Xt will represent the matrix of corrected absorbencies (m samples and w wavelengths).

- **Search principal components**

The first axis is the line for which the squares of deviations to the right are minimum. The criterion of least squares leads to maximizing the values of the orthogonal projections of the individuals on this line.

The projection coordinate $\mathbf{d_i}$ of a vector $\mathbf{x_i}$ on an axis is the scalar product of this vector with the unit vector $\mathbf{p_i}$ of this axis

$$d_i = x_i . p_i \qquad \text{Eq.23}$$

For all individuals the previous relation is written:

$$d = Xt . p_i \qquad \text{Eq.24}$$

With d is the vector of projections of m individuals on the axis 1.

For s the sum of the squares of the projections $\mathbf{s = d'.d = p_i'.Xt'.Xt.p_i}$

The first principal component is such that its unit vector $\mathbf{u_1}$ verifies:

- $\mathbf{p_1'.Xt'.Xt.\ p_1}$ est maximum
- $\mathbf{p_1}$ est un vecteur unitaire : $\mathbf{p_1'.\ p_1 = 1}$

Similarly, the second axis is such that its unit vector $\mathbf{p_2}$ satisfies the following three conditions:

- $\mathbf{p_2}'.\mathbf{Xt}'.\mathbf{Xt}.\mathbf{p_2}$ is maximum
- $\mathbf{p_2}$ is a unit vector: $\mathbf{p_2}'.\mathbf{p_2} = \mathbf{1}$
- $\mathbf{p_2}$ is orthogonal to $\mathbf{p_1}$: $\mathbf{p_2}'.\mathbf{p_1} = \mathbf{0}$

The following axes are defined in the same way

It is shown that the unit vectors that represents the solution of the problem are the eigenvectors of the initial covariance variance matrix **V** ($\mathbf{V} = \mathbf{X^{t}}'.\mathbf{X^{t}}$). The calculation is carried out by the diagonalization of the matrix **V**. The diagonalization of the matrix **V** gives two types of results: the matrix of the eigenvectors (called "loading") **P (w, a)** with a number of selected components and the matrix diagonal of the eigenvalues **L (a, a)**. To each eigenvector **p** is associated an eigenvalue **λ** which is the variance of the individuals on the corresponding axis.

- **Calculation of individuals coordinates**

The coordinates are calculated by projecting the individuals on the new selected axes:

$$T = X^t . p \qquad \text{Eq.25}$$

With T (m, a): matrix of factorial coordinates (also called scores).

It is possible to project in the same space, individuals who did not participate in the creation of the axes using the previous equation. These individuals are then called additional individuals.

2.2. Supervised methods

Exploratory techniques for data analysis, such as PCA, are unsupervised, meaning that they just show the data as they are. Conversely, supervised chemometric methods look for determined features within data, explicitly oriented to address particular issues. In particular, when a model is developed with the purpose of predicting a qualitative (PLS-DA) or quantitative (PLS) property of interest.

2.2.1. PLS-DA (Partial Least Square Discriminant Analysis)

Just like the PCA, Partial Least Squares Discriminant Analysis (PLS-DA) is a statistical method of data reduction, but it differs from the PCA in its supervised nature. PLSDA requires a predefinition to a membership group (class) of the different objects in the dataset [23–24].

PLS-DA is an extension of the PLS regression and can easily be adapted to the case of supervised classification especially when the number of explanatory variables is significantly greater than the number of individuals and / or when the variables are correlated.

It is necessary to reformulate the classification problem in the form of a regression equation. This is accomplished through the introduction of a dummy matrix Y, which codifies the belonging of the samples to a class. This matrix Y has as many columns as number of defined classes and as many rows as number of samples. For each observation in the training data set, its membership in a given class is binary encoded (1 for its class, 0 otherwise). For example, if 4 classes are defined and the i^{th} sample belongs to class 3 the i^{th} row (yi) of the matrix Y is:

$$y_i = [0 \quad 0 \quad 1 \quad 0] \qquad \text{Eq.26}$$

By using this binary coding for class membership, it is then possible to transform the classification problem into a regression problem through a function associating the X data matrix containing the variables measured on the samples and the dummy matrix Y.

$$Y = f(X) \qquad \text{Eq.27}$$

Assuming there is a linear relationship between these two matrices, the preceding equation becomes:

$$Y = XB \qquad \text{Eq.28}$$

B is a matrix of regression coefficients. The PLS approach can then be used to calculate the model even when the number of samples is less than the number of variables. The corresponding classification method is called discriminant PLS regression (PLS-DA).

The PLS regression assumes that the independent matrix **X** and the dependent matrix **Y** can be projected onto a reduced dimensional space and that a linear relationship exists between the scores of the two blocks. From a mathematical point of view, **X** and **Y**are decomposed into matrix of scores and loadings according to the following equations:

$$X = T.P^T + E_X \qquad \text{Eq.29}$$

$$Y = U.Q^T + E_Y \quad \text{Eq.30}$$

When T and U are respectively the matrices containing the scores of X and Y, P and Q are those containing the loadings of **X** and the loadings of **Y**, and E_X and E_Y containing the residuals.

In addition, the linear dependence between the **X** and **Y** scores previously assumed implies:

$$U = TC \quad \text{Eq.31}$$

C is the coefficients diagonal matrix. Based on the two previous equations (33 and 34), it is possible to compute a matrix containing the regression coefficients, denoted B, which makes it possible to predict the values of the dependent matrix Y_n for the unknown samples to which are associated variables measured X_n.

$$Y_n = X_n B \quad \text{Eq.32}$$

It is important to emphasize that unlike the dummy matrix Y which is coded in binary, the predicted values on the unknown samples Y_n will be real numbers. Thus, if 4 classes are involved in the classification problem as was the case in our previous example, when the model is tested on an unknown sample, the result will be in the form of a vector of real values with 4 dimensions that can take values such as [0,05 -0,16 0,88 0,09]. The classification of the sample is then carried out by assigning this sample to the category (class) corresponding to the highest value of the predicted values, in our example to the class 3 (0,88)(72).

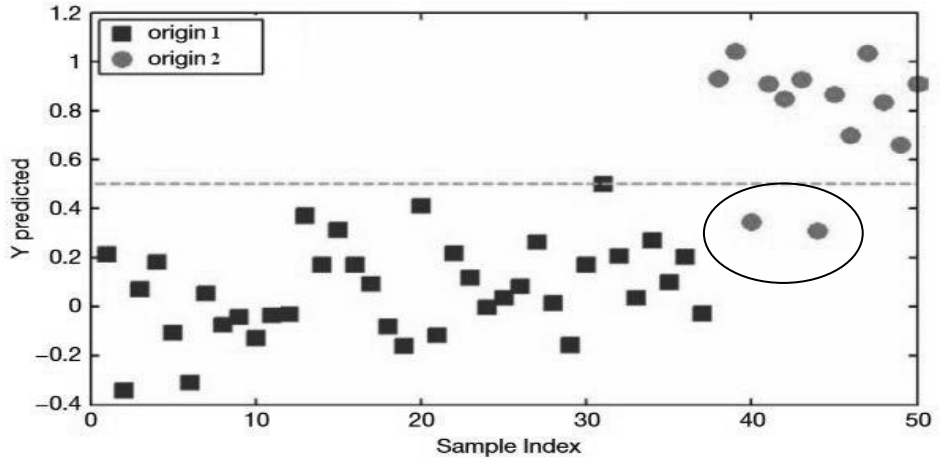


Figure 32: Illustration of how classification is accomplished in PLS-DA (estimated class values)

The figure 32 represents the estimated class values, samples are assigned to one or the other class based on the predicted y values and the green dashed line indicates the classification threshold (about 0.5). Accordingly, all samples except the two surrounded samples that fall below the threshold are correctly classified. This means that all individuals with a predicted value close to 0.8 were assigned to the origin 2 class, whereas individuals with predicted values of about 0 were attributed to the origin 1 class, and the threshold separates the two classes.

Since PLS-DA is a technique based on the number of components, when constructing the model, it is necessary to estimate the number of latent variables in order to obtain a reliable model and to avoid overestimation of the results. It is most often performed by cross-validation.

Once the model is calculated, it is essential to interpret the results to identify the initial variables that are most significant for the discrimination. Indeed, it is important to identify potential chemical markers and to ensure that this difference does not arise from data pre-processing errors or over-fitting (or over-adjustment) errors.

2.2.2. Partial least squares regression (PLS)

2.2.2.1. Introduction

PLS is a supervise method used in quantitative analysis, very widely used in infrared spectroscopy, Raman and fluorescence(73). It is considered as a standard tool to perform calibrations and predictions by modeling the relationship between two matrices, **X** and **Y**. The PLS is a recently developed generalization of the MLR multiple linear regression(74,75). It is a procedure that seeks to find the latent variables T (also called scores), which are linear combinations of both the predictor variables **X** and the variables that we are trying to predict **Y**(52).The construction of the relationship between predictive variables and responses is currently done using the NIPALS algorithms.

2.2.2.2. History of the PLS

The PLS approach was introduced for the first time by the Swedish statistician Herman Wold(76,77), in the framework of the modeling of structural relationships on latent variables for the analysis of several blocks of variables observed on the same individuals. Herman Wold's son, Svante Wold, with a number of Scandinavian scientists, including Harald Martens,

advocated its use in chemistry as part of two variable blocks(78,79). They have slightly modified the simple model of PLS to deal with the complex data to better respond to science and technology data where ordinary regression has been difficult or impossible to apply. Other PLS studies have been completed and reviewed by Tenenhaus(80).

2.2.2.3. Theory

The PLS method consists of a regression of the variable to be predicted, on latent variables which are linear combinations of the initial predictor variables. The latent variables are determined taking into account not only the predictor variables but also the responses(81).

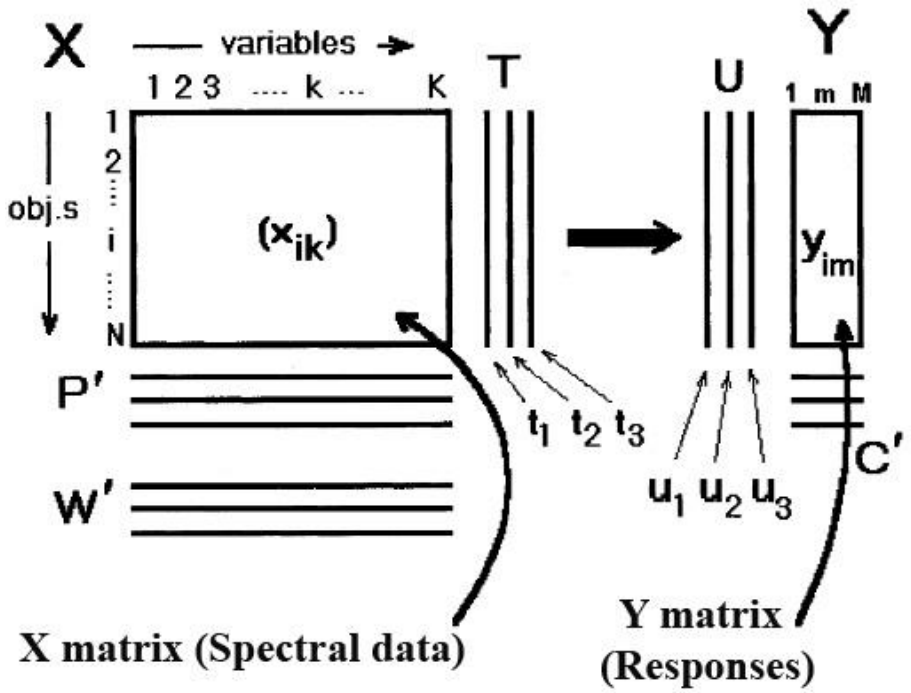


Figure 33: Data of a PLSR can be arranged as two tables, matrices, X and Y(79).

As displayed in figure 33, the linear PLSR model finds a few new variables, which are estimates of the LV's or their rotations. These new variables are called X-scores and denoted by $\mathbf{t_a}$ (a= 1,2,...,A). The X-scores are predictors of **Y** and also model **X** (Eqs. 34 and 36 below), i.e., both **Y** and **X** are assumed to be, at least partly, modeled by the same LV's.

The X-scores are "few" (A in number), and orthogonal. They are estimated as linear combinations of the original variables X with the coefficients, "weights", W^* (a= 1,2, . . . ,A). These weights are sometimes denoted by r_{ka}(82,83). Below, formulas are shown matrix form:

$$T = XW^* \quad \text{Eq.33}$$

(a) They are, multiplied by the loadings P', good "summaries" of X, so that the X-residuals, E, are small

$$X = TP' + E \quad \text{Eq.34}$$

With multivariate Y (when M>1), the corresponding "Y-scores" are, multiplied by the weight C, good "summaries" of Y, so that the residuals, G, are "small":

$$Y = UC' + G \quad \text{Eq.35}$$

(b) The X-scores are good predictors of Y, i.e.:

$$Y = TC' + F \quad \text{Eq.36}$$

The Y-residuals, F express the deviations between the observed and modeled responses, and comprise the elements of the Y-residual matrix, F.

Because of Eq. 33, the Eq. 36 can be rewritten to look as a multiple regression model:

$$Y = XW^*C' + F = XB + F \quad \text{Eq.37}$$

The "PLS-regression coefficients", B, can be written as:

$$B = W^*C' \quad \text{Eq.38}$$

2.2.2.4. Interpretation of the PLSR model

One way to see PLSR is that it forms "new x-variables" (LV estimates), t_a, as linear combinations of the old x's, and thereafter uses these new t's as predictors of Y. Hence, PLSR is based on a linear model. Only as many new t's are formed as are needed, as are predictively significant.

All parameters, t, u, w (and w*), p, and c, are determined by a PLSR algorithm. For the interpretation of the PLSR model, the scores, t and u, contain the information about the objects and their similarities/dissimilarities with respect to the given problem and model.

The weights W or the closely similar W*, and C, give information about how the variables combine to form the quantitative relation between X and Y, thus providing an interpretation of the scores, T and U. Hence, these weights are essential for the understanding of which X-variables are important (numerically large W-values), and which X-variables that provide the same information (similar profiles of W-values).

The PLS weights W express both the "positive" correlations between X and Y, and the "compensation correlations" needed to predict Y from X clear from the secondary variation in X. The latter is everything varying in X that is not primarily related to Y. This makes W difficult to interpret directly, especially in later components (a>1). By using an orthogonal expansion of the X-parameters in O-PLS, one can get the part of W that primarily relates to Y, thus making the PLS interpretation more clear

The part of the data that are not explained by the model, the residuals, are of diagnostic interest. Large Y-residuals indicate that the model is poor, and a normal probability plot of the residuals of a single Y-variable is useful for identifying outliers in the relationship between T and Y, analogously to MLR. In PLSR we also have residuals for X; the part not used in the modeling of Y. These X-residuals are useful for identifying outliers in the X-space, i.e., molecules with structures that do not fit the model, and process points deviating from "normal" process operations. This, together with control charts of the X-scores, T, is used in multivariate Statistical Process Control(84).

2.2.2.5. Geometric interpretation

PLSR is a projection method and thus has a simple geometric interpretation as a projection of the X-matrix (a swarm of N points in a K-dimensional space) down on an A-dimensional hyper-plane in such a way that the coordinates of the projection (ta, a=1,2, . . . ,A) are good predictors of Y. This is indicated in Fig. 34.

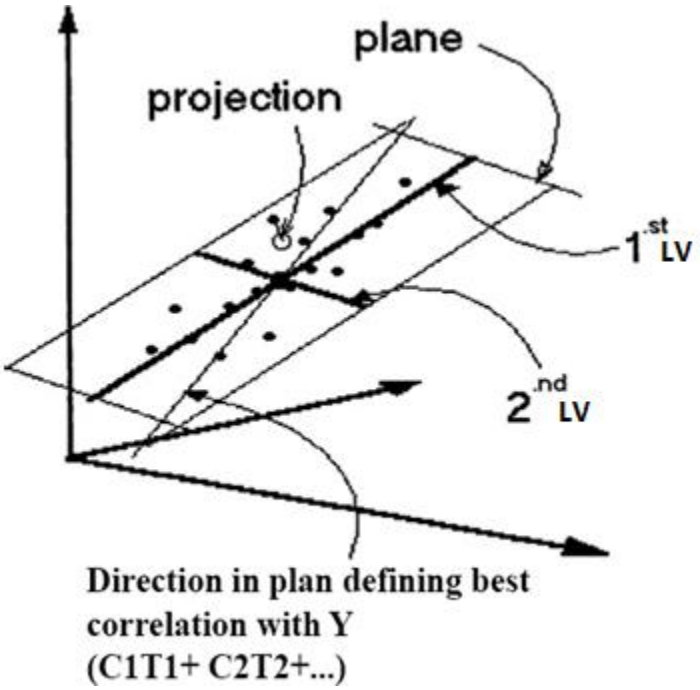


Figure 34: The geometric representation of PLSR(79)

The X-matrix can be represented as N points in the K dimensional space where each column of X (x_k) defines one coordinate axis. The PLSR model defines an A-dimensional hyper-plane, which in turn, is defined by one line, one direction, per component. The direction coefficients of these lines are p_{ak}. The coordinates of each object, i, when its data (row I in X) are projected down on this plane are t_{ia}. These positions are related to the values of Y(79).

2.2.2.6. The partial least squares regression of the 2nd order (PLS2)

The PLS2 algorithm(85), considered as the extension of the originally PLS algorithm developed by Wold et al.(75) and Martens and Næs(86), which is useful in the types of data set which is composed of two matrices, **X** and **Y**. the matrix **Y** corresponds to the responses to be predicted which contains more than one response (87), so in this case the prediction will be multiple, the model will search to correlate the input variables with a Y matrix that contains several columns.

2.2.2.7. The number of PLS components (LVs)

In any empirical modeling, it is essential to deter-mine the correct complexity of the model. With numerous and correlated X-variables there is a substantial risk for "over-fitting", i.e.,

getting a well fitting model with little or no predictive power. Hence, a strict test of the predictive significance of each PLS component is necessary, and then stopping when components start to be non-significant.

Cross-validation CV is a practical and reliable way to test this predictive significance(88). This has become the standard in PLSR analysis, and incorporated in one form or another in all available PLSR software. Good discussions of the subject were given by Wakeling and Morris (89)and Clark and Cramer(90).

Basically, CV is performed by dividing the data in a number of groups, G, say, five to nine, and then developing a number of parallel models from reduced data with one of the groups deleted.

After developing a model, differences between actual and predicted Y-values are calculated for the deleted data. The sum of squares of these differences is computed and collected from all the parallel models to form the Root Mean Squared Error of Cross Validation RMSECV, which estimates the predictive ability of the model(91).

2.3. Statistical criteria for assessing the quality of models

The robustness of a model is its ability to maintain a good predictive ability when exposed to environmental variations (such as temperature) or samples.

To estimate the performance of predictive models, we use several statistical criteria. The calibration error or the root mean square error of calibration (RMSEC). The cross-validation error or the root mean square error of cross-validation RMSECV. The prediction error or the root mean square error of prediction (RMSEP). The correlation coefficient R^2.

y and ŷ the vectors giving the reference values and the predicted values for N samples. A is the number of dimensions used to calculate the model, and C the centering value, 1 if centered model, 0 if non-centered model. The RMSEP calculation is:

$$RMSEP = \sqrt{\frac{(\hat{y}-y)'(\hat{y}-y)}{N}} = \sqrt{\frac{\sum_{i=0}^{N}(\hat{y}_i-y_i)^2}{N}} \qquad \text{Eq.39}$$

The same formula is often used to calculate the RMSEC and RMSECV. However, a more accurate calculation is obtained by calculating more precisely the degrees of freedom, ie: ddl = N-A-C. This gives:

$$RMSEC\ or\ RMSECV = \sqrt{\frac{(\hat{y}-y)'(\hat{y}-y)}{N-A-C}} = \sqrt{\frac{\sum_{i=0}^{N}(\hat{y}_i-y_i)^2}{N-A-C}} \qquad \text{Eq.40}$$

The Precise calculation is especially useful when the number of observations is small.

The correlation coefficient R^2 is the quantification of the difference between the real values and those obtained by the model.

3. A review of Chemometric tools applied to the field of petroleum products

3.1. Classification studies

Many studies have been conducted for classification and determination of the origin of petroleum products, using analytical techniques such as FTIR, GC-MS, and physico-chemical proprieties, coupled with some multivariate analysis such as Principal Component Analysis (PCA), Partial Least Square Discriminate Analysis (PLS-DA) and Soft Independent Modeling of Class Analogy (SIMCA)(13)(14).

Roman M. Balabin and Ravilya Z. Safieva have compared two Chemometric tools which are Regularized Discriminant Analysis (RDA) and Multi-Layer Perceptron (MLP) for the classification of motor oils by some physicochemical properties. Comparing the errors of the two techniques showed that the error of RDA was the double of MLP's one (57% vs. 26%) which shows that the latter is the best. The authors noted that the errors were very high and therefore proposed the use of another more sophisticated method like Support Vector Machines (SVM)(13).P.M.L. Sandercock and E. Du Pasquier have used chromatographic data (full-scan Gas Chromatography-Mass Spectrometry (92,93)(GC-MS) and GC-MS in Selected Ion Mode (SIM)) coupled to PCA and then Linear Discriminant Analysis (LDA) to establish the common origin of two or more liquid gasoline samples. The results showed that tested samples could be successfully classified into 32 different groups(94).Paulo J. S. Barbeira et al have shown that by

the use of Hierarchical Cluster (HC) and LDA applied on physicochemical properties, the origin of 1328 samples of gasoline marketed in the Brazilian market can be identified(95). Roman M. Balabin and Ravilya Z. Safieva Have compared three classification tools, namely, LDA, MLP and Soft Independent Modeling of Class Analogy (SIMCA), 382 samples of gasoline were tested to identify the most reliable method of classification. And as a result, they found that the MLP is the most effective method of classification(14).

3.2. Quantification studies

Several studies have focused on the prediction of some quality parameters of petroleum products using the coupling of chemometric tools and especially the PLS with spectroscopic techniques.

A. Borin and R. J. Poppi investigated the usefulness of iPLS (Interval Partial Least-Squares Regression) models for the quantification of some lubricants adulterant (gasoline, ethylene glycol and water) they obtained successful models characterized by a root mean square error of prediction (RMSEP) of 0.34% for gasoline, 0.037% for ethylene glycol and 0.023% for water(96). J. C. L. Alves et al compared the results of two types of modelization; the support vector regression (SVR) and the PLS coupled to near infrared spectra. The obtained values of RMSEP with the SVR models are 1.98°C and 0.453 for flash point and cetane number, respectively. The SVR provided significantly better results when compared with PLS (3.77°C and 0.556 for flash point and cetane number, respectively)(97). Jérémy Laxalde in his doctoral thesis dealt with the characterization of heavy oils using near-infrared spectroscopy, especially the prediction of saturates, aromatics, resins and asphaltens by PLS regression. The obtained RMSEP results in% (w / w), were 1.51 for saturates, 1.59 for aromatics, 0.77 for resins and 1.26 for asphaltenes. Which proves that this methodology is effective(98,99). Helga G. Aleme, Paulo J.S. Barbeira have used Partial least squares regression (PLS) on distillation curves data (ASTM-D86) for the prediction of flash point and cetane index of diesel, in this work they demonstrated that distillation curves together with PLS regression were effective to predict flash point in the 34.3–74.3 °C range and cetane index in the 41.2–50.8 range, regardless origin and type of diesel, with low RMSEP values 0.69 and 0.3 respectively for FP and CI(100).

4. Key findings

In this chapter we have recalled some essential notions of multivariate techniques, data preprocessing tools applied to the spectral data. We have developed the theory of the chemometric tools used in this thesis and we defined the statistical criteria for assessing the quality of models. This chapter has been completed by a bibliographical study on the history of the use of chemometric tools in the oil field.

EXPERIMENTAL PART:

THE APPLICATION OF FTIR SPECTROSCOPY AND CHEMOMETRIC TOOLS IN THE QUALITY CONTROL OF PETROLEUM PRODUCTS MARKETED IN MOROCCO

I. Introduction

Petroleum derivatives obtained from the crude oil refining are complex mixtures of many chemical compounds characterized by specific boiling points. They are basically composed of saturated and aromatic hydrocarbon; these contents depend on to two main criteria, the crude oil geographical origin and the quality of the refining unit. These products are the main source of energy and especially in the field of transport.

Engines are designed and manufactured to work on specified fuel. When the fuel characteristics are changed, it can deteriorate the engine and increase the emission of more pollutants. In this way the use of low quality, non-compliant and adulterated fuels usually leads to engine's life and performance reduction and increase the emission of harmful pollutants (101,102).

The Fuel Quality Control Law requires refineries and importers and even at the retail level to test the quality of the fuel prior to its marketing. These requirements are intended to ensure that these products meets all the mandatory specifications at retail stations by installing internal control laboratories that must be equipped with specific instruments based on physic-chemical analysis for the determination of different quality parameters piloted by qualified personnel (table 8).

Quality control tests must be carried out based on standardized test procedures or internally developed and validated methods. These methods require equipment characterized by their sophistication, their expensiveness and demands a specialized staff; this has led specialists to look for simpler, reliable and cheaper methods to evaluate the fuel quality inside the refineries.

Table 8: example of diesel property information(103)

Property	ASTM D	Minimum	Maximum	Average
API gravity	1298	31.0	48.0	39.5
Density	1298	0.7880	0.8719	0.82995
API gravity	4052	30.7	48.2	39.45
Kinematic viscosity. cSt. 40 °C	445	1.14	4.05	2.595
Cloud point. °C	2500	260.5	2.1	131.3
Freeze point. °C	2386	-59.4	6.6	-26.4
Pour point. °C	97	-78	-3	-40.5
Flash point. °C	93	23	96	59.5
Initial boiling point. °C	86	99	229	164
Boiling point at 10% recovery. °C	86	158	256	207
Boiling point at 50% recovery. °C	86	182	297	239.5
Boiling point at 90% recovery. °C	86	223	347	285
Boiling point at 95% recovery. °C	86	230	375	302.5
Final boiling point. °C	86	240	388	314
Cetane number	613	36.9	61.3	49.1
Calculated cetane index	976	37.0	59.8	48.4
Calculated cetane index	4737	37.4	14.33	25.865
Carbon. wt %	5291	84.74	87.61	86.175
Hydrogen. wt %	5291	12.29	14.33	13.31
Carbon/hydrogen	5291	5.95	7.09	6.52
Net heat of combustion. MJ/Kg	240	42.25	43.46	42.855
Net heat of combustion. BTU/lb	240	18 116	18 684	18400
Monocyclic aromatics. wt %	5186	7.7	38.9	23.3
Dicyclic aromatics. wt %	5186	0.4	12.8	6.6
Polycyclic aromatics. wt %	5186	0.0	3.4	1.7
Total aromatics. wt %	5186	8.3	47.2	27.75

To meet these requirements, the following objectives of this thesis are discussed in this experimental part:

- FTIR fingerprints associated to a PLS-DA model for rapid detection of smuggled non-compliant diesel marketed in Morocco.
- Discrimination and classification of diesel fuels using FTIR and GC-MS analysis and Chemometrics methods.
- Discrimination and quantification of Moroccan gasoline adulteration with diesel using Fourier Transform Infrared Spectroscopy and Chemometric tools.
- The association of the FTIR spectroscopy with PLS2 and PLS1 algorithms for the prediction of ten quality parameters of diesel fuels

II. Materials and methods

1. GC-MS analysis

A Thermo Electronics gas chromatograph coupled with mass spectrometer ISQ 7000 Single Quadrupole (Thermo Fisher Scientific wissenschaftlicheGeräte GmbH, Inc., Austria) was used to collect the GC-MS data. An apolar capillary column TR-5 (60m x 0.32mm x 0.25 μm) was used. The constant velocity mode (1,5ml/min) with Helium carrier gas was selected. The inlet temperature was set to 270°C with an injection volume of 0.1μl. The GC oven profile was set to start at 50°C (held for 4 min), then increased to 110°C at a rate of 8°C/min (held for 0.1 min) and increased again to 270°C at a rate of 15°C/min (held for 1min). The mass spectrometer source was operated at a temperature of 300°C in electron ionization mode (70 eV) with a scan range of m/z 35–400(104,105).

2. FTIR spectra acquisition

A Perkin Elmer Spectrum One FTIR spectrometer equipped with a Zinc Selenide (ZnSE) ATR system was used to record all the spectra between 4000 and 400 cm^{-1}. Every spectrum was the mean of 24 scans with a resolution of 4 cm^{-1}, the number of resulting variables was 2559.

3. Sample collection

3.1. Rapid detection of smuggled non compliant Diesel products

Nineteen (19) authentic diesel samples were purchased from the major brands of petroleum products in the Moroccan capital Rabat (Total and Afriquia). Eleven (11) other smuggled diesel samples were collected directly from smugglers in the eastern zone (Oujda city). The samples were stored in the fridge at 4-9° C to avoid the risk of component degradation.

3.2. Classification of diesel samples by suppliers

80 diesel samples were purchased in different regions of Morocco from different stations owned by the four most important brands (Afriquia, Shell, Total and Winxo). They were conditioned in 30ml dark glass bottles and stored in a fridge at 4-9° C to avoid their degradation.

3.3. Discrimination and quantification of gasoline adulteration with diesel

Gasoline and diesel samples were purchased from Total Gas Company in the city of Rabat. A group of 100 mixtures (1 mL; from 0 to 98%) were prepared by adding diesel to gasoline in small vials. Vials were then stored in the fridge to avoid evaporation of the components and especially gasoline components.

3.4. Prediction of diesel quality parameters

Fifty diesel samples were collected from the Saybolt Lab Morocco owned by Total Morocco. Corresponding to different batches of diesel collected during the control of imported petroleum products. The samples conditioning was done in opaque glass vials that were stored in a refrigerator at 4-9 ° C to prevent the degradation of petrol components.

4. Software and data processing

The data were processed using the PLS-Toolbox software version 8.2.1 (Eigenvector Research, Wenatchee, WA, USA) on Matlab version 15.a (The Math-Works, MA, USA) and the free and open source software R-package.

5. The quality control test

The interest properties of the diesel were determined by the Saybolt Total Laboratory owned by the Total Company in the city of Mohammedia according to the ASTM and ISO guidelines as shown in the table 9.

Table 9: Testing methods for diesel quality parameters determination

Parameter	Testing method
Color	ASTM D 1500
Density	ASTM D 1298
Boiling point at40% recovery	ASTM D 86
FlashPoint	ASTM D 93
Pour Point	ASTM D 97
Cold Filter Plugging Point	ASTM IP 309
Viscosity at 40°C.	ASTM D 445
Cetane Number	ASTM D 976
Conductivity at 20°C.	ASTM D 2624
Water contain	NF EN ISO 12937

III. Results and discussion

1. FTIR fingerprints associated to a PLS-DA model for rapid detection of smuggled non-compliant diesel marketed in Morocco

1.1. The aim

The aim of this work was the development of a rapid and non-destructive method in order to detect and classify smuggled diesel from the authentic fuels based on the FTIR spectra coupled to PCA and PLS-DA methods.

1.2. Methodology

The multivariate model was set up using the total of the acquired spectra (X matrix), whereas two categorical classes (authentic and smuggled) were considered as responses matrix (y matrix). The samples set was divided into training set and test set. The first one corresponds to a random selection of 16 authentic and 9 smuggled samples, while three samples from each group were considered as a test set.

1.3. GC-MS characterization of diesel samples

The GC-MS analyses of diesel samples were performed as a qualitative control by identifying their chemical composition in order to verify and avoid the samples adulteration. GC-MS chromatograms (Figure 35) and Table 10 showed that all samples were essentially composed of a series of hydrocarbons (C_8H_{18} to $C_{24}H_{50}$) in addition to some volatile products such as benzene, toluene, benzene-ethyl, dimethyl and trimethyl. This composition found is in concordance with the results obtained by Sink et al. (106,107). GC-MS results obtained certify that the diesel samples were pure and do not contain foreign products.

Table 10: List of main compounds of diesel with their retention time and monitored ions

N°	Compound	Retention Time, min	Monitored ions, m/z
1	Toluene	1.508	51, 65, 91
2	Ethylbenzene (EB)	2.691	77, 91, 105, 106
3	Dimethylbenzene (DMB)	3.086	77, 91, 105, 106
4	Trimethylbenzene (TMB)	5.662	77, 91, 105, 120
5	C_8H_{18}	1.719	43, 57, 71, 85, 99, 114
6	C_9H_{20}	3.204	43, 57, 71, 85, 99, 128
7	$C_{10}H_{22}$	5.916	43, 57, 71, 85, 99, 142
8	$C_{11}H_{24}$	8.309	43, 57, 71, 85, 99, 156
9	$C_{12}H_{26}$	10.394	43, 57, 71, 85, 99, 170
10	$C_{13}H_{28}$	12.161	43, 57, 71, 85, 99, 184
11	$C_{14}H_{30}$	13.457	43, 57, 71, 85, 99, 198
12	$C_{15}H_{32}$	14.505	43, 57, 71, 85, 99, 212
13	$C_{16}H_{34}$	15.364	43, 57, 71, 85, 99, 226
14	$C_{17}H_{36}$	16.164	43, 57, 71, 85, 99, 240
15	$C_{18}H_{38}$	16.909	43, 57, 71, 85, 99, 254
16	$C_{19}H_{40}$	17.606	43, 57, 71, 85, 99, 268
17	$C_{20}H_{42}$	18.276	43, 57, 71, 85, 99, 282
18	$C_{21}H_{44}$	18.881	43, 57, 71, 85, 99, 296
19	$C_{22}H_{46}$	19.475	43, 57, 71, 85, 99, 310
20	$C_{23}H_{48}$	20.075	43, 57, 71, 85, 99, 324
21	$C_{24}H_{50}$	20.631	43, 57, 71, 85, 99, 338

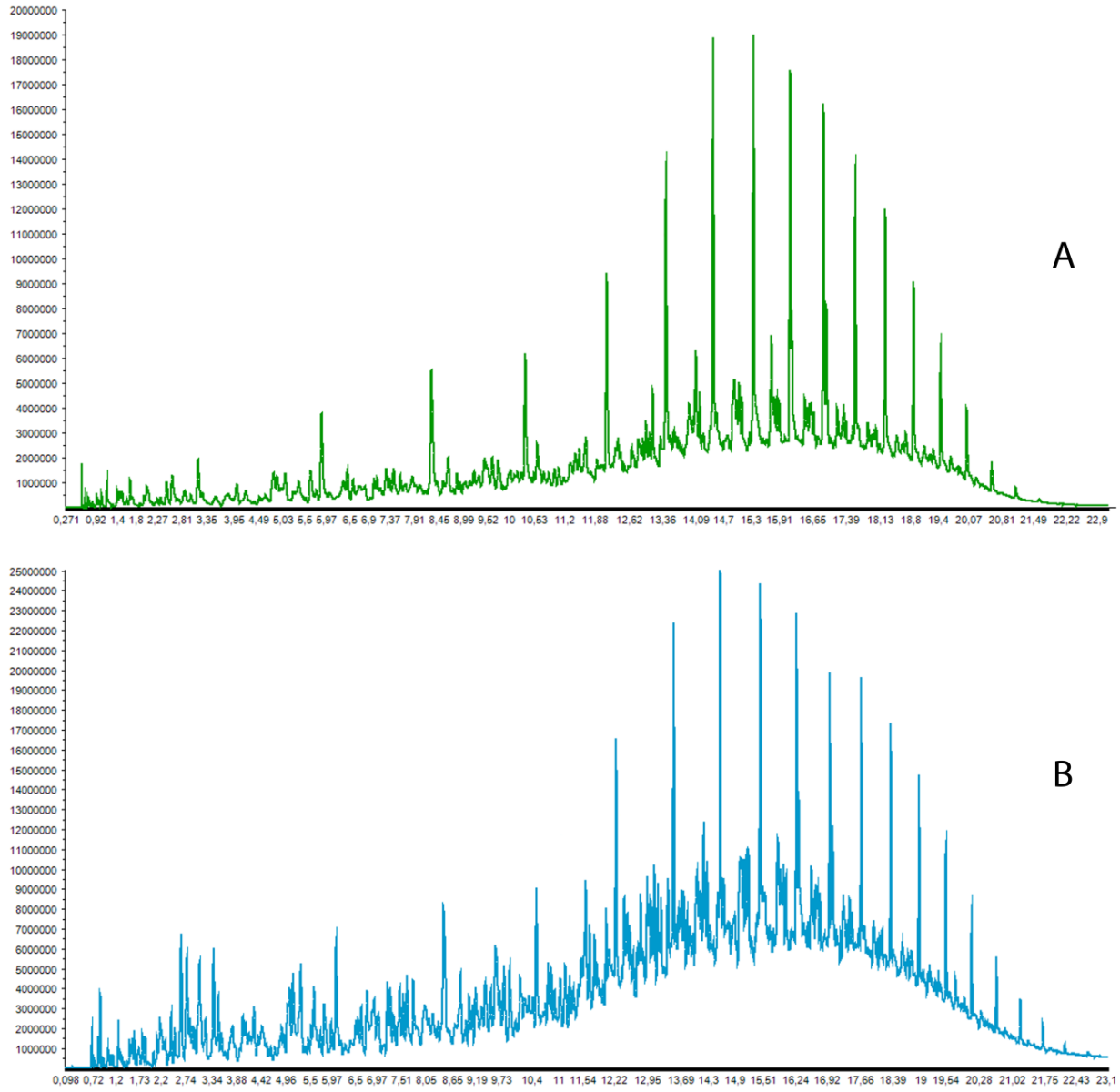


Figure 35: Examples of chromatograms obtained by GC-MS of authentic diesel (A) and smuggled diesel (B)

1.4. FTIR spectra

FTIR spectra of diesel samples are shown in Figure 36. The most important band at 2900 cm^{-1} was attributed to the symmetric CH_3 stretch. The second at 2830 cm^{-1} was generated by the symmetric CH_2 stretch. The band at 1445 cm^{-1} can be associated to the CH_3 asymmetric deformation and the absorbance at 1400 cm^{-1} assigned to the CH_2 scissor bending vibration. The group of bands between 900–700 cm^{-1} was due to $C_1{-}C_2$ stretching, CH_3rocking and C−O stretching(108).

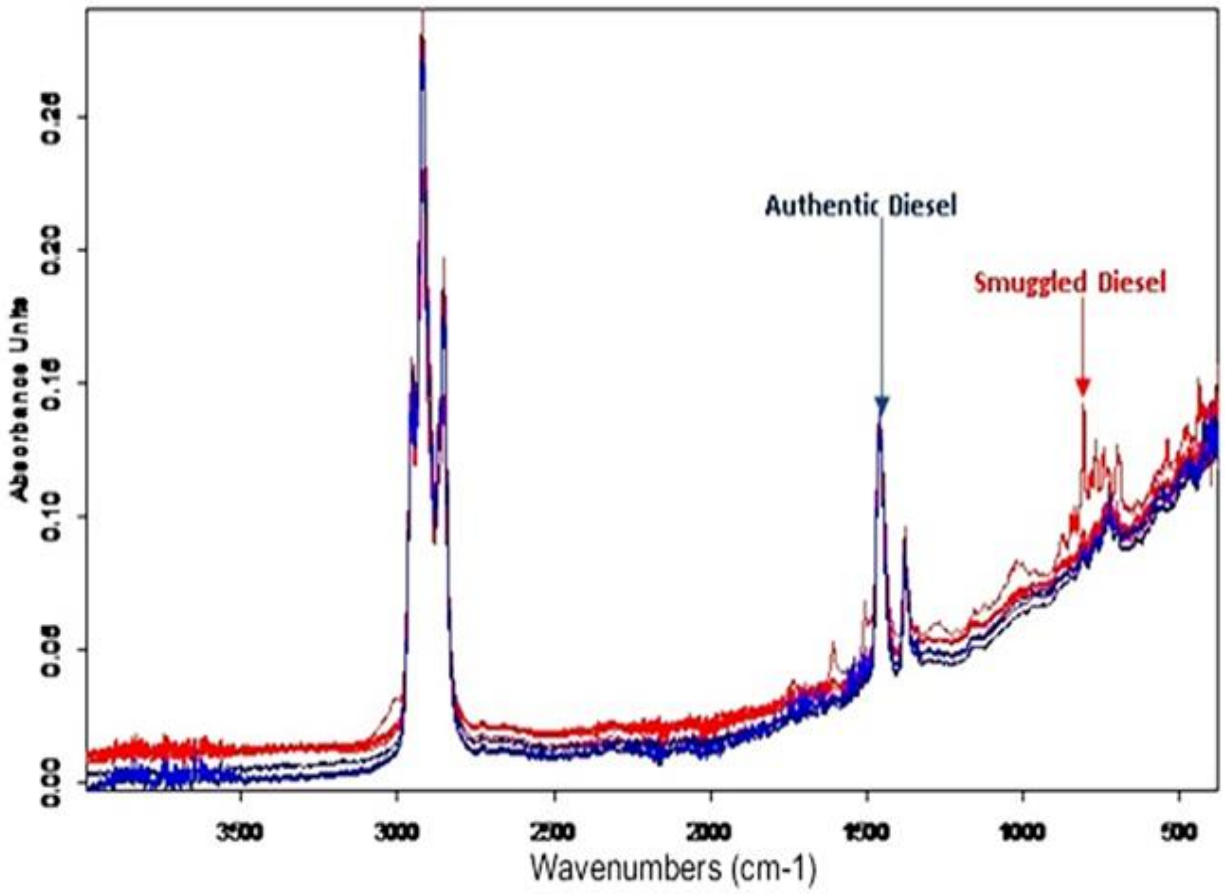


Figure 36: FTIR spectra of authentic diesel samples and smuggled samples

1.5. Data exploration by principal component analysis (PCA)

Principal Component Analysis was applied to the full spectral data, the entirety of spectra range (4000–600 cm^{-1}) was investigated using cross-validation (Venetian-blinds).

The first two PCs explained more than 68% of the total variance (51.09% for PC1 and 16.3% for PC2). PC1 vs. PC2 scores plot (Figure 37) indicated possible discrimination between the two classes of samples. The green class refers to the authentic diesel samples, whereas the red ones represented smuggled products. In addition, no outlier has been highlighted.

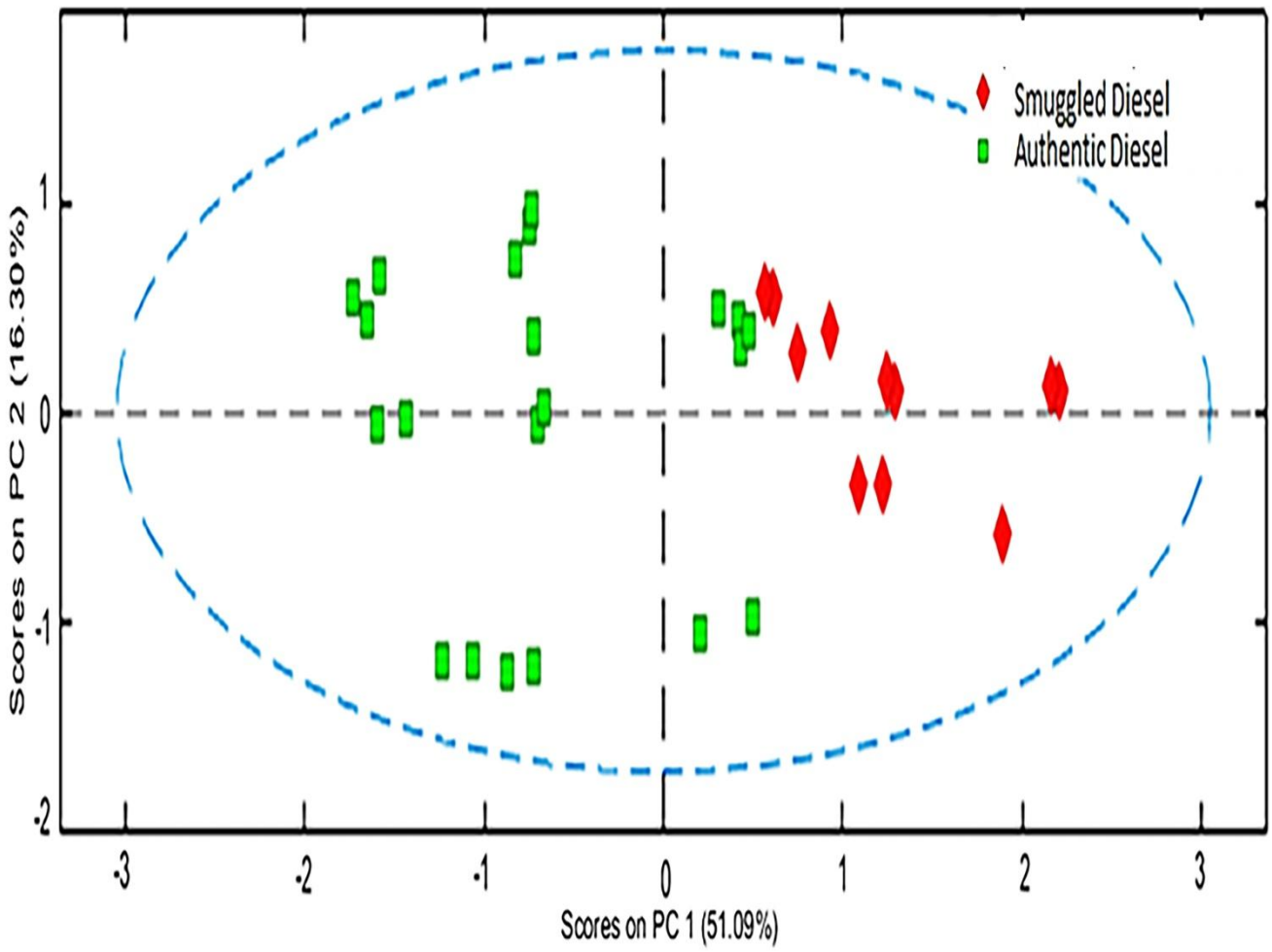


Figure 37: Scores plot of PCA model with two principal components showing smuggled diesel samples in red and authentic samples in green

The loadings plot (Figure 38) showed the main bands which allowed the discrimination between the two groups. These variables were situated between 3000 and 2800 cm^{-1} characteristic of CH_3 and CH_2, in addition to some absorbance located between 1700 and 1400 cm^{-1}.

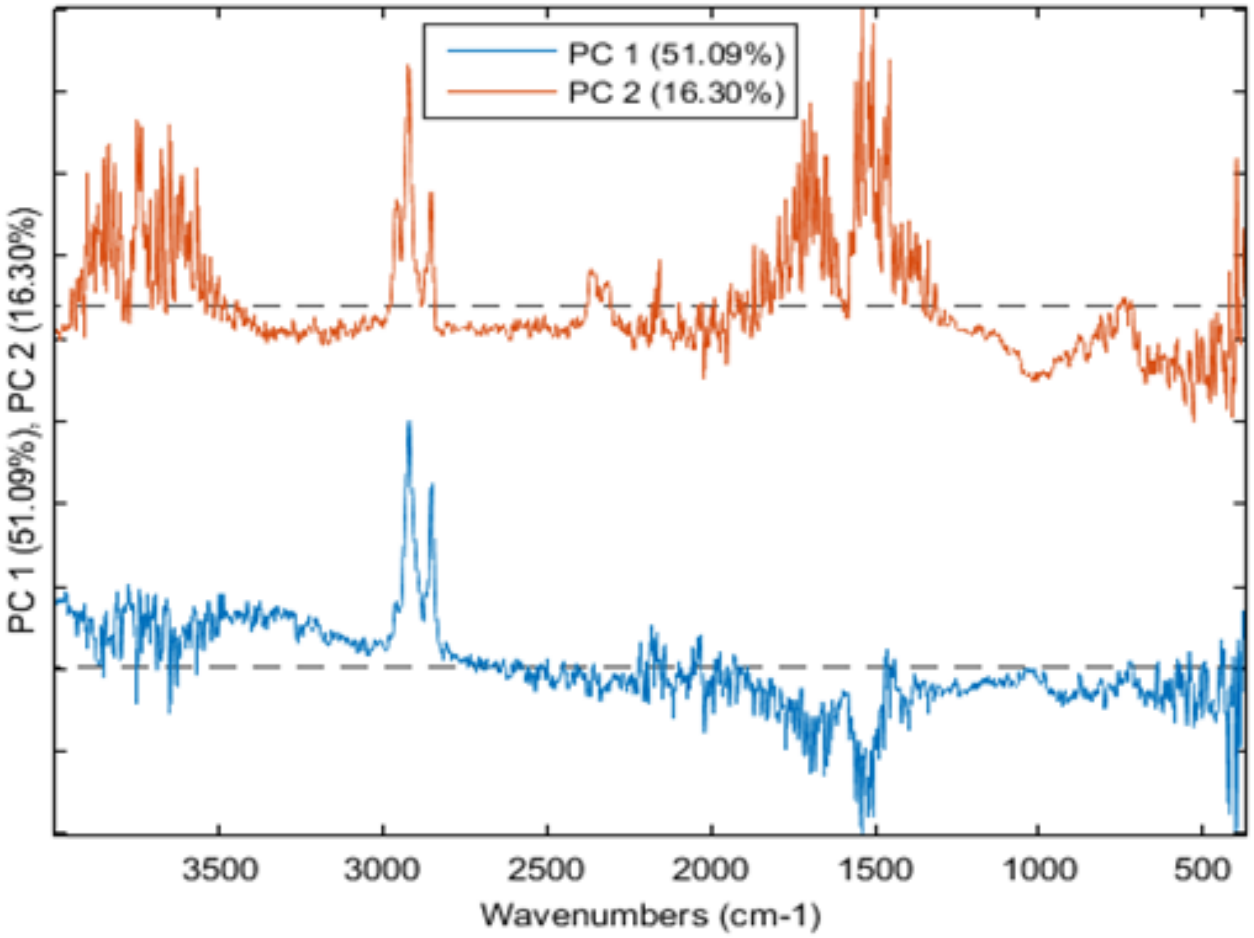


Figure 38: Plots of loadings of the PCA model

1.6. Classification by Partial Least Square Discriminate Analysis

The PLS-DA model allowed a clear differentiation between the two classes as indicated in the PLS-DA scores plot (Figure 39). Both groups were easily discriminated with the two first latent variables (LV1 explained 52.32% of the total variance and 9.83% by the second LV). The PLS-DA loadings plot (Figure40) indicated the variables responsible for the classification. The absorbance bands were especially located between 3000 and 2800 cm^{-1}, in accordance with the trends of the PCA results.

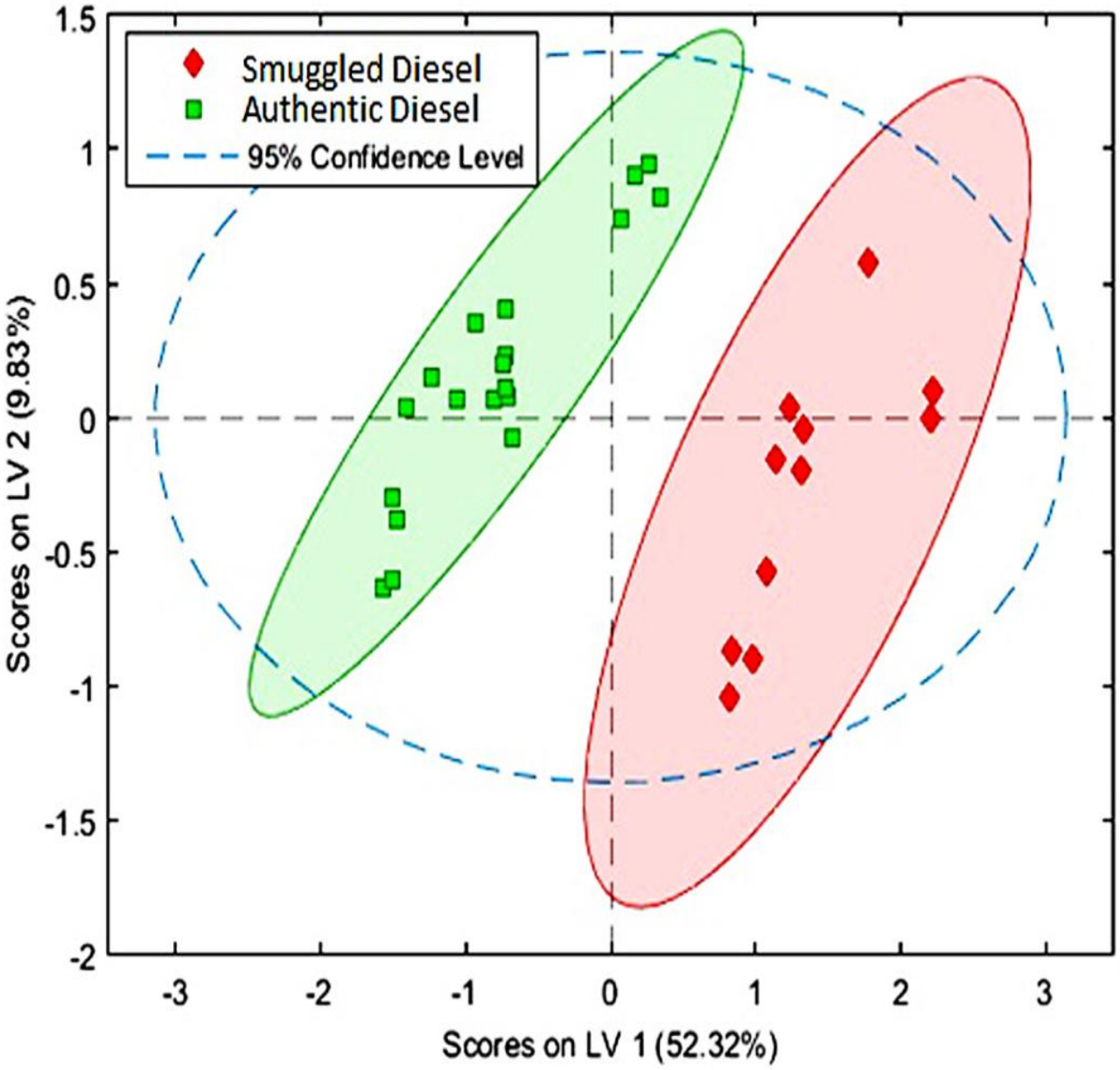


Figure 39: PLS-DA Scores plot with two latent variables (the first and the second) presenting two separate groups of samples. In red smuggled products and authentic samples in green

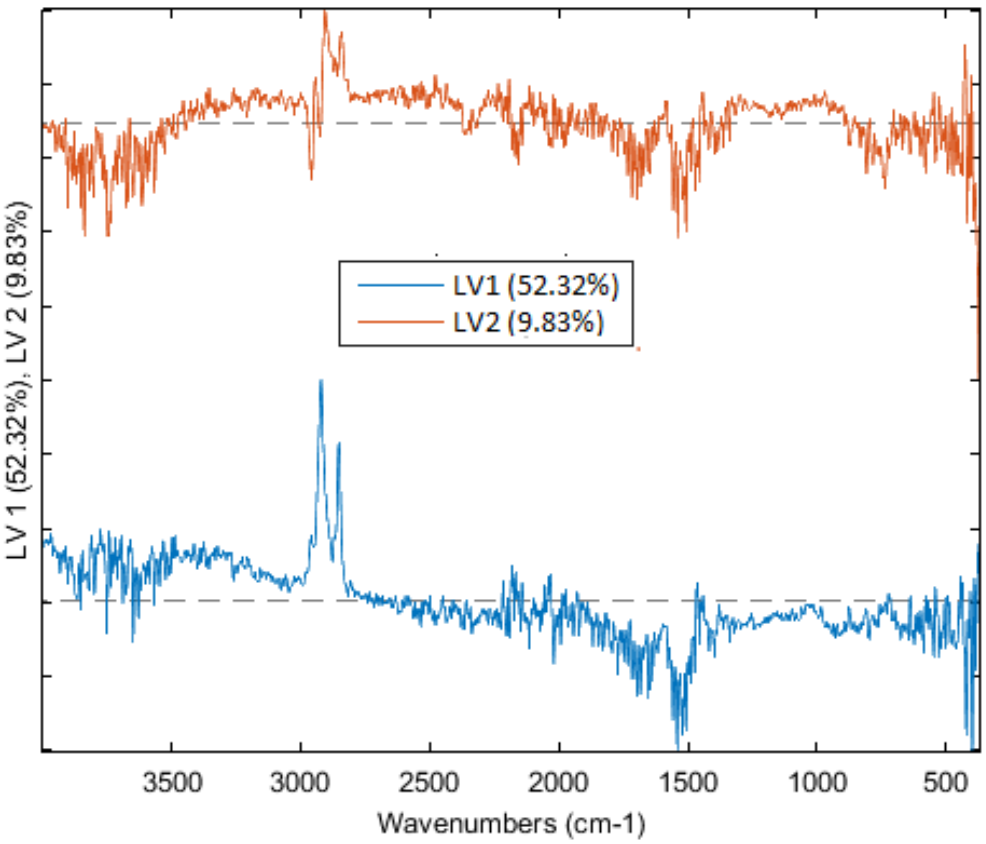


Figure 40: Plots of loadings of the PLS-DA model

Otherwise, Figure 41 showed a clear discrimination between the estimated values of the two test sets without ambiguity. This was confirmed by the clear separation between the samples belonging to each class by the threshold, as well as the validation individuals were successfully attributed to their original classes. These results prove that FTIR spectra contain a lot of useful chemical information that allows the PLS-DA model to classify the two products.

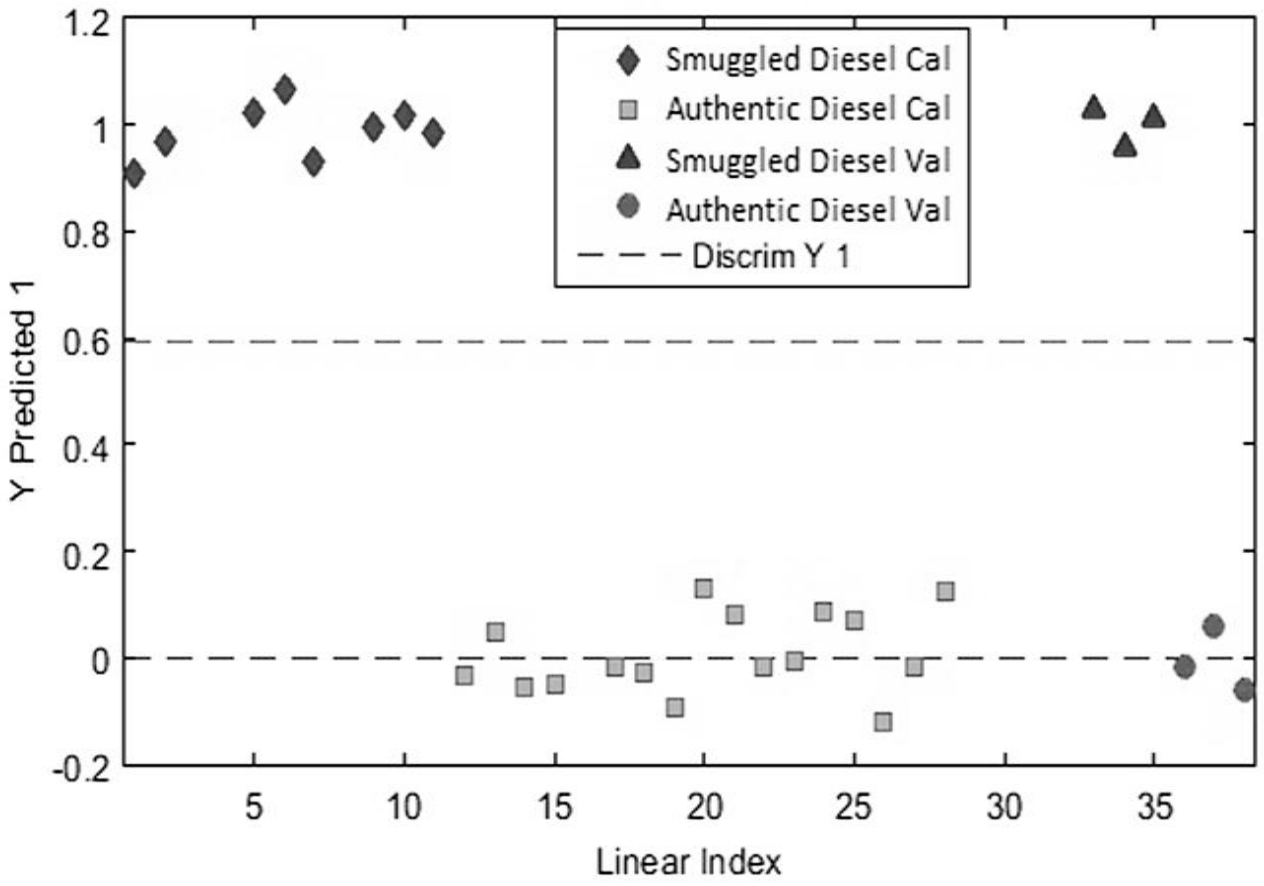


Figure 41: Estimated class values for calibration and validation sets obtained by the PLS-DA model

Table 11: Figures of merit of PLS-DA model

Class of diesel	Latent Variables	RMSEC	RMSEP	Specificity	sensitivity
Smuggled Diesel	3	0.1013	0.6924	1.00	1.00
Authentic Diesel	3	0.1013	0.6924	1.00	1.00

The Figures of merit summarized in the table 11, indicate that the classification model established in this work was characterized by a high sensitivity and specificity (100%), and acceptable values of errors (RMSEC and RMSEP).

The authenticity certification of a diesel sample requires performing at least six physic-chemical analysis including distillation fractions, density, flash point, cetane number, sulfur content and the determination of the chemical composition by GC-MS (16). Nevertheless, this study proves that a FTIR fingerprinting coupled with chemometric tools as a simple, rapid and non-destructive technique offers an alternative to the complex methods for certifying the authenticity of diesel.

1.7. Conclusion

The usefulness of coupling FTIR fingerprints to PCA and PLS-DA for the discrimination and classification of two diesel fuel products (authentic and smuggled) was demonstrated. In addition to its simplicity and rapidity, the FTIR spectra of diesel displayed very important information. PCA allowed a good discrimination between the two groups of samples, whereas the PLS-DA model was able to classify and predict the two classes with a high classification rate, sensitivity and specificity. This work proves that a simple FTIR fingerprint with chemometric tools can solve the authenticity problem of diesel fuels.

The results obtained in this study are quite promising, encouraging similar investigations in diesel fuel quality control. The current results require further validation and high number of samples in order to trace reliable models for rapid detection of smuggled non-compliant products.

2. Discrimination and classification of diesel fuels using FTIR and GC-MS analysis and Chemometrics methods

2.1. The aim

This work was aimedto compare the performances of GC-MS and FTIR techniques coupled with the PLS-DA model to discriminate diesel samples from four predefined suppliers. The proposed approach will solve the problem of diesel origin identification by resorting predictive PLS-DA models on the recorded FTIR and GC fingerprints.

2.2. Methodology

The multivariate models were set up using all the acquired data (X matrix) which pertain to the GC-MS analysis, on the one hand, and the FTIR spectra, on the other hand, measured on the fuel samples. For PLS-DA, the group variable, Y, corresponds to the origin of the fuel with four categories: Afriquia, Total, Shell and Winxo. The dataset was divided into a training set and a test set. The first one corresponds to a random selection of 64 samples with a balanced distribution in the four groups (i.e., 16 samples from each group). The test set was formed of 16 samples (4 samples from each group).

2.3. Data preprocessing

Regarding the FTIR data, three signal preprocessing were applied: the sample normalization, general least square weighting with α= 0.0002 and the mean centering(109).

The chromatograms were submitted to two preprocessing; The first strategy is the Correlation Optimized Warping (COW) which eliminates the shift-related artifacts from measurements by aligning the data from the various samples to best correlate to a reference(109,110). The second strategy is the baseline correction. This consists in a mean centering of each column. Thus, the data associated with each sample (*i.e.*, row) reflect how that sample differs from the average sample in the original data matrix(109).

2.4. FTIR spectra

Figure 42 shows the FTIR spectra of diesel samples. The highest absorbance at 2900 cm^{-1} was generated by the symmetric CH_3 stretch. The second at 2830 cm^{-1} was attributed to the

symmetric CH_2 stretch. Then some weak absorbance at 1445 cm^{-1} and 1400 cm^{-1} respectively associated to the CH_3 asymmetric deformation and the CH_2 scissor bending vibration(108).

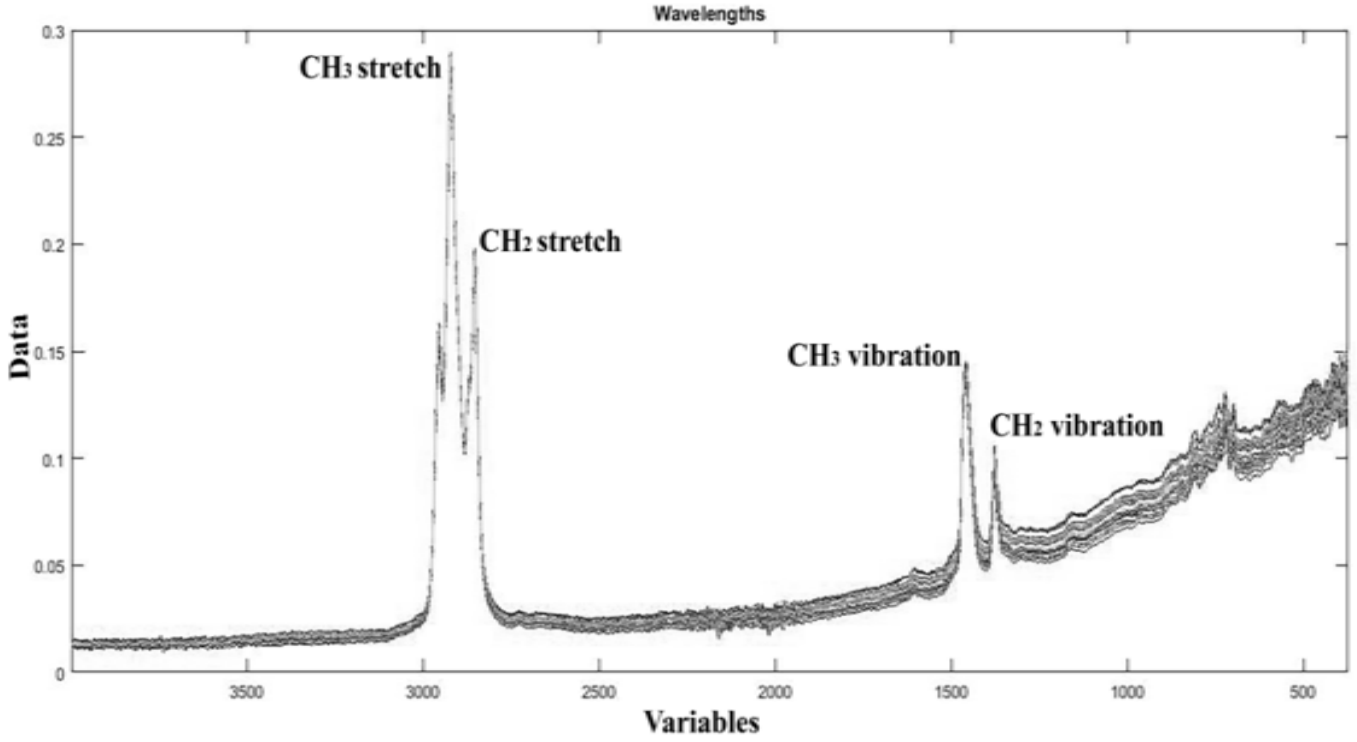


Figure 42: FTIR spectra of diesel samples

2.5. GC-MS characterization of diesel samples

The determination of the chemical composition of diesel samples was performed by the GC-MS technique. Prior to the classification step, the GC data can be used to verify and avoid the samples adulteration(111). GC-MS chromatograms Figure 43, revealed that diesel samples were mainly composed of a series of hydrocarbons (C_8H_{18} to $C_{25}H_{52}$) and low contents of some volatile products such as benzene, toluene, benzene-ethyl, dimethyl and trimethyl.

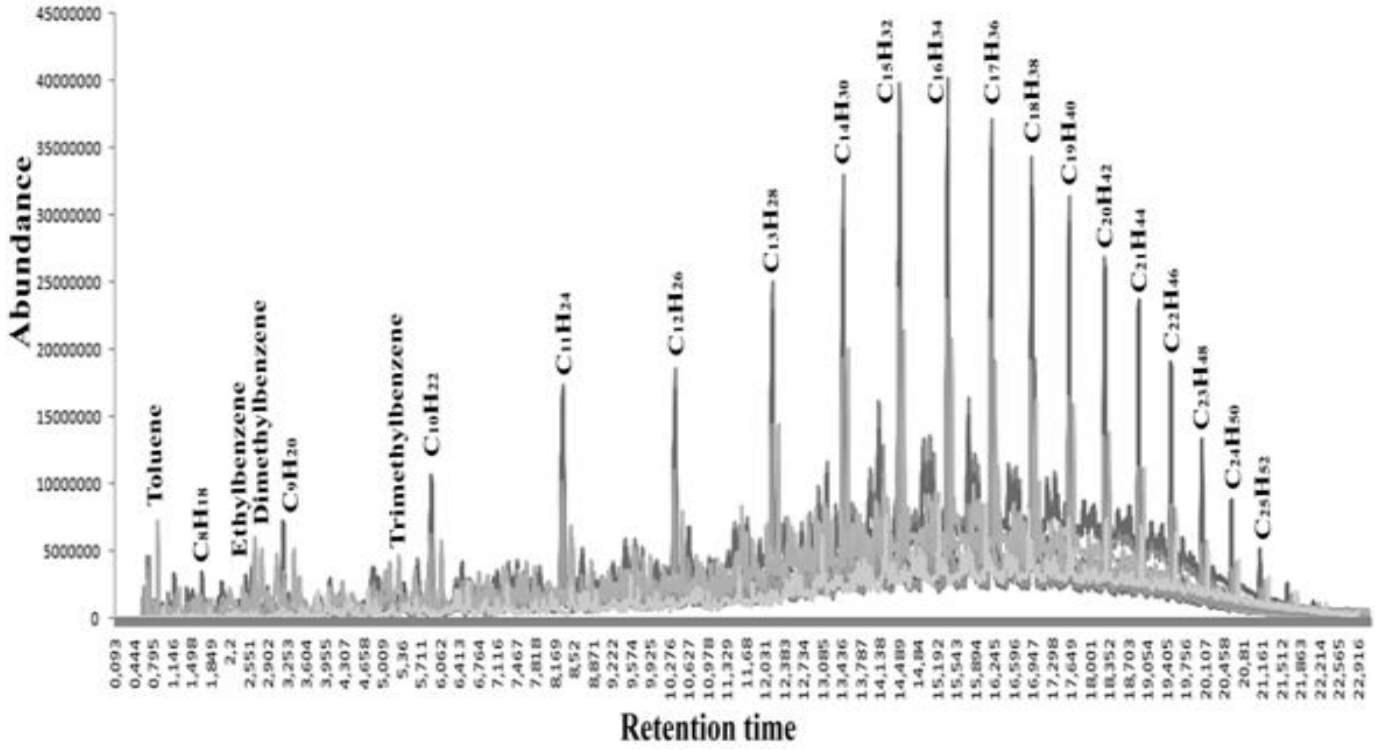


Figure 43: Chromatograms obtained by the GC-MS technique for Diesel fuels

2.6. Classification of diesel fuels by suppliers based on FTIR spectra

2.6.1. PCA Analysis

Once preprocessed, the FTIR spectra were submitted to PCA. Figure 44 shows the scores plot depicting the configuration of the diesel samples on the basis of the first two principal components (PCs), which explain around 50% of the total variance. Although PCA was not specifically aimed at discriminating the predefined groups, we can observe that the four groups associated with the four brands (Afriquia, Shell, Winxo and Total) are clearly identified.

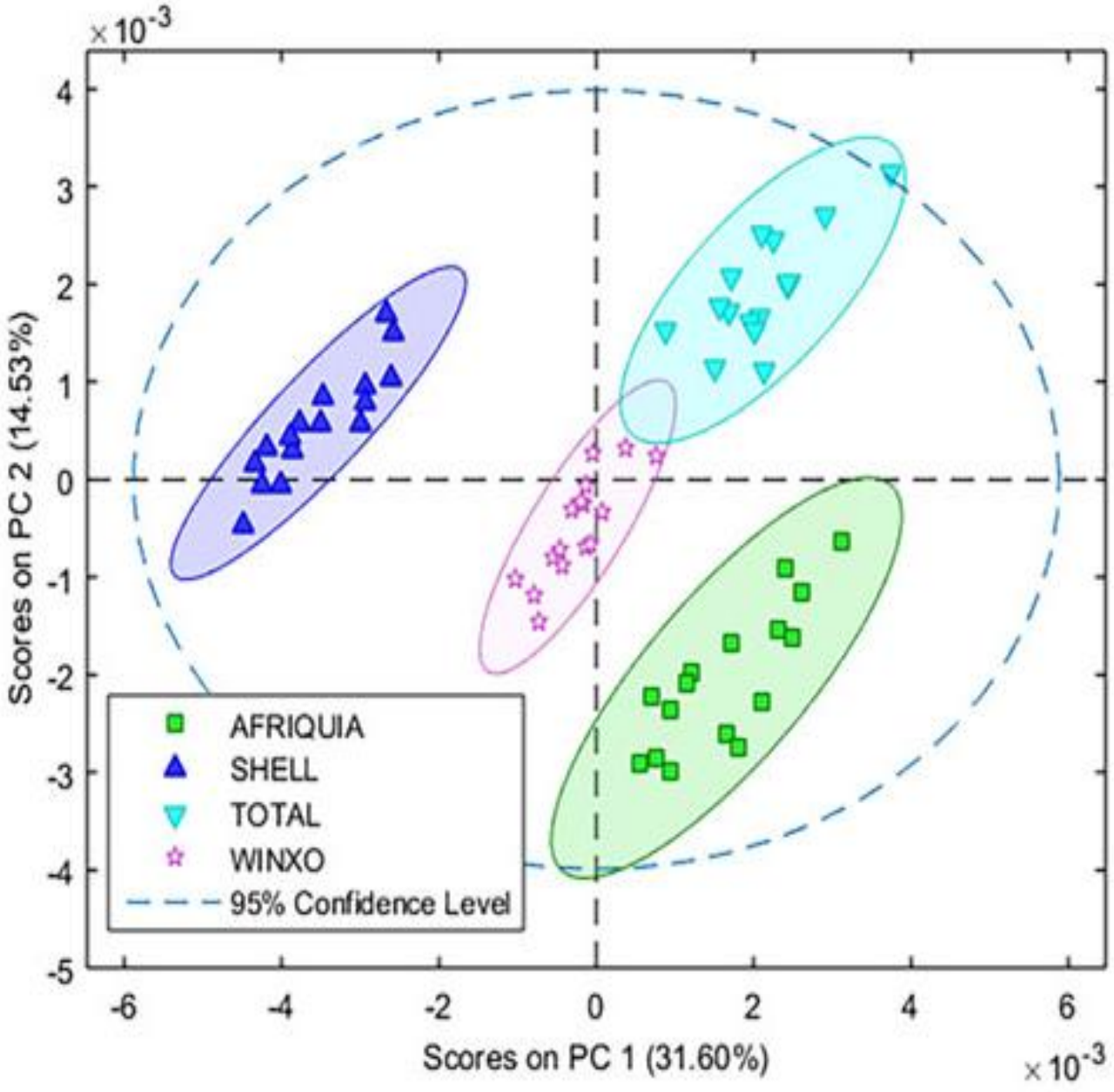


Figure 44: PCA Scores plot, PC1vs PC2, on the FTIR spectra

The loading plot (Figure 45) represents the weights of the variables (absorbencies) on the different PCs. It shows that the main absorbencies that define the first two PCs are located between 3050 and 2700 cm^{-1} corresponding to CH_2 and CH_3 stretch.

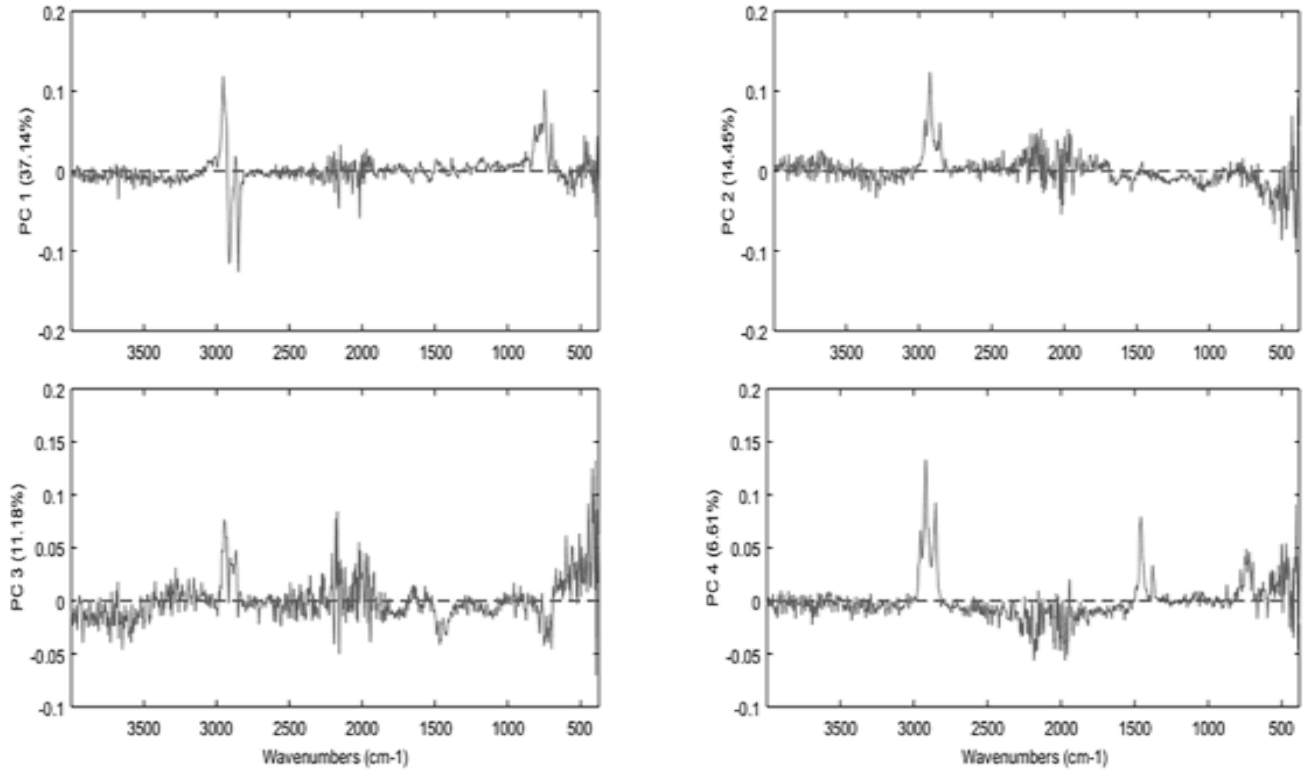


Figure 45: PC1 to PC4 loading plots of the PCA model on FTIR spectra

2.6.2. PLS-DA discrimination

In order to set up the PLS-DA model, the number of significant latent variables should be fixed. Using a cross validation procedure (112), three latent variables were selected.

Table 12 gives indicators that highlight the performance of the model. The values and differences between RMSEC and RMSEP are clearly small for all the fuel groups. Moreover, the good values of sensitivity and selectivity about 100%.

Table 12: Discrimination parameters of the FTIR PLS-DA models

Technique	Param\Fuel	Diesel Afriquia	Diesel Shell	Diesel Total	Diesel Winxo
FTIR	**LVS**	3	3	3	3
	RMSEC	0.11	0.06	0.10	0.09
	RMSEP	0.13	0.09	0.12	0.15
	Sensitivity (calibration)	1.00	1.00	1.00	1.00
	Specificity (calibration)	0.973	1.00	1.00	1.00
	Sensitivity (prediction)	1.00	1.00	1.00	1.00
	Specificity (Prediction)	1.00	1.00	1.00	1.00
	Threshold	0.35	0.56	0.45	0.49

The 3D scores plot Figure 46 shows that the PLS-DA model allows an unambiguous separation of different fuel groups. The loading plot (Figure47) indicates that the most significant absorbencies are respectively located between 3050-2700 cm^{-1} and 550-450 cm^{-1} corresponding to the C—H hydrocarbon bonds.

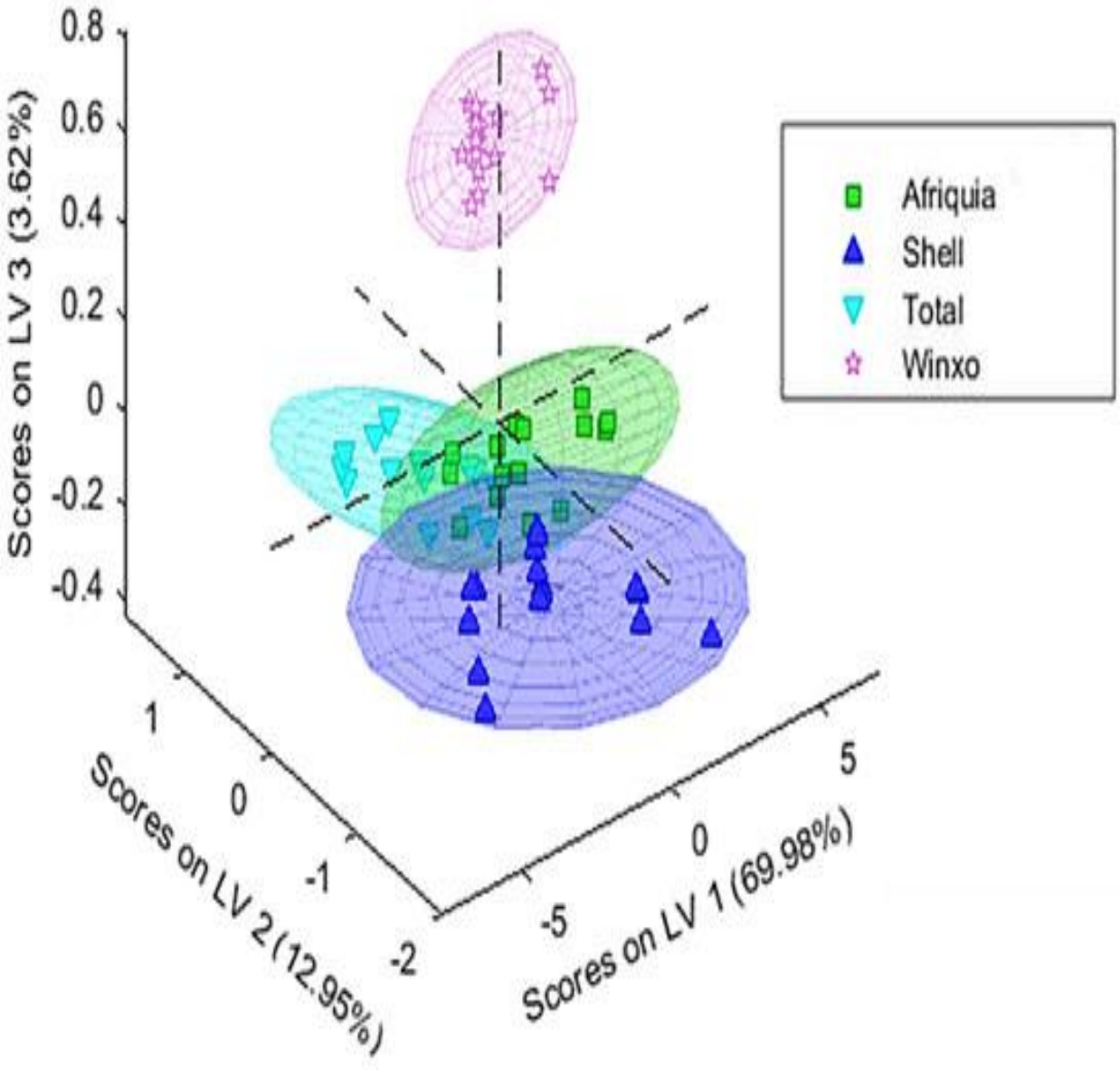


Figure 46: 3D PLS-DA scores plots of the FTIR model

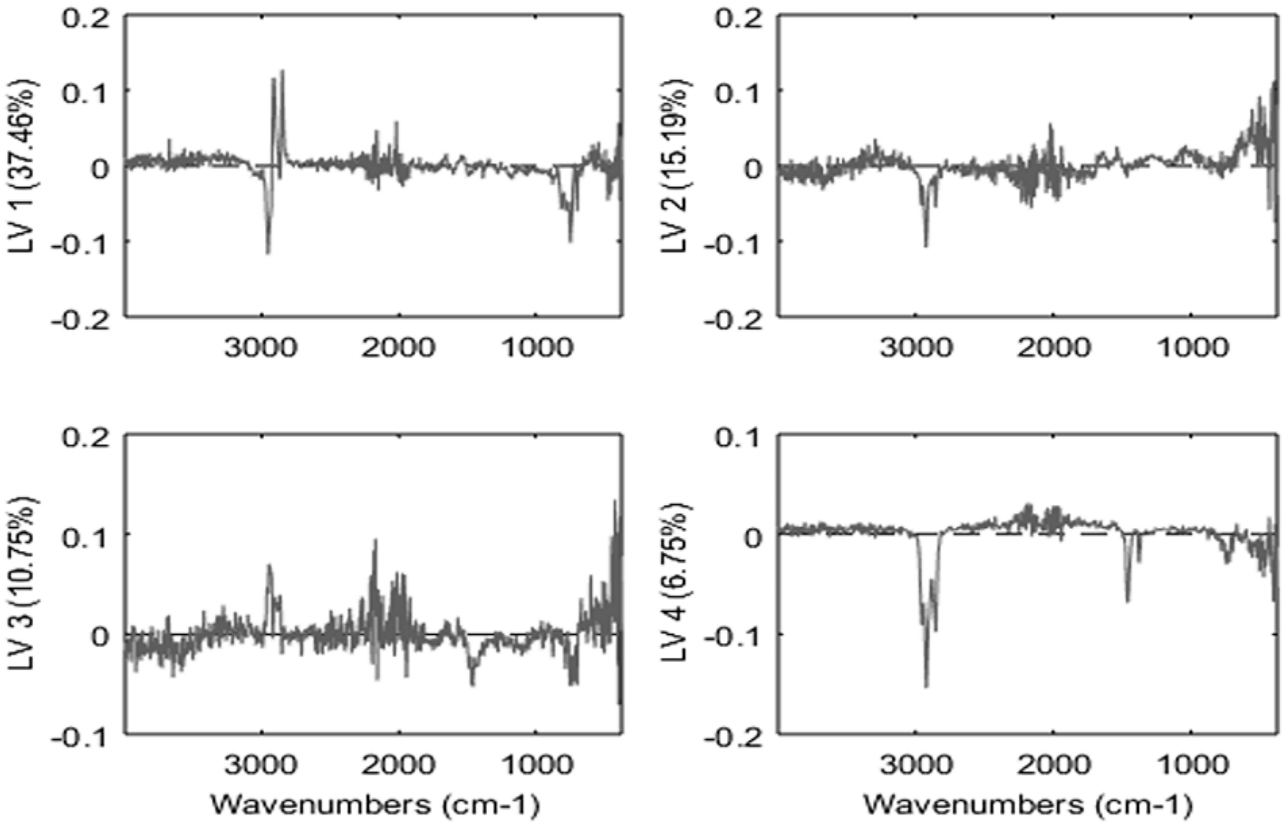


Figure 47: LV1 to LV4 loading plots of PLS-DA-FTIR model

By inspecting the estimated class values of the training and test sets Figure 48, it is noticed that all validation samples are clearly assigned respectively to their proper classes. Thus, the accuracy of the PLS-DA models was confirmed.

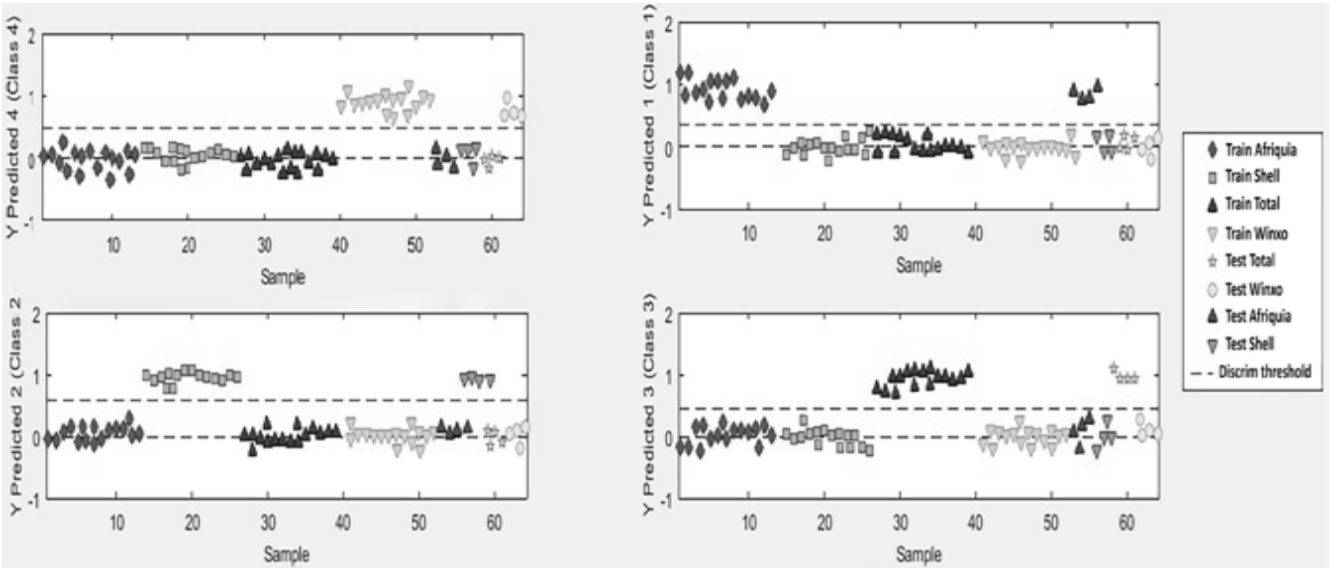


Figure 48: Estimated Class Values of training and test sets obtained by PLS-DA model models established on FTIR spectra

2.7. Classification of diesel fuels by suppliers based on GC-MS Data

2.7.1. PCA analysis

PCA was applied to the pretreated GC-MS data. The first three PCs explained more than 99% of the total variability. From Figure 49, which depicts the configuration of the samples on the first two PCs, we can observe that the four groups of diesel samples are not as clearly identified as in the case of FTIR data. Only the samples from the Shell can be singled out.

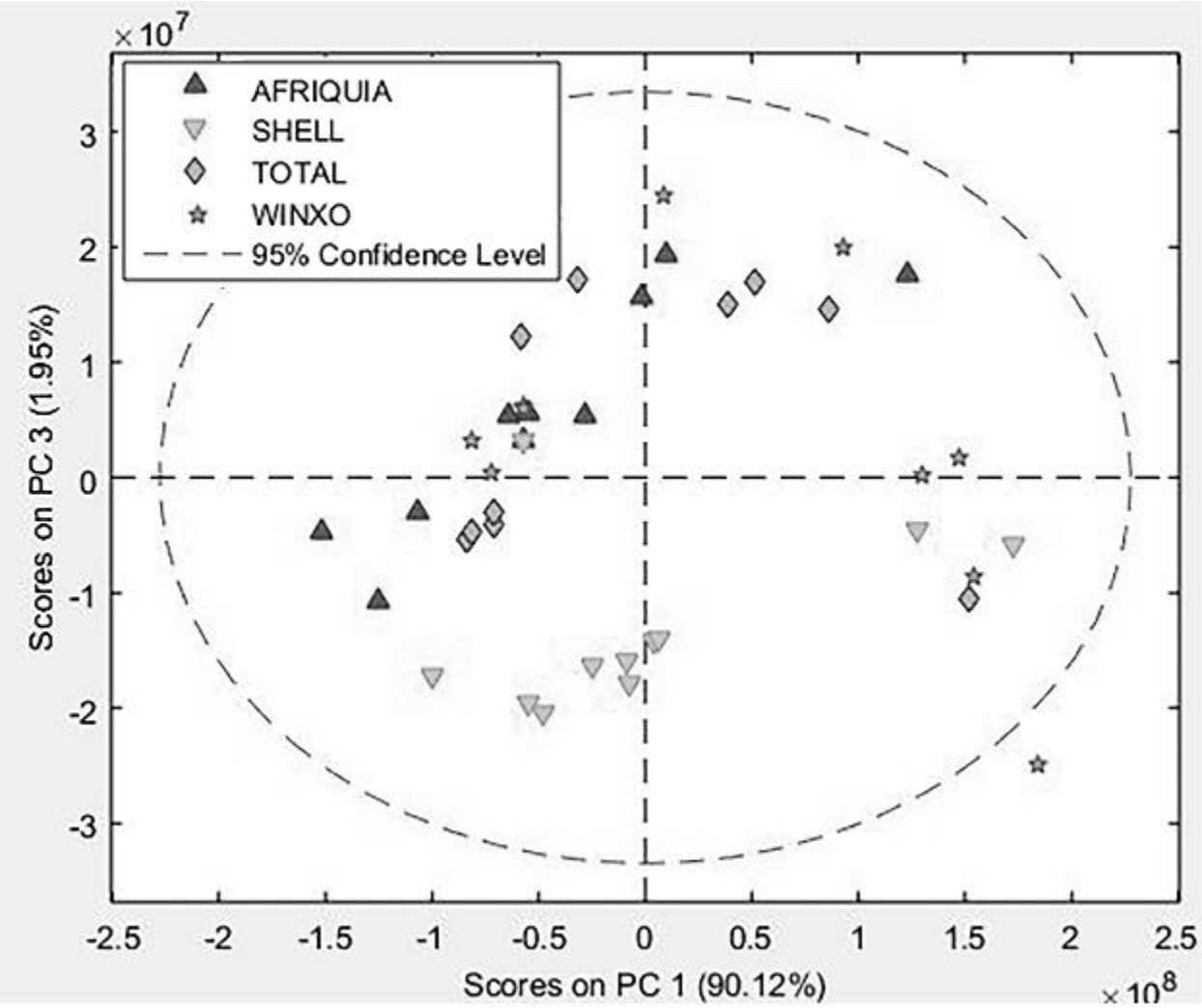


Figure 49: PCA Scores plot, PC1vs PC3, on the GC-MS data

The loadings plot (Figure50) shows that the hydrocarbon chain had the most important weight for PC1. Some chemical compounds as benzene, toluene, benzene-ethyl, dimethyl and trimethyl seem not to have a significant weight for the first PC, but they significantly contribute to the determination of the other PCs.

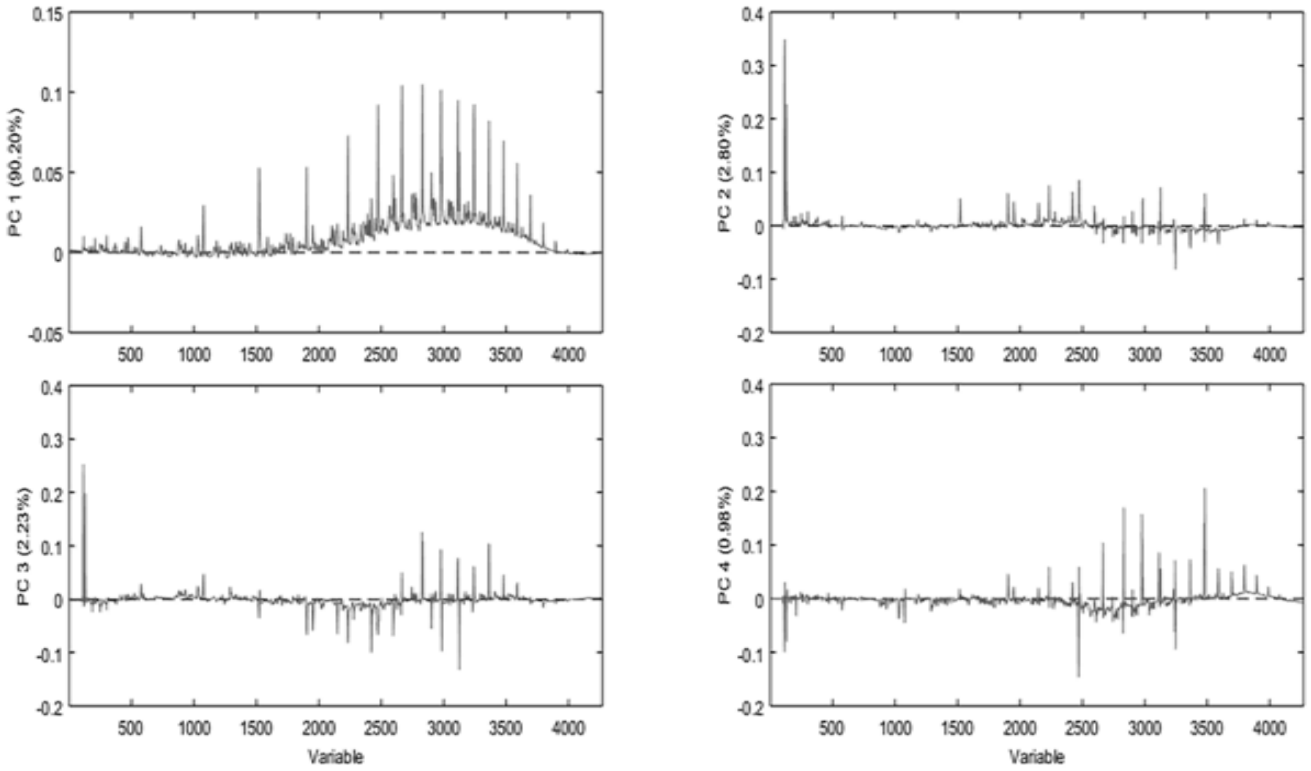


Figure 50: PC1 to PC4 loading plots of the PCA model on GC-MS data

2.7.2. PLS-DA Discrimination

To highlight better the differences between the four suppliers using the CG-MS data, a PLS-DA model was established using the previously pretreated chromatographic data. As expected, the hydrocarbon chain was singled out as having the larger weights in the classification (Figure 51). This finding agrees with the PCA results. The main difference between the samples belonging to the different suppliers is their hydrocarbon composition, which indicates that these brands used different crude oils which are probably originated from different geographical areas.

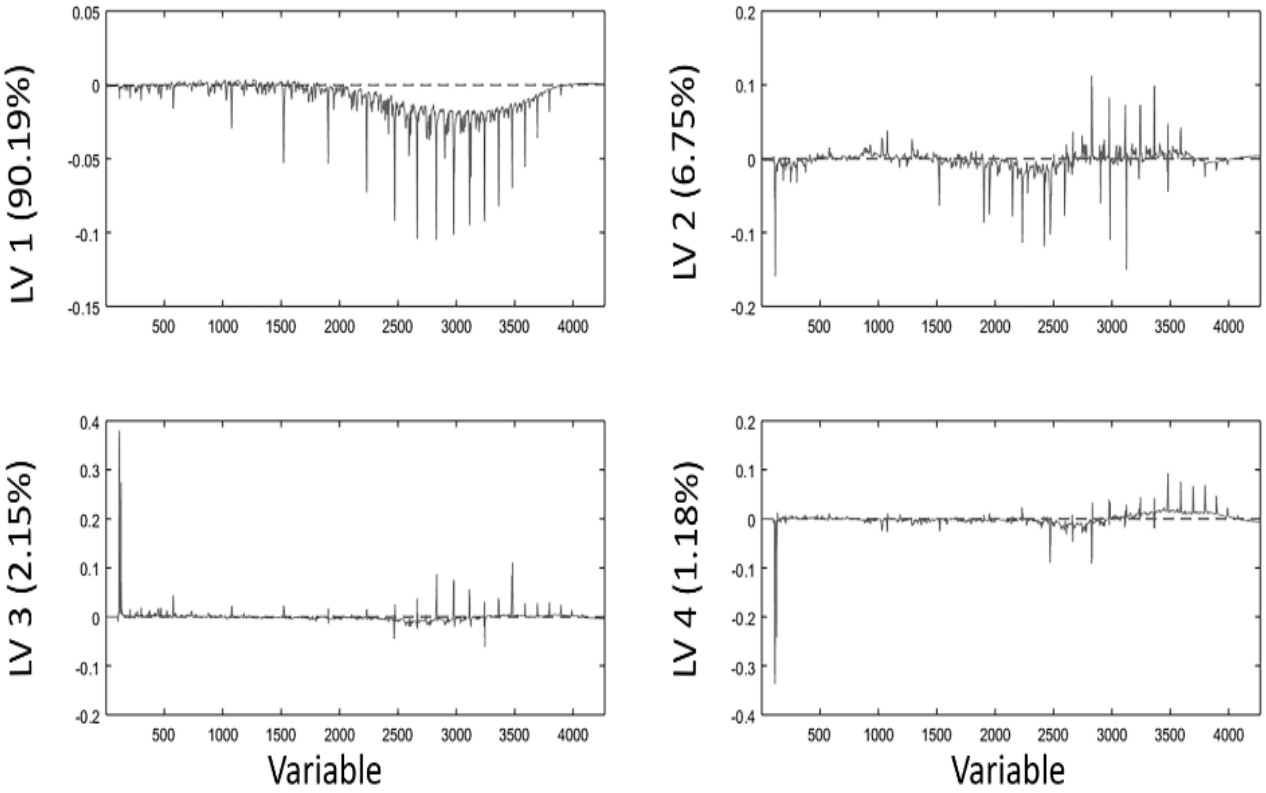


Figure 51: LV1 to LV4 loading plots of the PLS-DA-GCMS model

The 3D PLS-DA scores plot Figure 52 shows a clear separation between the four groups of samples. As indicated in Figure 53 and the Table 13, the 16 test samples were clearly assigned to their own classes with good predictive parameters

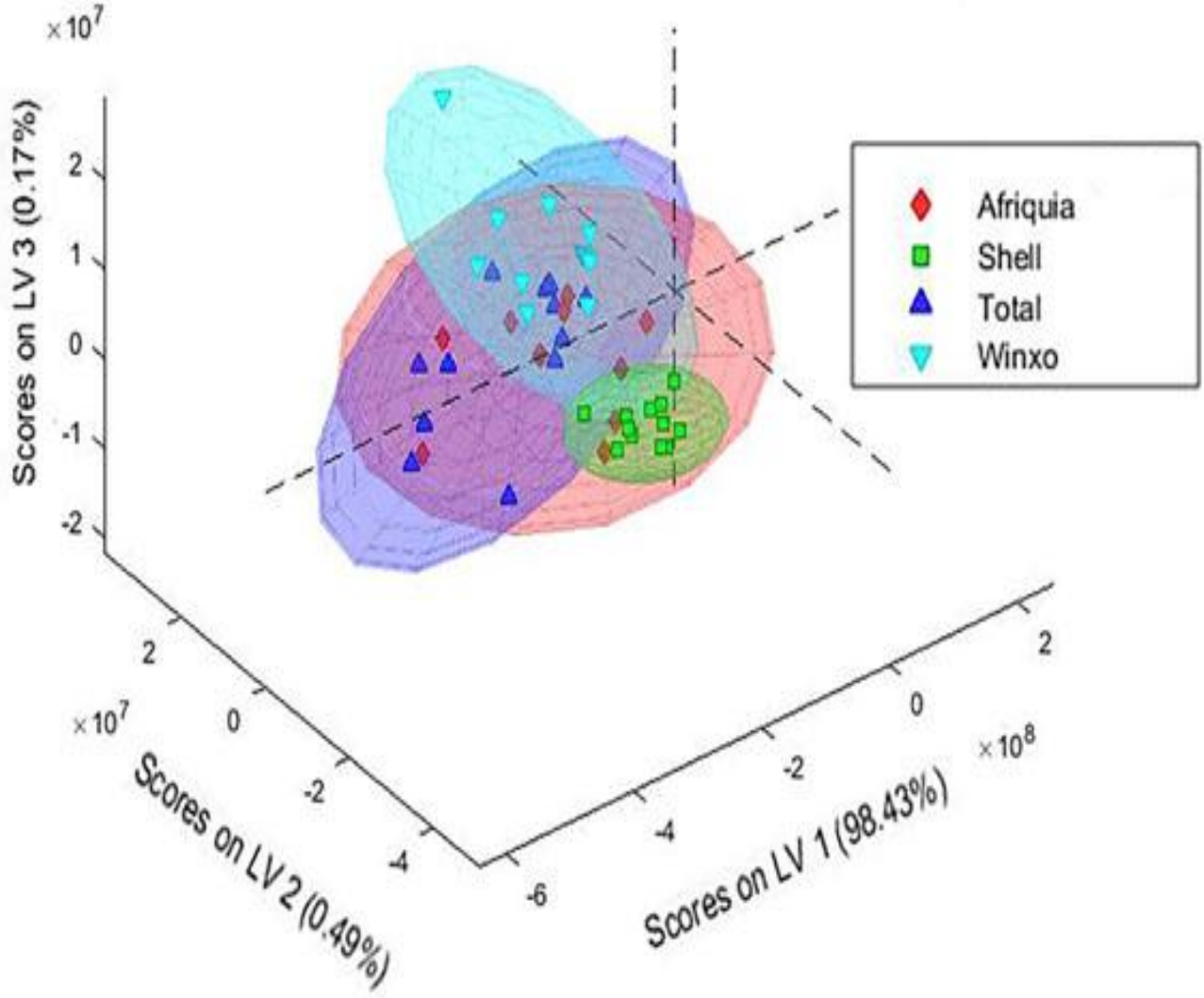


Figure 52: 3D PLS-DA scores plots of the GC-MS model

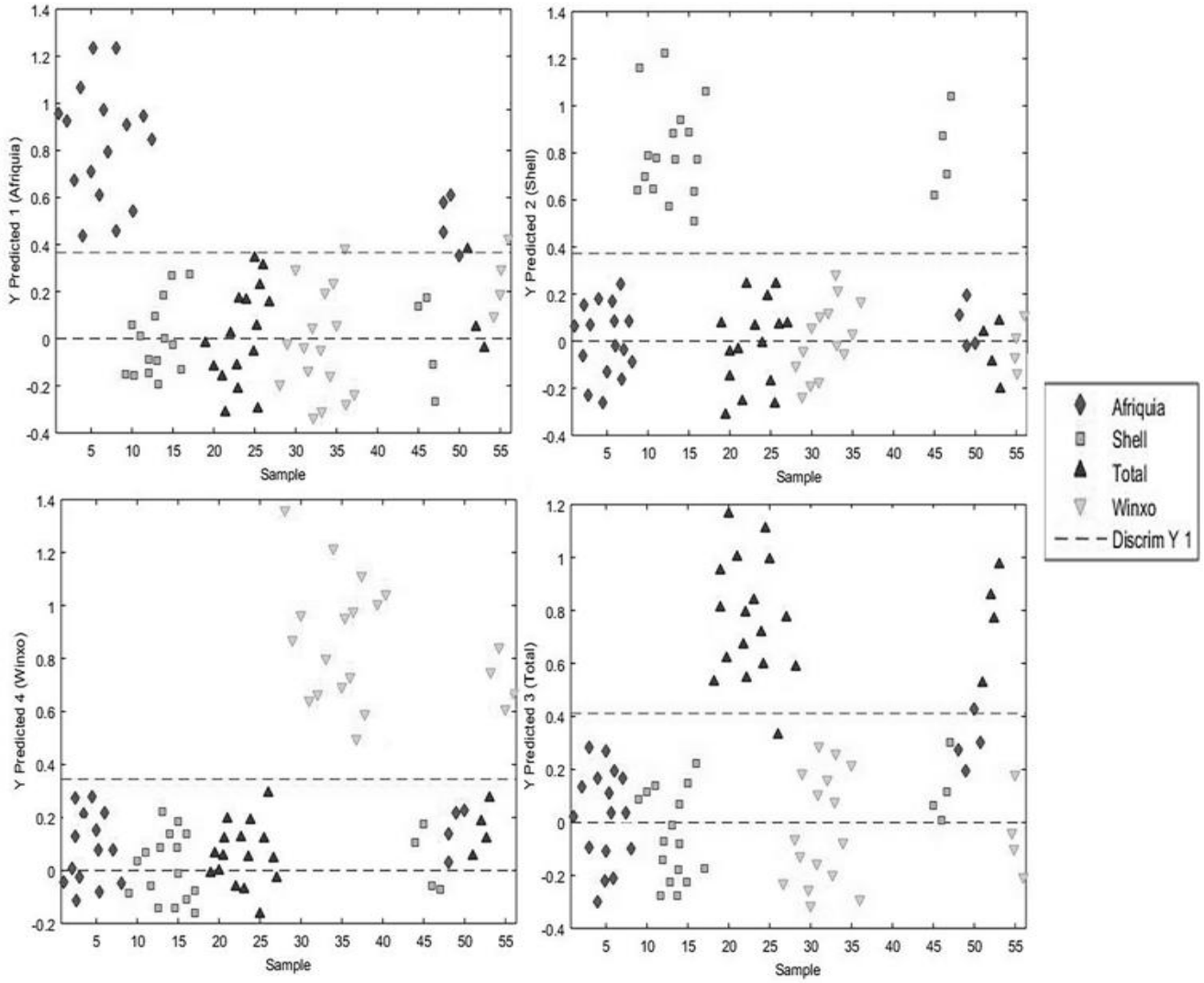


Figure 53: Estimated Class Values of training and test sets obtained by PLS-DA model models established on GC-MS data

Table 13: Discrimination parameters of the GC-MS PLS-DA models

Technique	Param\Fuel	Afriquia	Shell	Total	Winxo
GC-MS	**LVS**	7	7	7	7
	RMSEC	0.20	0.12	0.21	0.17
	RMSEP	0.44	0.15	0.46	0.25
	Sensitivity (calibration)	1.00	1.00	0.89	1.00
	Specificity (calibration)	0.96	1.00	1.00	1.00
	Sensitivity (prediction)	0.87	1.00	1.00	1.00
	Specificity (Prediction)	0.95	1.00	0.88	1.00
	Threshold	0.38	0.39	0.36	0.41

2.8. Conclusion

The present work investigated the possibility of classifying diesel fuels by suppliers. 80 diesel samples were studied, their chemical composition was characterized both by recording their FTIR and GC-MS fingerprints.

PCA and PLS-DA models on FTIR spectra successfully discriminated the four groups of products. Moreover, all the test samples were unambiguously predicted. The obtained PLS-DA models were characterized by the high sensitivity, selectivity and predictive statistical values.

Nevertheless, the acquisition of the FTIR spectra is much easier, cheaper and faster than the GC-MS analysis of the samples. Thus, the FTIR technique should be selected as a fast and reliable solution for the classification of diesel fuels by suppliers.

3. Discrimination and quantification of Moroccan gasoline adulteration with diesel using Fourier Transform Infrared Spectroscopy and Chemometric tools

3.1. The aim

In this work a supervised chemometric model was developed, by coupling FTIR spectra to PCA and PLSR calibration, for the detection and quantification of unleaded gasoline adulteration with diesel.

3.2. Methodology

After obtaining pure unleaded gasoline and diesel samples, a group of 100 mixtures (1ml) (from 0% to 98%) were prepared by adding diesel to gasoline in small vials. Then in order to build the multivariate calibration model, the adulterated gasoline is set to 80 samples for training and 20 for validation.

3.3. Data preprocessing

To improve the predictions quality, two types of data pretreatments were applied, namely the Mean Centering and the Standard Normal Variate transformation.

3.4. FTIR spectra

By simple visual inspection of spectra corresponding to different ratio of adulterated gasoline samples shown in Figure 54, it can be seen that it is difficult to distinguish differences between spectra.

Also, it can be observed the existence of three clusters of spectral bands. The first group appears in the spectral region between 2800 cm^{-1} and 3000 cm^{-1}, where the prominent band at 2920 cm^{-1} is attributed to symmetric CH_3 stretch. The band at 2850 cm^{-1} attached to symmetric CH_2 stretch and the band at 2950 cm^{-1} corresponds to asymmetric CH_2 stretch. The second cluster of bands in the spectral region between 1200 cm^{-1} and 1500 cm^{-1}, where the most important band at 1450 cm^{-1} can be assigned to the CH_3 asymmetric deformation, the signal at 1400 was attributed to the CH_2 scissor bending vibration and the 1380 cm^{-1} assigned to the CH_3 symmetric

deformation. The last spectral bands group localized at 900–700 cm^{-1} region is due to C_1—C_2 stretching (108,113).

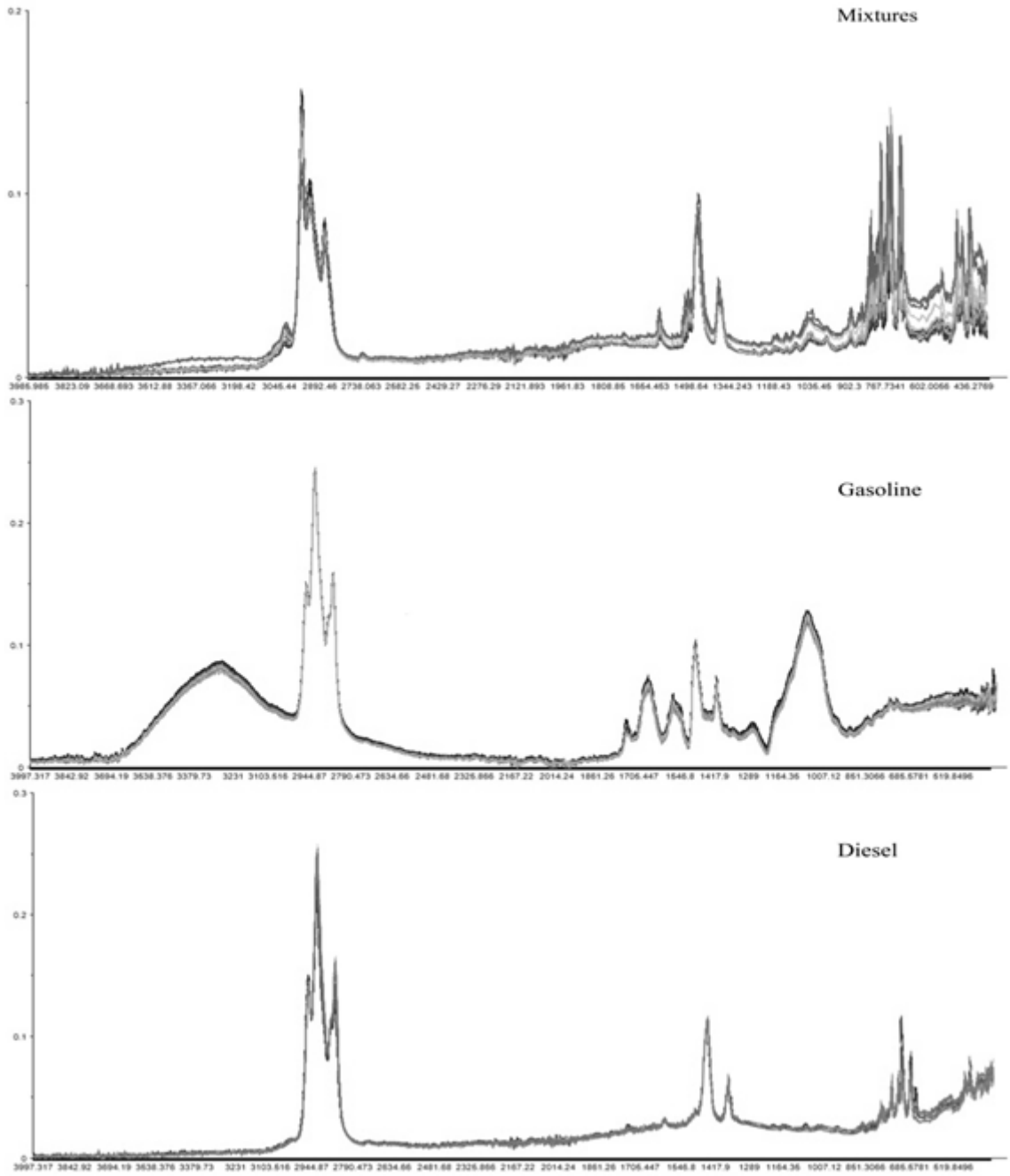


Figure 54: Fourier transform infrared spectra of three types of samples: gasoline, diesel, and adulterated gasoline samples with diesel

3.5. Principal Component Analysis

The first part of this study treats the visualization of data and verification of the possibility of distinguishing between the three types of products: two pure products of gasoline and diesel, the thirds are adulterated samples by the PCA, this is why principal component analysis with full cross validation was applied to the data set of pure and adulterated samples exploring the full acquired spectral range (4000–400 cm^{-1}).

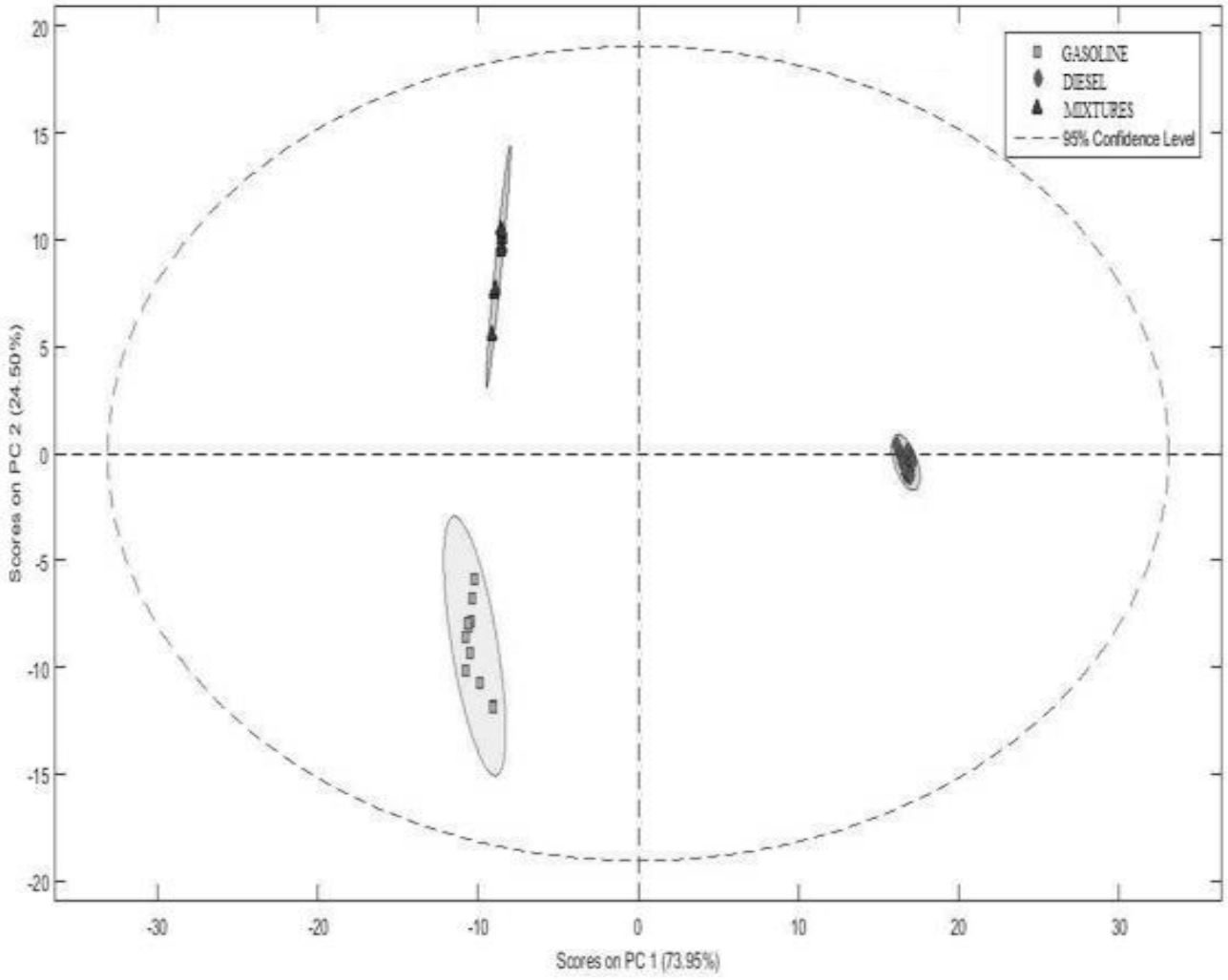


Figure 55: Scores graph for the first two principal components obtained for the discrimination of three classes of samples: Gasoline, Diesel and Adulterated samples

The scores graph obtained (Figure 55) shows that the two first principal components PC1 and PC2 explained more than 98% of the original variability, and displays the formation of three groups, one formed by unadulterated gasoline samples, the second by diesel ones and the third by adulterated samples.

3.6. Quantitative determination of gasoline adulteration with diesel amount by PLSR model

After detecting the adulteration of a gasoline sample by PCA, the level of adulterant was predicted by a PLSR model. Calibration model development was performed using pre-processed spectra to eliminate noise and scattering by applying two types of data pretreatments Mean Center Algorithms and Standard Normal Variate transformation (SNV). A set of twenty validation samples with known content of diesel was used to evaluate model's performance by studying four figures of merit (Number of retained latent variables, RMSEC, RMSEP and percentile of explained variance) presented in Table 14.

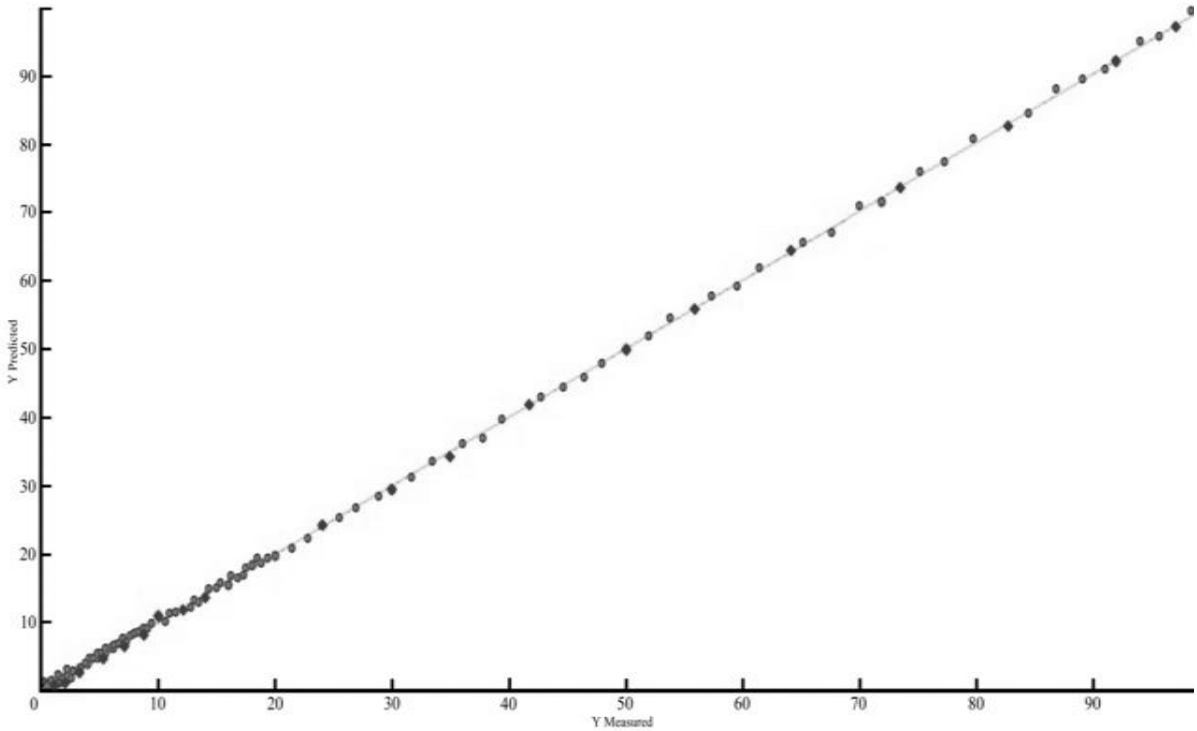


Figure 56: Predicted versus real adulterant contents graph obtained by the PLS model for calibration set (black) and test set (Red).

Table 14: Performance of PLS model for determination of adulterant levels in gasoline samples

	N° of LVs	R^2	RMSEC	RMSEP
PLSR	5	0.99	0.63	0.69

The R^2 (99,25 %) value explains the good linearity of the model built (Figure 56) and indicate a good agreement between predicted and real levels of adulterants in gasoline for calibration and validation data sets. The reduced RMSEC and RMSEP values (0,63 and 0,69) obtained with five Latent Variables, indicate a high accuracy of the model in terms of calibration and prediction.

3.7. Conclusion:

This work showed that the combination of FTIR with chemometric tools have a high usefulness to discriminate and quantify gasoline adulteration with diesel. The detection of adulterated samples was easily highlighted by the use of PCA. On the other hand, PLSR model has shown a good accuracy to predict the amount of adulterant in the range 0% to 98% w/w.

This approach based on coupling FTIR to PLSR can be useful in routine quality control tests of adulteration problems in the field of petroleum products.

4. The association of the FTIR spectroscopy with PLS2 and PLS1 algorithms for the prediction of ten quality parameters of diesel fuels

4.1. The aim

The operation of controlling the quality parameters of petroleum derivatives has become an exigency in order to ensure the good quality of the marketed products.

Besides the official methods known by their complexity, expensiveness and the need of highly qualified personnel, several research projects have been conducted to produce prediction models based on spectroscopic techniques coupled with chemometrics tools and especially PLS2 models, which aim to facilitate and simplify the quality control of petroleum products by the option to predict multiple parameters simultaneously as shown in the table 15(87,114–116).

In this work, multivariate calibration technique (PLS2) was applied on the FTIR spectra for the simultaneous prediction of ten essential quality parameters of diesel fuels. Depending on the result, then the PLS1 strategy will be used in order to improve the prediction of the parameters that were not predicted perfectly.

Table 15: Overview of the models developed for the prediction of diesel quality parameters using NIR spectroscopy

Property	Farquharson et al.(114)		Morris et al.(115)	
	RMSEC	R^2	RMSECV	R^2
Density (15C), g/mL	0.0034	0.88	0.0024	0.96
Cetane Index	0.69	0.92	1.5	0.82
Viscosity (40C), cSt	0.068	0.88	0.195	0.85
Flash Point,°C	1.3	0.94	8.9	0.22
Boiling point at 50%	2.8	0.85	6.1	0.80
Boiling point at90%	3.5	0.75	8.1	0.46
Pour point, °C	-	-	5.0	0.30

4.2. Methodology

The total of the acquired X matrix data (FTIR spectra) ware used to set up The PLS2 models. the responses matrix Y corresponds to the ten responses to be predicted, which are the viscosity, density, color, flash point, conductivity, cetane number, cold filter plugging point, pour point, 40% of the distillation fraction and water contain in diesel fuel. The Excel RANDBETWEEN Function was used to divide the dataset into a training set and a test set. The first calibration set contained 40 samples. Whereas the test set was formed of 10 samples.

4.3. FTIR spectra

As shown in the FTIR spectra of diesel samples Figure 57, two highest absorbencies between 2900 and 2830 cm^{-1} correspond respectively to the symmetric $CH_3$and the symmetric CH_2 stretches. Other two weak absorbencies between 1445 and 1400 cm^{-1} respectively attributed to the CH_3 asymmetric deformation and the CH_2 scissor bending vibration(108).

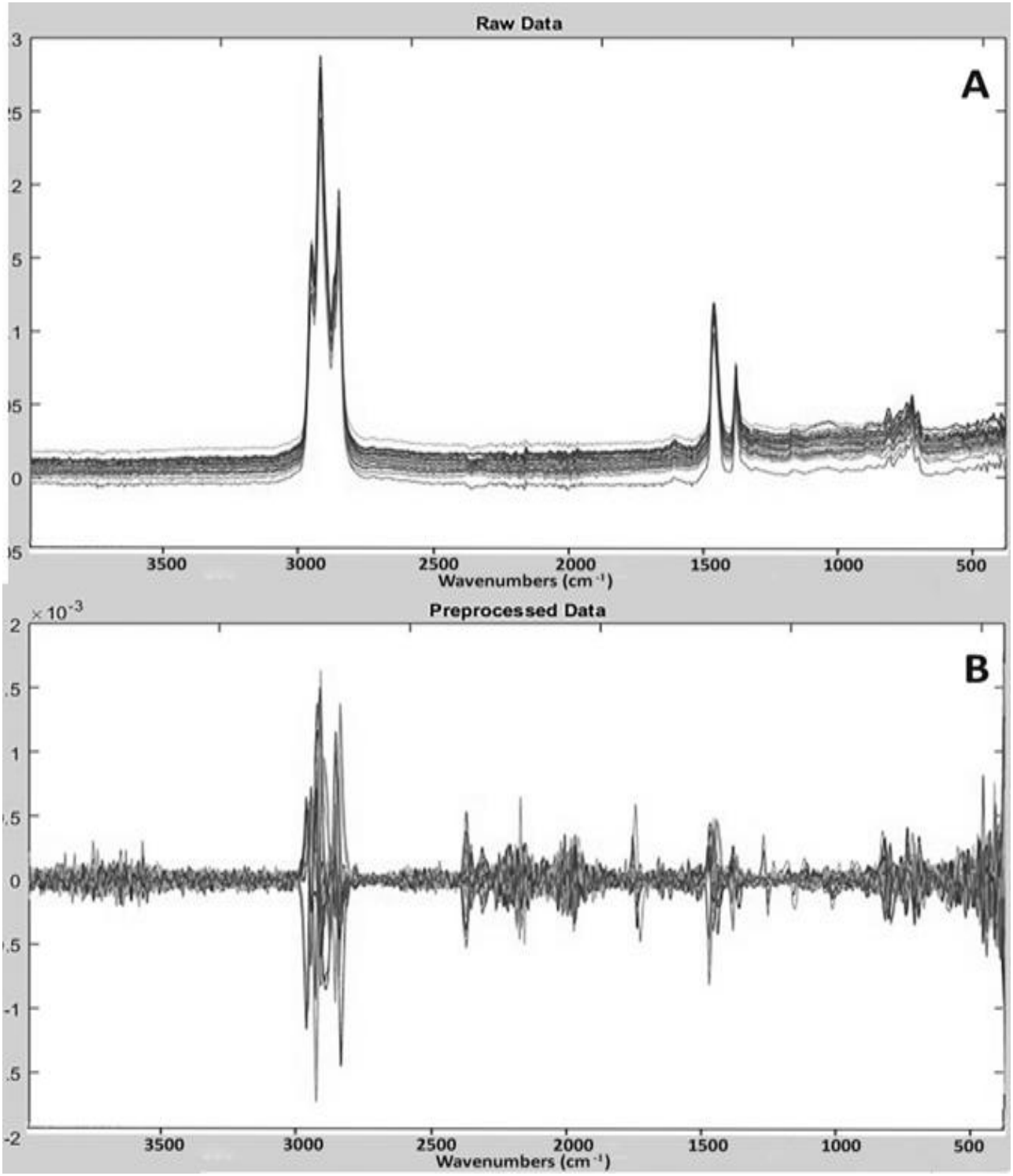


Figure 57: FTIR spectra of diesel samples before preprocessing (A) and after applying the 1st derivative and the mean-centering pretreatments (B)

4.4. Data preprocessing

Two data preprocessing were applied to the crude FTIR data in order to improve the predictive abilities of the PLS2 models (117). Namely, the 1st Savitzky-Golay Derivative and mean-centering (47). Whereas, the responses data were Autoscaled in order to give them the same magnitude order. The statistical pretreatment were described in the part 1.2 of the chapter IV.

4.5. Data visualization by PCA

Once preprocessed, the PCA was used to reveal the correlations between the **Y** variables. Figure 58 shows the scores plot depicting the structure of the ten variables based on the first two principal components (PCs), explaining around 54% of the total variance. It pointed out the existence of some correlations especially between the four parameters related to the temperature (cetane number, cold filter plugging point, pour point and 40% of the distillation fraction) those variables were mainly explained by the 1st PC. The color was totally explained by the 2nd PC.

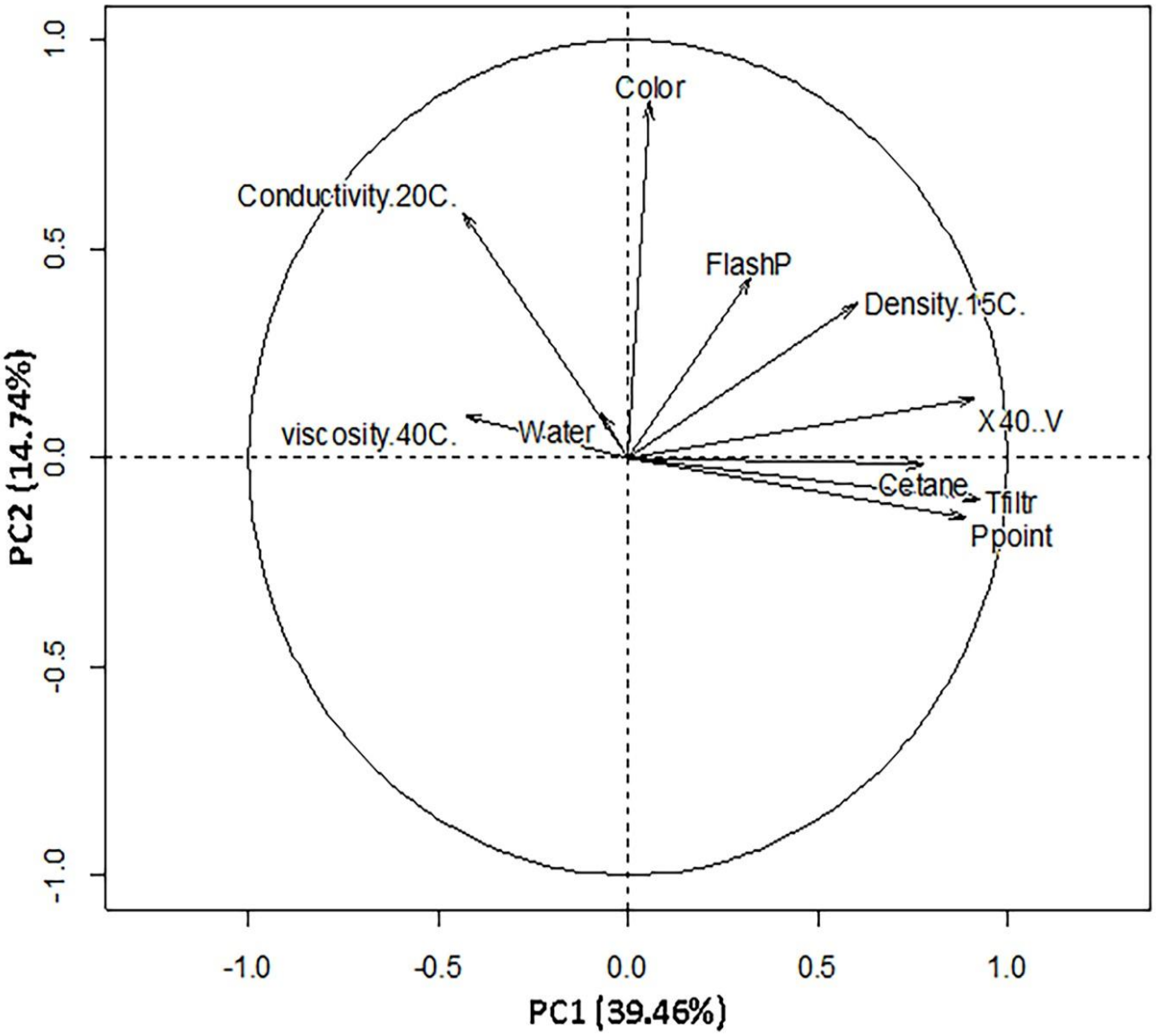


Figure 58:PCA scores plot (PC1 Vs PC2) of the ten responses

On the other hand, the variability behavior and the outliers assessment were performed by projecting the preprocessed FTIR data on PCA. The PC1 Vs PC2 scores plot, Figure 59, explains nearly 85% of the total variance and indicates that the dispersion of the swarm of individuals is normal and depicts the existence of two groups of samples scattered from left to right on the PC1, this information too important when dividing the data into the calibration and

test sets. The choices must be performed randomly in such way that the test samples must belong to both groups and not be chosen only from the group on the left or vice versa.

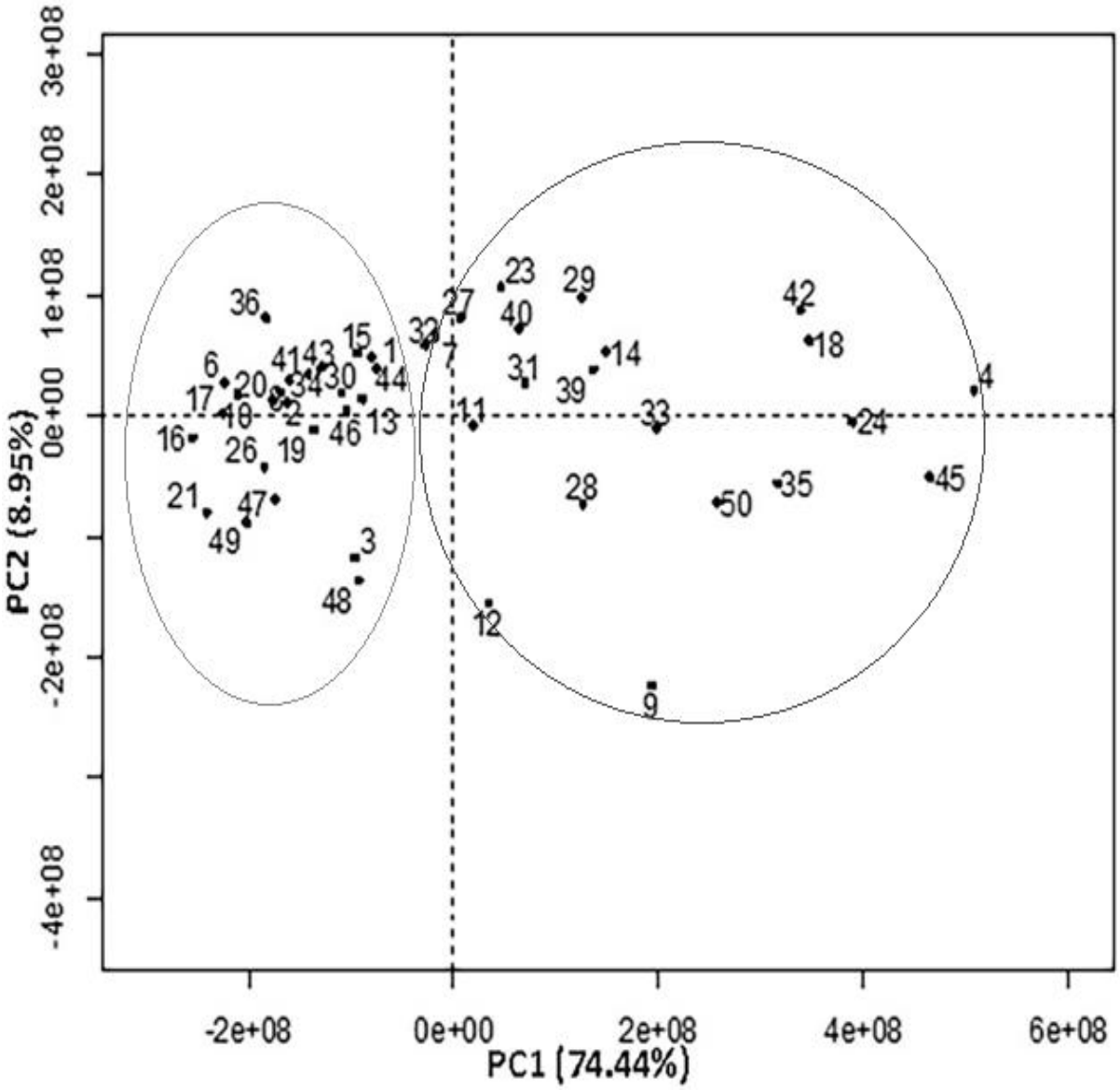


Figure 59: PCA scores plot (PC1 Vs PC2) based on FTIR spectra of the diesel samples

4.6. Simultaneous prediction of ten quality parameters of diesel fuels using the PLS2 models

In order to perform a simultaneous rapid prediction of ten diesel fuel parameters, namely, the viscosity, density, color, flash point, conductivity, cetane number, cold filter plugging point, pour point, 40% of the distillation fraction and water contain, PLS2 models were set up.

The training of the multivariate models was done based on the preprocessed spectral data of 40 diesel samples chosen randomly. Ten Latent Variables were selected.

The quality assessment of the model was done based on three parameters which are the modeling errors (RMSECV and RMSEP) the linear correlation coefficients of predicted versus measured responses.

It is possible to note that the multivariate calibration allows a satisfactory determination of the sought responses. On the one hand the correlation coefficients between real and predicted values were generally around 0.90. On the other hand, similar and low RMSECV and RMSEP values table 19 were obtained in comparison with those presented in the bibliography (table 16) indicating that this model can be accepted for future predictions of the considered parameters in new diesel samples.

According to the figure 60 a good linear correlation has been noted between the real and the predicted values by the model especially for the viscosity, CFPP, flash point, density, water contain and for the temperature needed for the 40% of atmospheric distillation.

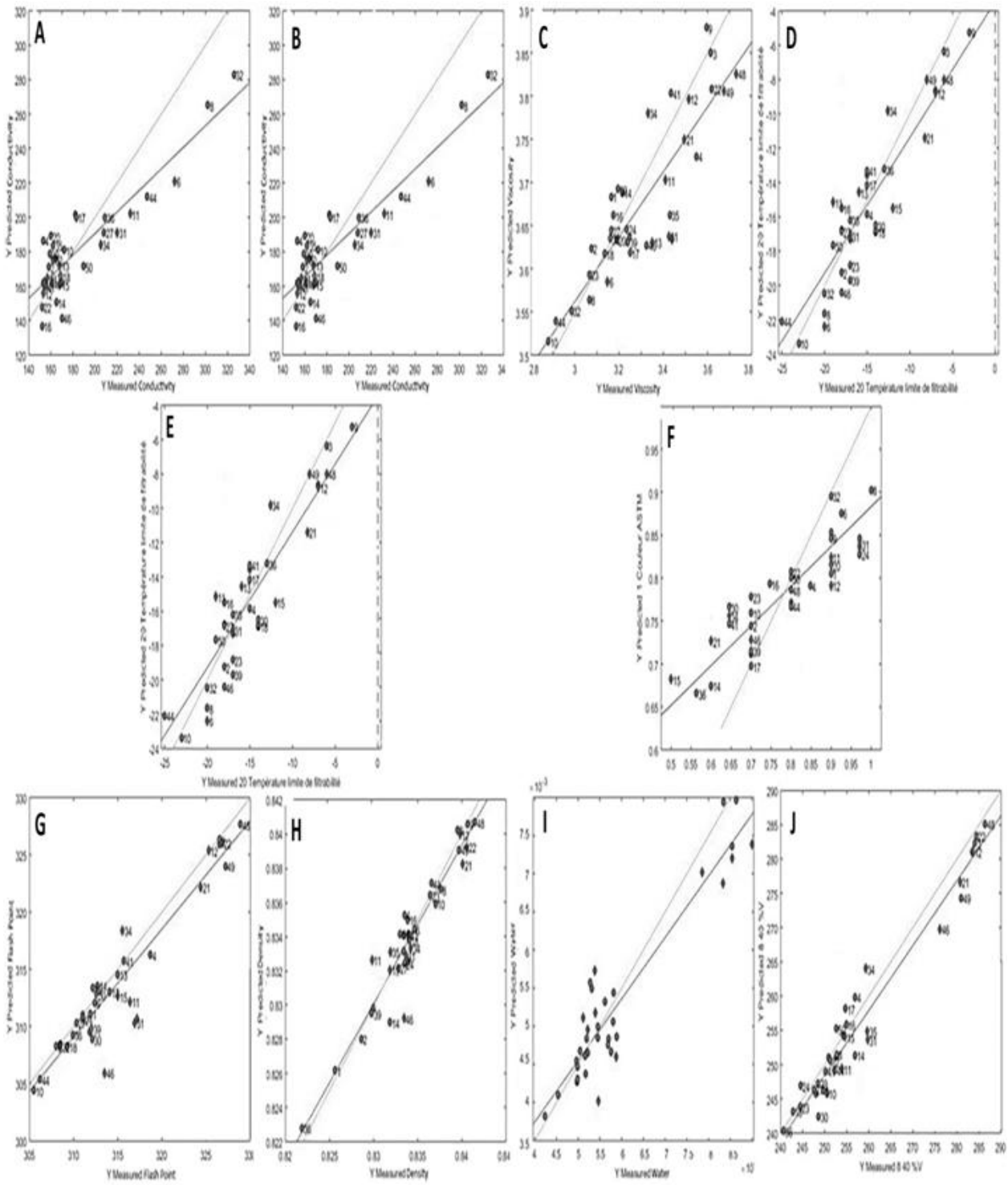


Figure 60: Predicted versus real values of: (A): Conductivity, (B): Pour Point, (C) Viscosity, (D): CFPP, (E): Flash Point, (F): Color, (G): Cetane Number, (H): Density, (I): Water contain, (J): 40% Distillation

Table 16: Results obtained by the PLS2 model

Parameter	LVs	RMSECV	RMSEP	R^2
Color	10	1,06	1.06	0.90
Density	10	0.73	0.73	0.89
Boiling point at40% recovery	10	0.52	0.52	0.91
FlashPoint	10	1.16	1.16	0.92
Pour Point	10	1.05	1.04	0.92
Cold Filter Plugging Point	10	0.99	0.98	0.92
Viscosity at 40°C.	10	1.19	1.18	0.92
Cetane Number	10	1.07	1.06	0.83
Conductivity at 20°C.	10	1.10	1.10	0.91
Water contain	10	1.19	1.19	0.78

4.6.1. PLS1 model for the cetane number prediction

The cetane number (CN) characterizes the self-ignition ability of a diesel fuel. In diesel engine, the fuel is injected under high pressure into the compressed air and the ignition occurs spontaneously after a very short time, in the order of 1 millisecond. This process is favored by increasing the compression ratio which can vary between 15 and 22, but it is also necessary that

the fuel has a chemical structure favorable to auto-ignition. This quality is expressed by the CN(118–121).

The importance of this essential parameter for certifying the quality of diesel has prompted us to search to improve the prediction of CN. In order to predict the cetane number of diesel fuels in the range 49-59, a Partial Least Square Regression 1 model was established based on the same FTIR data matrix previously used in the PLS2 model.

The selection of the optimal number of latent variables required for the good prediction is done based on the Root Mean Squared Error of Prediction values which are calculated using external validation set. In our case 8 LVs were selected with RMSEP of 0.49.

The RMSEP value obtained by our PLS model was lower than those presented by Helga G. Aleme et al.(100) (RMSEP=0.50), Fodor et al.(103)(RMSEP=1.3), Soyemi et al.(122) (RMSEP=1.23), Santos Jr. et al.(123) (RMSEP values of PLS models were 0.71, 0.67 and 0.58 respectively for FTIR-ATR, FTNIR and FT-Raman), which reflects the good predictive quality of the model set up using the coupling of FTIR spectroscopy and PLS multivariate calibration.

On the other hand, the high determination coefficient R^2 (0.99) explains a good correlation between the real and predicted cetane number values figure 59and table 17.

Table 17: PLS model results for prediction of cetane number in diesel samples

	Samples splitting		N° of Latent Variables	R^2	RMSEC	RMSEP
	Calibration	Test				
PLSR	40	10	8	0.99	0.08	0.49

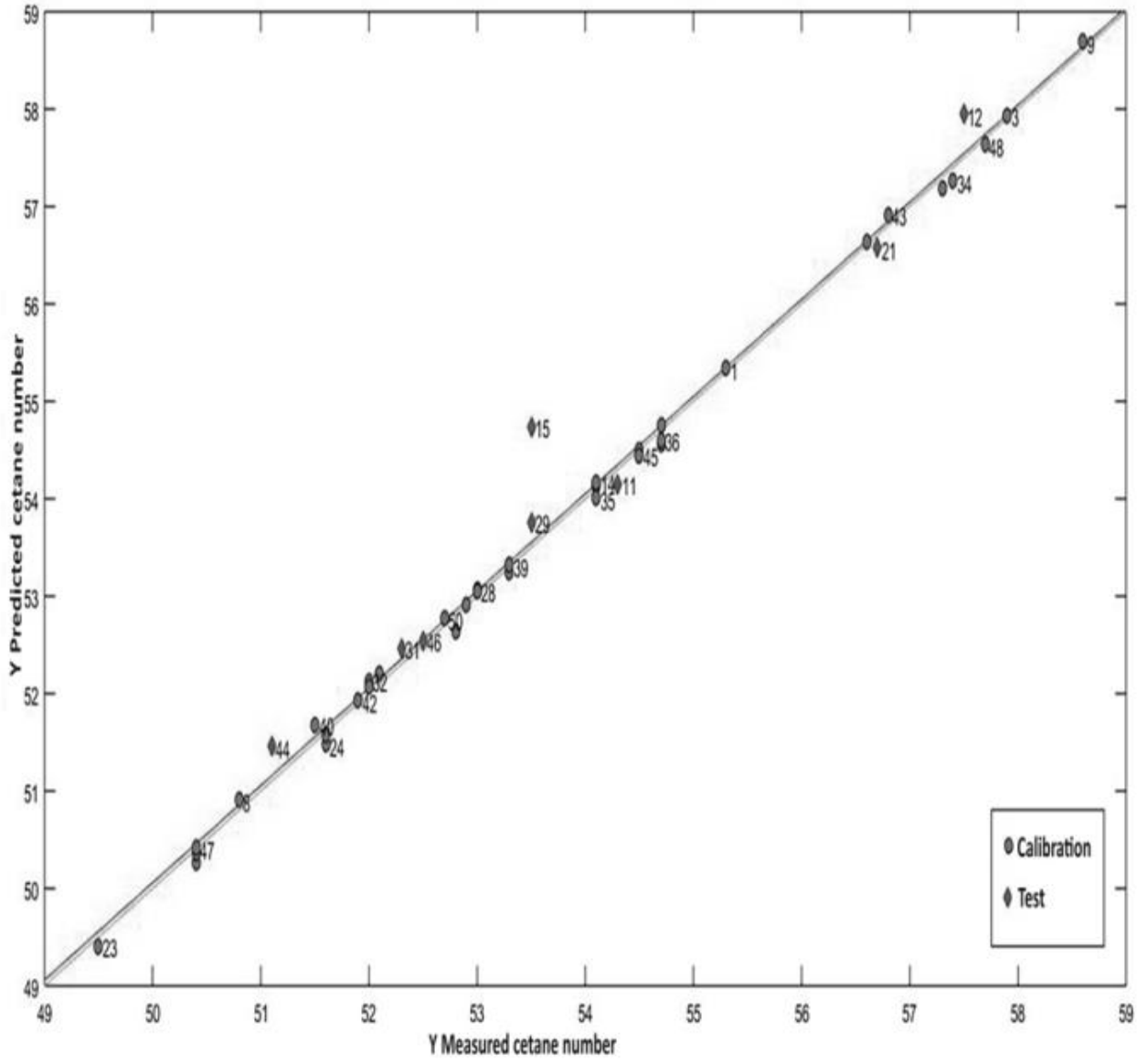


Figure 61: Predicted versus real Cetane number values graph obtained by the PLS-FTIR model

4.7. Conclusion

The present study investigated the development of a multivariate analysis based on FTIR spectroscopy as a technique of choice to answer the analytical needs in the field of quality control of petroleum products.

The predictive models that were calibrated using preprocessed spectral data showed an accurate prediction of ten quality control parameters of diesel fuels. On the one hand, the PLS2 models were characterized by high determination coefficients between the real and the predicted values and acceptable errors values (RMSECV and RMSEP) especially for the Boiling point at40% recovery, Flash Point, Pour Point, CFPP and the Viscosity at 40°C in comparison with similar multivariate models developed based on spectral techniques.

On the other hand, In order to improve the prediction of the cetane number, as one of the most important parameters to attest the quality of diesels, a PLS1 model was calibrated. The quality assessment of this multivariate model was considered perfect, characterized by very significant determination coefficient (0.99) and low prediction error (0.49).

Proving that the proposed approach can replace the usual laboratory methods based on carrying out the different tests one at a time, especially when monitoring the quality of the products during production in the refinery.

The low number of samples seems to be the main cause of prediction weakness of some parameters (water contain and the cetane number). A collection of new samples should therefore be highly recommended to improve the predictive quality of the models.

CONCLUSION AND PERSPECTIVES

The present work has as purpose to develop chemometric models based on the FTIR spectroscopy technique in order to authenticate and predict the quality parameters of petroleum products marketed in Morocco. On the other hand, to crack down on the frauds in this vital field and make the mission of quality control of these products easier and faster.

The set up of multivariate analysis applied on the Fourier Transform Infrared Spectroscopy data was investigated in order to meet the analytical requirements.

As a result, the FTIR technique has many advantages. First, the rapidity in terms of spectra acquisition, inexpensive and does not requires sample preparation especially when equipped with the ATR accessory. And secondly, the obtained spectra contain huge hidden information for hydrocarbon which has been assessed for the characterization and the prediction of diesel fuel Moroccan's parameters.

However, responding to the objectives of this thesis, our approach was therefore articulated in four axes:

First of all, a PLS-DA model was established for the rapid detection of smuggled diesels. The samples used in this study were divided into two parts, a calibration group that was used to set up the model and a 2^{nd} part for testing the model. The model allowed an unambiguous classification of the two classes of products (authentic and contraband), and was characterized by a good sensitivity and selectivity.

Then in order to verify the potential of the FTIR spectroscopy for the classification of diesel samples from four suppliers which will allow to ensure the traceability of diesel products marketed in Morocco, a comparison was made between a PLS-DA model established on the basis of the FTIR spectra and another based on the gas chromatographic data. The results of this study showed that the two models led to a good classification of the four classes of diesel, but it was noticed that the coupling FTIR-PLSDA was better because the prediction of the different classes was made with 100 % accuracy.

Afterward the problem of the detection and the quantification of unleaded gasoline adulteration with diesel was treated by the set up of a PLS regression model built on FTIR spectra. It was noticed that the prediction of the amount of adulterant in the range 0% to 98% w/w was so successful with a reduced error of prediction value (RMSEP=0.69) and a significant determination coefficient (R^2=0.99).

The last part of our thesis concerns the application of the PLS2 calibration approach combined with the FTIR spectroscopy for the prediction of ten important properties for the certification of diesel quality. Multivariate calibrations of properties of interest which are the viscosity, density, color, flash point, conductivity, cetane number, cold filter plugging point, pour point, 40% of the distillation fraction and water contain were considered satisfactory in terms of prediction error and the correlations. Regarding the cetane number which is a very important parameter that provides information on the ability for self-ignition of diesel fuels, and in order to improve its prediction, A PLS1 model was set up based on the same spectral data. The model was considered successful, characterized by a goodness of fit (R^2 = 0.99) and reflects a good agreement between the real and predicted values, and a low value of prediction error (RMSEP = 0.49) which is of the order of uncertainty of the measuring apparatus. The laboratory application of this new analysis approach of diesel will generate a saving of analysis time of about 4 to 5 hours.

In conclusion, the association of the FTIR spectroscopy with proper chemometric data analysis demonstrated its potential to be a perfect solution in the quality control of petroleum products.

Regarding the following of this work, an improvement of the predictive models by feeding the database by more samples is in progress before transferring them for routine use in the laboratory.

References

1. International Energy Agency. World energy outlook 2013. OECD. 2013;92.
2. Ward JW. Hydrocracking processes and catalysts. 1993;35:55–85.
3. Wauquier J-P. Le raffinnage du pétrole. Editions T. 1994. 26-50 p.
4. Ancheyta J, Rana MS, Sa V, Diaz JAI. A review of recent advances on process technologies for upgrading of heavy oils and residua. 2007;86:1216–31.
5. Chikhi S, Boughedaoui M, Kerbachi R, Joumard R. On-board measurement of emissions from liquefied petroleum gas , gasoline and diesel powered passenger cars in Algeria. JES [Internet]. 2014;26(8):1651–9. Available from: http://dx.doi.org/10.1016/j.jes.2014.06.005
6. Kharbach M, Kamal R, Bousrabat M, Alaoui Mansouri M, Barra I, Alaoui K, et al. Characterization and classification of PGI Moroccan Argan oils based on their FTIR fingerprints and chemical composition. Chemometrics and Intelligent Laboratory Systems [Internet]. 2017;162:182–90. Available from: http://dx.doi.org/10.1016/j.chemolab.2017.02.003
7. Barra I, Mansouri MA, Bousrabat M, Cherrah Y, Bouklouze A, Kharbach M. Discrimination and Quantification of Moroccan Gasoline Adulteration with Diesel using Fourier Transform Infrared Spectroscopy and Chemometric Tools. Journal of AOAC International. 2018 Oct 24;
8. Al-Ghouti MA, Al-Degs YS, Amer M. Application of chemometrics and FTIR for determination of viscosity index and base number of motor oils. Talanta [Internet]. 2010;81(3):1096–101. Available from: http://dx.doi.org/10.1016/j.talanta.2010.02.003
9. Bassbasi M, De Luca M, Ioele G, Oussama A, Ragno G. Prediction of the geographical origin of butters by partial least square discriminant analysis (PLS-DA) applied to infrared spectroscopy (FTIR) data. Journal of Food Composition and Analysis [Internet]. 2014;33(2):210–5. Available from: http://dx.doi.org/10.1016/j.jfca.2013.11.010
10. Bassbasi M, Hafid A, Platikanov S, Tauler R, Oussama A. Study of motor oil

adulteration by infrared spectroscopy and chemometrics methods. Fuel [Internet]. 2013;104:798–804. Available from: http://dx.doi.org/10.1016/j.fuel.2012.05.058

11. Linker R, Shmulevich I, Kenny A, Shaviv A. Soil identification and chemometrics for direct determination of nitrate in soils using FTIR-ATR mid-infrared spectroscopy. Chemosphere. 2005;61:652–8.
12. Osborne BG. Near-infrared Spectroscopy in Food Analysis. Encyclopedia of Analytical Chemistry. 2006;1–14.
13. Balabin RM, Safieva RZ. Motor oil classification by base stock and viscosity based on near infrared (NIR) spectroscopy data. Fuel. 2008;87:2745–52.
14. Balabin RM, Safieva RZ. Gasoline classification by source and type based on near infrared (NIR) spectroscopy data. Fuel. 2008;87:1096–101.
15. Mckenna AM, Purcell JM, Rodgers RP, Marshall AG. Heavy Petroleum Composition . 1 . Exhaustive Compositional Analysis of Athabasca Bitumen HVGO Distillates by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry : A Definitive Test of the Boduszynski Model. 2010;44(8):2929–38.
16. AFNOR. NF EN 590 Automotive Fuels - Diesel Engine Fuels (Diesel) - Requirements and Test Methods. 2005;(1):5–6.
17. Speight JG. Handbook of Petroleum Product Analysis. 2nd ed. Wiley, New York,; 2015. 4-19 p.
18. Kraus RS. Oil and Natural Gas [Internet]. PETROLEUM REFINING PROCESS. Available from: www.ilocis.org/documents/chpt78e.htm
19. Jacqueline B. Le pétrole au Maroc. In: Cahiers d'outre-mer. N° 73 - 19e année. 1966. 73-81 p.
20. Samir. Rapport Annuel 2014 [Internet]. 2014. p. 38. Available from: http://www.samir.ma/images/stories/pdf/RapportAnnuel2014VF.pdf
21. Commerce M of. Official Bulletin of the Kingdom of Morocco No. 6454 on the liberalization of the fuel market. 2016. p. 443–8.
22. Berrada R. Carburant: les principaux chiffres du rapport parlementaire [Internet]. 2018. Available from: https://www.medias24.com/MAROC/ECONOMIE/ECONOMIE/182941-

Carburant-les-principaux-chiffres-du-rapport-parlementaire.html

23. Mounadi D. Hydrocarbons: Large legislative dusting. Aujourd'hui le Maroc, [Internet]. 2016; Available from: http://aujourdhui.ma/economie/hydrocarbures-gros-depoussierage-legislatif-120875
24. International Organization for Standardization. ISO 3170 Petroleum liquids — Manual sampling. 2004;
25. American Society for Testing and Materials. ASTM D 1500 – 07 Standard Test Method for ASTM Color of Petroleum Products (ASTM Color Scale. 2009;1–5.
26. American Society for Testing and Materials. ASTM D1298 − 12b Standard Test Method for Density , Relative Density , or API Gravity of Crude Petroleum and Liquid Petroleum Products by Hydrometer Method. 2013;1–8.
27. American Society for Testing and Materials. ASTM D 445 – 06 Standard Test Method for Kinematic Viscosity of Transparent and Opaque Liquids (and Calculation of Dynamic Viscosity). 2009;1–10.
28. American Society for Testing and Materials. ASTM D 97 – 05 Standard Test Method for Pour Point of Petroleum Products. 2:1–9.
29. American Society for Testing and Materials. ASTM D 613 – 05 Standard Test Method for Cetane Number of Diesel Fuel Oil. 2007;
30. American Society for Testing and Materials. ASTM D 976 – 91 Standard Test Methods for Calculated Cetane Index of Distillate Fuels. 2000;91(Reapproved):4–6.
31. American Society for Testing and Materials. ASTM D 86 - 07 Standard Test Method for Distillation of Petroleum Products at Atmospheric Pressure. 1996;
32. American Society for Testing and Materials. ASTM D 6371 – 99 Standard Test Method for Cold Filter Plugging Point of Diesel and Heating Fuels. 1999;
33. American Society for Testing and Materials. ASTM D 4294 – 08a Standard Test Method for Sulfur in Petroleum and Petroleum Products. 2009;1–9.
34. Herschel W. Investigation of the Powers of the prismatic Colours to heat and illuminate Objects. Philosophical Transactions of the Royal Society of London. 1800;90:255.

35. Colthup NB, Daly LH, Wiberley SE. Introduction to Infrared and Raman Spectroscopy. Third Edit. 1990. 547-580 p.
36. Cohen-Tannoudji C, Diu B, Laloe F. Mécanique quantique. tome1 ed. Paris: Hermann; 1996. 480-495 p.
37. Dalibart M, Servant S. Spectroscopie dans l‟infrarouge. Techniques de l‟ingénieur. 2000;2845:1–26.
38. Osborne B, Fearn T, Hindle P. Pratical NIR spectroscopy with application in food and beverage analysis. 2nd ed. Ltd A-WL, editor. Harlow: Prentice Hall; 1993. 273 p.
39. Khoshhesab ZM. Reflectance IR Spectroscopy [Internet]. 1ST ed. (Ed.) PTT, editor. Infrared Spectroscopy - Materials Science. InTech; 2012. 233-244 p. Available from: http://www.intechopen.com/books/infrared-spectroscopy-materials-science-engineering-and- technology/fundamental-of-reflectance-ir-spectroscopy
40. Theophanides T. Introduction to Infrared Spectroscopy. In: Theophile PT, editor. Infrared Spectroscopy - Materials Science, Engineering and Technology [Internet]. 1st ed. InTech; 1993. p. 1–10. Available from: http://www.intechopen.com/books/infrared-spectroscopy-materials-science-engineering-and-technology/introduction-to-infrared-spectroscopy%0AInTech
41. Stuart B. Infrared spectroscopy fundamentals and applications. 1st ed. Hoboken N., editor. J. Wiley: Chichester, Eng; 2004. 45-70 p.
42. Günzler H, Heise HM. IR spectroscopy: an introduction. 1st ed. Wiley G, editor. Wiley-VCH: Weinheim; 2002. 200-219 p.
43. LANTERI P, LONGERA R. Chimiométrie: le mariage réussi entre les sciences analytiques et l'informatique. Analusi. 1996;24:17–27.
44. Dhanoa MS, Lister SJ, Sanderson R, Barnes RJ. The link between Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV) transformations of NIR spectra. 1994;47:43–7.
45. Griffiths PETERR. Effects of Sample Dilution and Particle Size / Morphology on Diffuse Reflection Spectra of Carbohydrate Systems in the Near- and Mid-Infrared

. Part I : Single Analytes. 1993;47(6).

46. The I, Munk K. Piece-Wise Multiplicative Scatter Correction Applied to Near-Infrared Diffuse Transmittance Data from Meat Products. 1993;47(6).
47. Eigenvector. mean centring [Internet]. 2012. Available from: http://wiki.eigenvector.com/index.php?title=Advanced_Preprocessing:_Variable_Centering#Mean_Centering
48. Multispectral Analysis for Quantitative Measurements of Myoglobin Oxygen Fractional Saturation in the Presence of Hemoglobin Interference. 1992;46(12).
49. Ozaki Y, Mizuno A, Sato H, Kawauchi K, Muraishi S. Biomedical Application of Near-Infrared Fourier Transform Raman Spectroscopy. Part I: The 1064-nm Excited Raman Spectra of Blood and Met Hemoglobin. APPLIED SPECTROSCOPY. 1992;46(3).
50. Levillain P, Pompeydie D. Derivative spectrophotometry principles, advantages and limitations, applications. P Levillain, D Pompeydie. 1986;14:1–20.
51. eigenvector. Advanced Preprocessing: Noise, Offset, and Baseline Filtering [Internet]. Available from: http://wiki.eigenvector.com/index.php?title=Advanced_Preprocessing:_Noise,_Offset,_and_Baseline_Filtering#Derivative_.28SavGol.29
52. Bertrand D, Vigneau E. Prétraitement des données spectrales dans la spectroscopie infrarouge et ses applications analytique. In: Dufour E, editor. La spectroscopie infrarouge et ses applications analytiques. 2° edition. Tech & doc; 2006. p. 248.
53. BARNES RJ, DHANOA MS, SUSAN J. LISTER. Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra. APPLIED SPECTROSCOPY. 1989;43(5):772–7.
54. Roggo Y. Détermination de la qualité de la betterave sucrière par spectroscopie infrarouge et chimiométrie. Université des Sciences et Technologies de Lille, France; 2003.
55. Geladi P, Macdougall D. Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat. 1985;39(3).
56. Naes T, Isaksson T. Locally Weighted Regression and Scatter Correction for Near-

Infrared Reflectance Data. Analytical Chemistry. 1990;62(7):664–73.

57. Enzécri JPB. Histoire et préhistoire de l'analyse des données. In: Les cahiers de l'analyse des données. 1977. p. 9–40.
58. Jolliffe IT. Principal Component Analysis, Second Edition. Encyclopedia of Statistics in Behavioral Science [Internet]. 2002;30(3):487. Available from: http://onlinelibrary.wiley.com/doi/10.1002/0470013192.bsa501/full
59. Fisher RA. Studies in crop variation. Journal of Agricultural Science,. 1923;13:311–20.
60. Wold H. Nonlinear estimation by iterative least squares procedure. Research Papers in Statistic, Wiley, New York. 1966;411–44.
61. Hotelling H. Analysis of a compfex of statistical variables into principal components. Baltimore, editor. Warwick and York Inc; 1933. 417-441. p.
62. MANDEL J. Use of the Singular Value Decomposition in Regression Analysis. The American Statistician. 1982;36(1):15–24.
63. Kolmogoroff A. Uber die analytisehen Methoden in der Wahrseheinliehkeitsreehnung. Annales AcademiaeScientiarum Fennicae. 1947;137.
64. Pla LE. Análisis Multivariado : Método de Componentes Principales. ecretaríade la Organización de Estados Americanos (OEA), Washington, DC. 1986;
65. Pinto RC. Development of new chemometric methods - application to the spectroscopic characterisation of food quality. AgroParisTec; 2009.
66. Bro R, Age KS. Principal component analysis. Analytical Methods. 2014;6(2812):433–59.
67. Herrera A, Ballabio D, Navas N, Todeschini R, Cardell C. Principal Component Analysis to interpret changes in chromatic parameters on paint dosimeters exposed long-term to urban air. Chemometrics and Intelligent Laboratory Systems [Internet]. 2017;167(15):113–22. Available from: http://dx.doi.org/10.1016/j.chemolab.2017.05.007
68. Vanhatalo E, Kulahci M, Bergquist B. On the structure of dynamic principal component analysis used in statistical process monitoring. Chemometrics and

Intelligent Laboratory Systems [Internet]. 2017;167(May):1–11. Available from: http://dx.doi.org/10.1016/j.chemolab.2017.05.016

69. Ditta A, Nawaz H, Mahmood T, Majeed MI, Tahir M, Rashid N, et al. PT. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy [Internet]. 2019;117173. Available from: https://doi.org/10.1016/j.saa.2019.117173
70. Ballabio D, Consonni V. Classification tools in chemistry. Part 1: Linear models. PLS-DA. Analytical Methods. 2013;5:3790–3798.
71. Berrueta LA, Alonso-salces RM. Supervised pattern recognition in food analysis. Journal of Chromatography A. 2007;1158:196–214.
72. Heude C. Développement de nouvelles méthodes analytiques dans l ' agroalimentaire par RMN. Université de Strasbourg,; 2015.
73. Harald M, Tormod N. Multivariate calibration. In: Kowalski BR, editor. Mathematics and Statistics in Chemistry. 1st ed. D. Reidel Pu blishing Company; 1984. p. 147–56.
74. Wold H, Jöreskog KG. Soft modeling thebasic design and some extensions, inSystems Under Indirect Observation. vols. I an. North-Holland, Amsterdam; 1982.
75. Wold S, Eriksson L, Trygg J, Kettaneh N. The PLS method -- partial least squares projections to latent structures -- and its applications in industrial RDP. Umea, Sweden; 2004.
76. Wold H. Estimation of principal components and related models by iterative least squares. In: Multivariate analysis. NewYork: Academic Press; 1966. p. 391–420.
77. Wold H. No Title. In: n Encyclopaedia of statistical sciences (ed Kotz, S, Johnson, NL). Wiley, New York,; 1985. p. 581–91.
78. Wold S, Trygg J, Berglund A, Antti H. Some recent developments in PLS modeling. 2001;
79. Wold S, Sjostrom M. PLS-regression : a basic tool of chemometrics. 2001;109–30.
80. Tenenhaus M. Partial Least Squares. Revue de statistique appliquée. 1995;1(47):2–55.
81. Cuny M. Food authentication by Proton NMR spectroscopy in combination with

chemometric tools. AgroParisTech; 2008.

82. Burnham AJ, MacGregor JF, Viveros R. Latent variable multivariate regression modeling. Chemometrics and Intelligent Laboratory Systems. 1999;48(2):167–80.
83. Burnham AJ, Viveros R, Macgregor JF. Frameworks for latent variable multivariate regression. Journal of Chemometrics. 1996;10(1):31–45.
84. Nomikos P. Multivariate SPC Charts for Batch Monitoring Processes. Technometrics. 2015;37(1):41–59.
85. Brereton RG. Introduction to multivariate calibration in analytical chemistry. Analyst. 2000;125(11):2125–54.
86. Martens H, T. Næs. Near-Infrared Technologyin Agricultural and Food Industries. In: Near-Infrared Technologyin Agricultural and Food Industries. 1987. p. 57.
87. Hanafi M, Ouertani SS, Boccard J, Mazerolles G, Rudaz S. Multi-way PLS regression: Monotony convergence of tri-linear PLS2 and optimality of parameters. Computational Statistics and Data Analysis [Internet]. 2015;83:129–39. Available from: http://dx.doi.org/10.1016/j.csda.2014.10.003
88. Höskuldsson A. PLS regression. Journal of Chemometrics. 1988;2(August 1987):581–91.
89. Wakeling IN, Morris JJ. A test of significance for partial least squares regression. Journal of Chemometrics. 2005;7(4):291–304.
90. Clark M, Cramer RD. The Probability of Chance Correlation Using Partial Least Squares (PLS). Quantitative Structure-Activity Relationships. 1993;12(2):137–45.
91. Shao J. Linear Model Selection by Cross-Validation. American Statistical Association [Internet]. 1993;88(422):486–94. Available from: http://www.jstor.org/stable/2290328
92. Mi M, Aye S, Zhang N. A novel methodology in transforming bulk properties of refining streams into molecular information. Chemical Engineering Science. 2005;60:6702–17.
93. Filipe L, Brandão P, Willian J, Braga B, Anselmo P, Suarez Z. Determination of vegetable oils and fats adulterants in diesel oil by high performance liquid chromatography and multivariate methods. Journal of Chromatography A.

2012;1225:150–7.

94. Sandercock PML, Pasquier E Du. Chemical fingerprinting of unevaporated automotive gasoline samples. Forensic Science International. 2007;134(2003):1–10.

95. Barbeira PJS, Pereira RCC, Corgozinho CNC, Horizonte B, Gerais M, August R V, et al. Identification of Gasoline Origin by Physical and Chemical Properties and Multivariate Analysis. Energy Fuels. 2007;18(2):2212–5.

96. Borin A, Poppi RJ. Application of mid infrared spectroscopy and iPLS for the quantification of contaminants in lubricating oil. 2005;37:27–32.

97. Cesar J, Alves L, Henriques CB, Poppi RJ. Determination of diesel quality parameters using support vector regression and near infrared spectroscopy for an in-line blending optimizer system. Fuel. 2012;97:710–7.

98. Laxalde J, Ruckebusch C, Devos O, Caillol N. Analytica Chimica Acta Characterisation of heavy oils using near-infrared spectroscopy : Optimisation of pre-processing methods and variable selection. 2011;705:227–34.

99. Laxalde J. Analyse des produits lourds du pétrole par spectroscopie vibrationnelle. Université des Sciences et Technologie de Lille - Lille I; 2012.

100. Aleme HG, Barbeira PJS. Determination of flash point and cetane index in diesel using distillation curves and multivariate calibration. Fuel [Internet]. 2012;102:129–34. Available from: http://dx.doi.org/10.1016/j.fuel.2012.06.015

101. Ale BB. Fuel Adulteration and Tailpipe Emissions. Journal of the Institute of Engineering. 2003;3(1):12–6.

102. Pasadakis N, Kardamakis AA. Identifying constituents in commercial gasoline using Fourier transform-infrared spectroscopy and independent component analysis. Analytica Chimica Acta. 2006;578(2):250–5.

103. Fodor GE, Mason RA, Hutzler SA. Estimation of Middle Distillate Fuel Properties by FT-IR. APPLIED SPECTROSCOPY. 1999;53(10).

104. Bai L, Smuts J, Schenk J, Cochran J, Schug KA. Comparison of GC-VUV , GC-FID , and comprehensive two-dimensional GC – MS for the characterization of weathered and unweathered diesel fuels. Fuel [Internet]. 2018;214(August

2017):521–7. Available from: https://doi.org/10.1016/j.fuel.2017.11.053

105. Tian J, Tan J, Hu N, Liu T, Wang Y, Zhong H, et al. Characteristics analysis for total volatile organic compounds emissions of methanol-diesel fuel. Journal of the Energy Institute [Internet]. 2017; Available from: http://dx.doi.org/10.1016/j.joei.2017.04.004
106. Taylor P, Sink CW, Hardy DR, Huang MA. Fuel Science and Technology International AN ACCURATE HYDROCARBON TYPE ANALYSIS OF ALL FUEL TYPES. (January 2015):37–41.
107. Vozka P, Kilaz G. How to obtain a detailed chemical composition for middle distillates via GC × GC-FID without the need of GC × GC-TOF / MS. Fuel [Internet]. 2019;247(January):368–77. Available from: https://doi.org/10.1016/j.fuel.2019.03.009
108. Miranda AM, Castilho-almeida EW, Martins EH, Moreira GF, Achete CA, Armond RASZ, et al. Line shape analysis of the Raman spectra from pure and mixed biofuels esters compounds. Fuel [Internet]. 2014;115(0016):118–25. Available from: http://dx.doi.org/10.1016/j.fuel.2013.06.038
109. Eigenvector. wiki eigenvector [Internet]. data Preprocessing. 2018. p. normalization, GLSW, mean centering, COW. Available from: http://wiki.eigenvector.com/
110. Berg F Van Den, Andersson C. Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. Journal of Chemometrics. 2004;(July):231–41.
111. Barra I, Alaoui M, Cherrah Y, Kharbach M, Bouklouze A. FTIR fingerprints associated to a PLS-DA model for rapid detection of smuggled non-compliant diesel marketed in Morocco. Vibrational Spectroscopy. 2019;101(January):40–5.
112. Westerhuis JA, Hoefsloot HCJ, Smit S, Vis DJ, Smilde AK, Velzen EJJ, et al. Assessment of PLSDA cross validation. Metabolomics. 2008;4(1):81–9.
113. Mazivila S, De Santana FB, Mitsutake H, Gontijo LC, Santos DQ, Neto WB. Discrimination of the type of biodiesel/diesel blend (B5) using mid-infrared spectroscopy and PLS-DA. Fuel. 2015;142:222–6.

114. Brouillette C, Smith W, Shende C, Gladding Z, Farquharson S, Morris RE, et al. Analysis of Twenty-Two Performance Properties of Diesel, Gasoline, and Jet Fuels Using a Field-Portable Near-Infrared (NIR) Analyzer. Applied Spectroscopy. 2016;70(5):746–55.
115. Robert E. Morris, Hammond MH, Cramer JA, Johnson KJ, Giordano BC, Kramer KE, et al. Rapid Fuel Quality Surveillance through Chemometric Modelling of NIR Spectra. Energy & Fuels. 2009;4(8):1610–8.
116. Ferrão MF, Viera MDS, Pazos REP, Fachini D, Gerbase AE, Marder L. Simultaneous determination of quality parameters of biodiesel/diesel blends using HATR-FTIR spectra and PLS, iPLS or siPLS regressions. Fuel. 2011;90(2):701–6.
117. Esbensen KH, Guyot D, Houmøller LP. An Introduction to Multivariate Data Analysis and Experimental Design. 5th editio. Camo, editor. Multivariate Data Analysis. Oslo, Norway: CAMO Process AS; 2004. 105-124 p.
118. Jean-Claude GUIBET. Indice de Cétane [Internet]. Encyclopædia Universalis. Available from: http://www.universalis.fr/encyclopedie/indice-de-cetane/
119. Jian-guang Y, Wu-gao Z, Zhen H. Effect of cetane number improver on heat release rate and emissions of high speed diesel engine fueled with ethanol – diesel blend fuel. 2020;83(2004):2013–20.
120. Min K, Valco DJ, Oldani A, Kim K, Temme J, Kweon CM, et al. Autoignition of varied cetane number fuels at low temperatures. Proceedings of the Combustion Institute [Internet]. 2018;000:1–9. Available from: https://doi.org/10.1016/j.proci.2018.05.078
121. García-martín JF, Alés-álvarez FJ, López-barrera MC, Martín-domínguez I, Álvarez-mateos P. Cetane number prediction of waste cooking oil-derived biodiesel prior to transesteri fi cation reaction using near infrared spectroscopy. Fuel [Internet]. 2019;240(November 2018):10–5. Available from: https://doi.org/10.1016/j.fuel.2018.11.142
122. Soyemi OO, Busch MA, Busch KW. Multivariate Analysis of Near-Infrared Spectra Using the G-Programming Language. 2000;1093–100.
123. Santos VO, Oliveira FCC, Lima DG, Petry AC, Garcia E, Suarez PAZ, et al. A

comparative study of diesel analysis by FTIR, FTNIR and FT-Raman spectroscopy using PLS and artificial neural network analysis. Analytica Chimica Acta. 2005;547(2):188–96.

Annex 1. Published Articles