

الجمهورية الجزائرية الديمقراطية الشعبية  
République Algérienne Démocratique et Populaire  
وزارة التعليم العالي و البحث العلمي  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique  
جامعة قاصدي مرباح ورقلة  
UNIVERSITE Kasdi Merbah - Ouargla  
Faculté des Sciences Appliquées  
Département de Génie des Procédés



**Mémoire présenté en vue de l'obtention de Master Académique**

Domaine : Sciences et Technologies  
Filière : Industrie Pétrochimique  
Spécialité : Génie de Raffinage

Présenté par : **Kaouter MIMMI**

## SUJET

# Contribution à la Prédiction des Indices de Cétane et d'Octane des composés et mélanges de carburant basée sur les modèle de Machine Learning

Soutenu publiquement, le 26 Juin 202, devant le jury :

M.	Djamal ZERROUKI	Professeur	Université de Ouargla	Président
M.	Kamel BENCHEIKH	Maître Assistant - A	Université de Ouargla	Examineur
M.	Mourad KORICHI	Professeur	Université de Ouargla	Encadreur

Année universitaire 2020/2021

# Remerciements

Je Veut Remercier Tout D'abord Notre Dieu  
Qui Ma A Donné La Force Et La Puissance Pour  
Terminer Ce Modeste Travail.

Tous Mes Remerciements Très Sincèrement A Mon  
Encadreur Mr. KORICHI.Mourad

Pour Son Grand Soutien Et Ses Conseils Considérables.

Je Veut Remercier Egalement Les Membres De Jury  
D'avoir Juger Ce Travail

Mr.Djamal ZERROUKI et Mr.Kamel BENCHEIKH

Je N'oublie Pas De Remercier Nos Enseignants Du  
Département Génie De Procède

Pour Les Efforts Qu'ils Sont Fournis Durant Notre  
Cursus Afin De Nous Amener Jusqu'au Bout De La  
Formation.

Je Profitant De L'occasion Pour Remercier Tous Ceux  
Qui Ont Collaboré De Près Ou De Loin  
À La Réalisation De Ce Mémoire.

MIMMI Kaouter





# *D*édicaces

*En Premier Lieu, Au Dieu Tout  
Puissant De M'avoir Accordé La Puissance  
Et La Volonté Pour Terminer Ce Travail.  
A Mes Parents Ma Les Plus Chers Que Ma Vie,  
ma Maman BRAHIMI Khiera Et Mon  
Père Qui Décédé Ahmida, Qui Ont Su  
M'apporter Amour Et Sentiment Tendre Et  
Qui Ont Fait De Moi Ce Que Je Suis Aujourd'hui Et  
Je Leur Serai Toujours Redevable.  
A Mes Frères, Mustapha, Ismaïl, Abd El Karim, Sliman  
Et Sans Oublier Mon Frère Qui Décédé Yahia.  
A Mes Sœurs, Zohra, Khadidja, Yamina, Asma Et  
Meriem.  
A Tous Ce Qui Me Sont Chers.*

## **Les abbreviation:**

**OI** : Indice d'Octane

**CN** : Indice de Cétane

**RON** : indice d'Octane de Recherche

**MON** : indice d'Octane de Moteur

**MTBE**: Methyl Tetra-Butyl Ether

**ETBE**: Ether Ethyl Tetra Butyl

**IQT**: Ignition Quality Tester

**HC** : Hydrocarbure

**HCO** : Oxyhydrocarbures

**MW**: Molecular Weight

## Listes des figures

### *Chapitre 1 : Etude bibliographique*

Figure I- 1 Filières des biocarburants de première génération [08] .....	2
Figure I- 2 Réaction de transestérification entre un triglycéride et du méthanol [09].....	2
Figure I- 3 Processus de fabrication empruntée par cette voie thermo-chimique[07] .....	4
Figure I- 4 Processus de fabrication des biocarburants de 3 <sup>ème</sup> génération [© Eric Maréchal] .....	5
Figure I- 5 Fonctionnement des moteurs de diesel .....	7

## Liste des graphiques :

### *Chapitre 3*

Le graphique III- 1 de la distance entre les prédictions et les observations de CN mesuré (sans MW) .....	21
Le graphique III- 2 de la distance entre les prédictions et les observations de CN mesuré (avec MW) .....	22
Le graphique III- 3 de la distance entre les prédictions et les observations de CN mesuré (sans MW) .....	26
Le graphique III- 4 de la distance entre les prédictions et les observations de CN mesuré (avec MW) .....	27
Le graphique III- 5 de la distance entre les prédictions et les observations de CN mesuré (sans MW). .....	31
Le graphique III- 6 de la distance entre les prédictions et les observations de CN mesuré (avec MW) .....	32
Le graphique III- 7 d'équivalence pour les valeurs mesurées et prédictives de CN par un modèle de régression d'auto-apprentissage pour l'ensemble de données de composés purs. ....	34
Le graphique III- 9 Valeurs prédites par rapport aux valeurs mesurées de CN pour les hydrocarbures oxygénés. ....	34
Le graphique III- 8 Valeurs prédites par rapport aux valeurs mesurées de CN pour les hydrocarbures. ....	34

## Liste des tableaux :

### *Chapitre2*

<b>Tableau II- 1</b> Groupes fonctionnelles de contribution.....	14
<b>Tableau II- 2</b> Données relatives à l'indice de cétones utilisés dans ce travail .....	15
<b>Tableau II- 3</b> Hypothèses utilisés dans la modélisation.....	18

### *Chapitre3*

<b>Tableau III- 1</b> Matrice de corrélation de base de données sans masse moléculaire .....	19
<b>Tableau III- 2</b> Matrice de corrélation de base de données avec masse moléculaire.....	19
<b>Tableau III- 3</b> Statistiques relatives à l'ajustement du modèle de régression.....	20
<b>Tableau III- 4</b> Matrice de corrélation de base de données sans masse moléculaire .....	24
<b>Tableau III- 5</b> Matrice de corrélation de base de données avec masse moléculaire.....	25
<b>Tableau III- 6</b> Statistiques relatives à l'ajustement du modèle de régression.....	26
<b>Tableau III- 7</b> Matrice de corrélation de base de données sans masse moléculaire .....	29
<b>Tableau III- 8</b> Matrice de corrélation de base de données avec masse moléculaire.....	30
<b>Tableau III- 9</b> Statistiques relatives à l'ajustement du modèle de régression.....	31
<b>Tableau III- 11</b> Statistiques relatives à l'ajustement du modèle de régression.....	33
<b>Tableau III- 12</b> la différence de $R^2$ entre les études .....	34

## Liste des références :

### Références bibliographique :

- [01] Casimir Togbe. Etude cinétique de l'oxydation de constituants de biocarburants et composés modèles : formation de polluants. Alimentation et Nutrition. Université d'Orléans, 2010. Français.
- [02] <http://biocarburants-parlons-chimie.e-monsite.com/pages/les-biocarburants/seconde-generation.html>
- [03] <https://www.techniques-ingenieur.fr/actualite/articles/les-biocarburants-de-troisieme-generation-6831/>
- [04] THESE du Doctorat: Diagnostic à base de modèle : application à un moteur diesel suralimenté à injection directe (spécialité automatique et traitement du signal par Zahi SABEH)
- [05] cours\_de\_carburant\_dr\_monsieur\_bouchlaleg\_lazhar\_Université de Batna 2 / Ecole Supérieure de Technologie pdf
- [06] Using machine learning with target-specific feature sets for structureproperty relationship modeling of octane numbers and octane sensitivity , Florian vom Lehn\*, Benedict Brosius, Rafal Broda, Liming Cai, Heinz Pitsch, *Fuel* 281 (2020) 118772
- [07] <http://www.rtflash.fr/biocarburants-deuxieme-generation-voie-lignocellulosique/article>
- [08] [IFP Énergies nouvelles](#)
- [9] Machine learning regression based group contribution method for cetane and octane numbers prediction of pure fuel compounds and mixtures Corresponding author. *E-mail address:* a.tsolakis@bham.ac.uk (A. Tsolakis). <https://doi.org/10.1016/j.fuel.2020.118589> Received 8 April 2020; Received in revised form 24 May 2020; Accepted 29 June 2020
- [10] A Group Contribution Method for Estimating Cetane and Octane Numbers William L. Kubic, Jr. July 2016
- [11] Compendium of Experimental Cetane Number Data September 2004



# Sommaire

Remerciement

Dédicaces

Abstract

Les abréviations

Lise de figures

Liste de graphiques

Liste des tableaux

Liste des références

*Introduction générale* ..... 1

## *Chapitre 1*

I) Généralité sur les Carburant : ..... 1

La biomasse : ..... 1

Les biocarburants : ..... 1

– Les biocarburants de 1<sup>ère</sup> génération : ..... 2

○ *Biodiesel* : ..... 2

○ *Bio alcool* : ..... 3

– Biocarburants de 2<sup>ème</sup> génération : ..... 3

○ *Voie Biochimique* : ..... 3

○ *Voie Thermochimique* : ..... 4

– Biocarburants de 3<sup>ème</sup> génération : ..... 5

II) Indice de cétane : ..... 5

Moteur diesel: ..... 6

III) Indice d'octane : ..... 7

Types des indices d'octane : ..... 8

– *Indice d'octane RON (indice recherche)* : ..... 8

– *Indice d'octane MON (indice moteur)* : ..... 8

IV) Méthodes expérimentales et normes de détermination des indices de cétane et d'octane : ..... 8

–	<b>Indice de cétane :</b> .....	8
–	<b>Indice d'octane :</b> .....	9
V)	<b>Conclusion:</b> .....	10

## *Chapitre 2*

I.	<b>Introduction :</b> .....	13
II.	<b>Description de la méthode contribution des groupes et modélisation :</b> .....	13
III.	<b>Base de données (base de calcul indice de cétane) :</b> .....	14
	□ <b>Hydrocarbures :</b> .....	14
	□ <b>Oxyhydrocarbures :</b> .....	15
IV.	<b>Description de la technique et la méthode utilisées pour l'indice de Cétane :</b> .....	15
	□ <b>Régression linéaire :</b> .....	15
V.	<b>Description de l'outil logiciel –XLSTAT :</b> .....	17
VI.	<b>Hypothèses de la modélisation :</b> .....	17

## *Chapitre 3*

<b>I.</b>	<b>Introduction :</b>	<b>19</b>
II.	<b>Présentation des résultats :</b> .....	19
II.1.	<b>Régressions linéaire de HC sans et avec masse moléculaire :</b> .....	19
	<b>Coefficients d'ajustement (Measured_CN) :</b> .....	20
	<b>R<sup>2</sup> :</b> .....	20
	<b>R<sup>2</sup> ajusté :</b> .....	20
	<b>Equation du modèle CN (sans masse moléculaire) :</b> .....	21
	<b>1. Equation du modèle CN (avec masse moléculaire) :</b> .....	21
	<b>Interprétation sans MW (Measured_CN) :</b> .....	22
	<b>Interprétation (Measured_CN) avec MW :</b> .....	22
II.2.	<b>Régressions linéaire de HCO sans et avec masse moléculaire :</b> .....	23
	<b>Coefficients d'ajustement (Measured_CN) :</b> .....	26
	<b>Equation du modèle CN (sans masse moléculaire) :</b> .....	26
	<b>Equation du modèle CN (avec masse moléculaire) :</b> .....	26

<b>Discussion :</b> .....	27
<b>Interprétation sans MW (Measured_CN) :</b> .....	27
<b>Interprétation avec MW (Measured_CN) :</b> .....	27
II.3. Régressions linéaire de tous les composés sans et avec masse moléculaire :.....	28
<b>Coefficients d'ajustement (Measured_CN) :</b> .....	31
<b>Equation du modèle CN (sans masse moléculaire) :</b> .....	31
<b>Equation du modèle CN (avec masse moléculaire) :</b> .....	31
<b>Discussion :</b> .....	32
<b>Interprétation sans MW (Measured_CN) :</b> .....	32
<b>Interprétation avec MW (Measured_CN) :</b> .....	32
<b>La comparaison avec des autres études :</b> .....	33
<b>Article [9] :</b> .....	33
<b>Article [10] :</b> .....	34

# Introduction générale

## Introduction :

Le raffinage du pétrole est un processus nécessaire au traitement du pétrole brut, Pour en extraire la substance désirée ; La matière extraite du pétrole brut est transformée en produits propres à la consommation. En tant que carburant, il s'agit d'un mélange extrait par le processus de raffinage du pétrole, et ce dernier nous fournit l'énergie thermique que nous utilisons dans de nombreux domaines de la vie, comme la production d'électricité et la cuisson, ainsi que le fonctionnement de différents moteurs tels que les moteurs de voiture, les avions et les trains, et les types de carburant sont divisés en deux types:

- Carburant naturels, tels que le charbon, le pétrole et le gaz naturel sont des gisements dans le sol qui se sont formés au cours de millions d'années à partir des restes d'animaux et de plantes, c'est pourquoi on l'appelle combustible fossile, qui est une énergie non renouvelable.
- Carburant industriel, qui est extrait de déchets végétaux et animaux, où nous pouvons utiliser ces déchets pour produire de l'énergie, après avoir subi des processus de fermentation. Ce combustible a été appelé biocarburant et est classé comme énergie renouvelable.

Pour utiliser du carburant dans des moteurs à essence ou diesel, nous devons connaître les propriétés physiques et chimiques de chaque carburant. Parmi ces caractéristiques ; le nombre de cétane (qui nous sommes concentrés sur l'estimation dans notre travail) d'un moteur diesel, qui est généralement estimé en calculant son indice dans le gazole, le nombre d'octane pour les moteurs à essence, qui est estimé en calculant le rapport dans le mélange d'essence.

De ce point de vue, nous mentionnons l'importance de l'indice de cétane et de l'indice d'octane, qui est généralement de maintenir la continuité du moteur ainsi que d'améliorer la qualité du carburant (essence et diesel).

Dans ce contexte, notre étude vise à trouver une méthode de calculer le nombre de cétane utilisé dans les moteurs diesel depuis le Machine Learning, ce qui permet d'estimer le nombre de cétane à partir de l'étude de ses propriétés chimiques et de ses groupes de contribution.

Afin d'accomplir ce travail, nous allons expliquer certaines des étapes adoptées selon la séquence suivante :

- Tout d'abord, dans le premier chapitre, nous aborderons certains concepts tels que carburant et biocarburant, qu'est-ce que le cétane et l'octane, quels sont les mécanismes et méthodes utilisés pour estimer les indicateurs d'octane et de cétane, et à la fin du chapitre, une petite conclusion à guider le travail.
- Deuxièmement, nous discuterons dans le deuxième chapitre de la méthode de modélisation utilisant le programme XLSTAT pour prédire l'indice de cétane, et nous

présenterons brièvement le contenu de la base de données et les groupes utilisés dans la modélisation, et au en fin de chapitre, nous présenterons les hypothèses de modélisation que nous avons adoptées dans ce travail.

- Et dans le dernier chapitre, nous présenterons certains des résultats obtenus et nous comparerons les résultats obtenus à partir de la modélisation avec les résultats d'autres travaux qui ont été étudiés auparavant.



## **I) Généralité sur les Carburant :**

Le carburant est le premier approvisionnement énergétique au monde. Combustibles naturels comme le charbon Le pétrole et le gaz naturel sont des gisements dans le sol qui se sont formés pendant des millions d'années à partir de restes d'animaux et de plantes, c'est ce qu'on appelle les combustibles fossiles.

Le carburant est un matériau inflammable qui interagit avec l'oxygène .La combustion produit de l'énergie thermique et se transforme en énergie mécanique pour faire fonctionner des moteurs thermiques à pistons (moteurs à essence et diesel) ou des moteurs à flux continu (avions et turbines à gaz).

L'essence et le diesel sont deux types de carburant (carburant conventionnel) classés comme les carburants les plus utilisés, et tous deux sont extraits par des procédés de raffinage du pétrole .

Aujourd'hui, la demande croissante d'énergie a épuisé les réserves de pétrole, pollué l'atmosphère et causé de graves dommages à l'atmosphère. Ces carburants ont été critiqués et responsables des émissions de gaz à l'origine du réchauffement climatique. Ces phénomènes ont conduit à la recherche d'autres sources d'énergie, et l'une des solutions a été de développer des carburants à base de biomasse, car les biocarburants raffinés issus de la biomasse sont respectueux de l'environnement et sont également considérés comme des énergies renouvelables.

### **La biomasse :**

Est connue comme une source d'énergie majeure, allant de la combustion du bois pour le chauffage au raffinage des matières végétales et animales en combustible. En raison de la forte demande d'énergie dans tout l'univers et de l'épuisement des réserves de pétrole ainsi que des dommages environnementaux causés par les combustibles fossiles, la production de combustibles a eu tendance à dépendre de la biomasse.

### **Les biocarburants :**

C'est le carburant qui est extrait des matériaux vitaux et est l'une des sources d'énergie renouvelable les plus importantes, contrairement aux ressources naturelles telles que le pétrole

et les combustibles fossiles. On distingue deux principales filières de production de biocarburants : les filières de la 1<sup>ère</sup> et la 2<sup>ème</sup> génération :

– **Les biocarburants de 1<sup>ère</sup> génération :**

Ce biocarburant est tiré des betteraves, des céréales et de la canne à sucre pour l'éthanol, mais le biodiesel est dérivé du colza, du tournesol, du palmier et du soya. Ils sont principalement de deux filières :

○ **Biodiesel :**

Le biodiesel est un carburant qui peut être utilisé pur et/ou mélangé dans le gazole pour les moteurs Diesel. Il est constitué d'esters méthyliques d'huiles végétales (EMHV), eux-mêmes 17 provenant d'huiles végétales (tournesol, colza, palme, soja, etc.). En effet, les esters sont des hydrocarbures obtenus par réaction de transestérification entre les triglycérides et les alcools. Dans notre cas, l'alcool utilisé est le méthanol, d'où l'appellation « esters méthyliques ». Les esters méthyliques ont ainsi pour formule R-COOCH<sub>3</sub>, où R est un groupement alkyle. La figure présente une réaction de transestérification entre un triglycéride et le méthanol.[01]

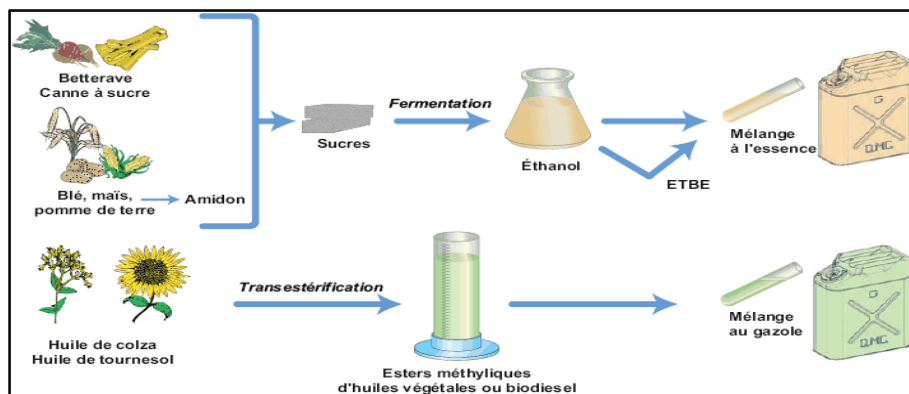


Figure I- 1 Filières des biocarburants de première génération [08]

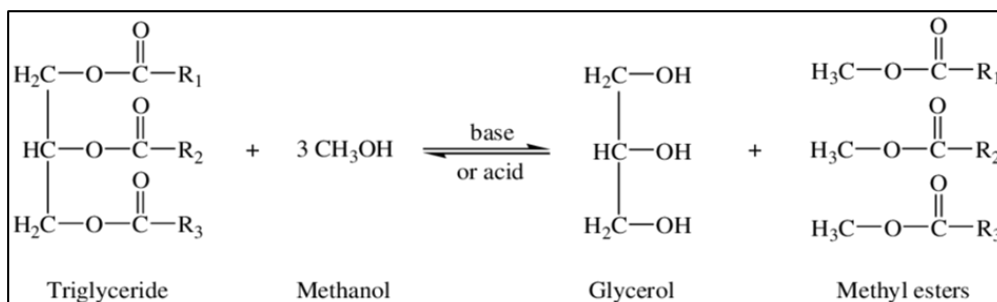


Figure I- 2 Réaction de transestérification entre un triglycéride et du méthanol [09]

Ils sont actuellement développés et utilisés comme carburant alternatif. Ils sont incorporés dans la composition des gazoles jusqu'à 5% en volume car ils présentent des caractéristiques à la combustion similaires à celles des gazoles (viscosité, indice de cétane, etc.). L'indice de cétane est l'un des principaux indicateurs de la qualité du gazole et caractérise le délai d'allumage du carburant au moment de son injection dans la chambre de combustion.[01]

- **Bio alcool :**

Le bio-alcool (y compris le bio-éthanol) est principalement produit à partir de la fermentation de sucreries (canne à sucre, betterave à sucre) et de céréales (blé, maïs et orge). Le bioéthanol brut devient un composant de l'essence à différents pourcentages selon les carburants. Cet alcool, formulé par ROH, a beaucoup de propriétés intéressantes, y compris des indicateurs d'indice d'octane élevé.

Il aussi permet de diminuer les émissions de gaz à effet de serre, par contre aux carburants fossiles .mieux que, les matières végétales(ou agricole) utilisées pour sa fabrication absorbent le CO<sub>2</sub> en quantité importante.

Les perspectives de développement de ces biocarburants de première génération sont réelles, mais leur production soulève parfois de sérieuses interrogations dans la société par la concurrence qu'ils créent avec les besoins alimentaires des populations. De plus, les quantités produites de nos jours restent encore limitées et ne peuvent se substituer qu'à une faible proportion du carburant d'origine fossile.

Les voies de production dites de seconde génération promettent une augmentation de la production de ces carburants, sans concurrence avec les utilisations alimentaires, qui demeureront toujours prioritaires.

- **Biocarburants de 2<sup>ème</sup> génération :**

Ils sont obtenus à partir de la biomasse lignocellulosique comme le bois ou la paille. Il existe deux types de voies de transformation. Une pour le domaine de la production d'éthanol correspondant à la voie biochimique de la cellulose et l'autre pour le secteur du gazole et du kérosène de synthèse correspondant à la voie thermochimique appelée BtL (Biomass to Liquid).

- **Voie Biochimique :**

Permettant la production d'éthanol (filière essence). La biomasse lignocellulosique est constituée de trois polymères : la cellulose, les hémicelluloses et la lignine. D'abord, il faut séparer la lignine des deux autres polymères par hydroliquéfaction par pyrolyse. Ensuite, la cellulose libérée va être transformée en glucose via une hydrolyse enzymatique. Le glucose produit est alors fermentable en éthanol (dit de seconde génération). Puis comme pour la première génération, l'éthanol est purifié par distillation puis sera ajouté dans des proportions différentes à l'essence.

Notons que les hémicelluloses et la lignine ne sont pas fermentescibles mais les hémicelluloses peuvent être valorisées et transformées en gaz de synthèse (voir "la filière thermochimique").

### o Voie Thermochimique :

Permettant la production de biogazole (biodiesel). La première partie du procédé consiste à gazéifier l'hémicellulose, ce qui permet de casser ces molécules en un gaz de synthèse, standard l'utilisation de procédés à hautes températures. Le gaz de synthèse ou syngas (mélange de CO et H<sub>2</sub>) peut, en présence de catalyseur tels que le fer ou le cobalt, se transformer en eau et en hydrocarbures liquides (molécules similaires à celles produites avec le pétrole) : c'est la réaction de synthèse Fischer-Tropsch (deux chimistes allemands).[02]

- Avec un catalyseur fer :  $n (2 H_2 + CO) \rightarrow C_n H_{2n} + n H_2O$
- Avec un catalyseur cobalt :  $(2n + 1) H_2 + n CO \rightarrow C_n H_{2n+2} + n H_2O$

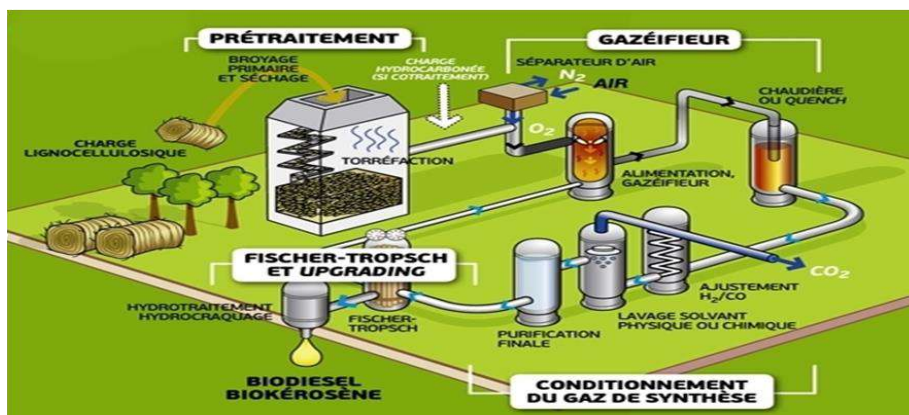


Figure I- 3 Processus de fabrication empruntée par cette voie thermochimique[07]

– **Biocarburants de 3<sup>ème</sup> génération :**

Les biocarburants de troisième génération sont produits par des microalgues. Il existe deux méthodes de cultivation. D'une part avec des procédés dit « intensifs » où les microalgues croissent dans des photobioréacteurs fermés où l'on contrôle les paramètres physico-chimiques et d'autre part, les procédés dit « extensifs » dans des bassins en extérieur. Le milieu de culture circule grâce à des roues à aubes. Les éléments nutritifs sont apportés de manière à garantir, dans les conditions standards, une croissance optimale des algues.[03]

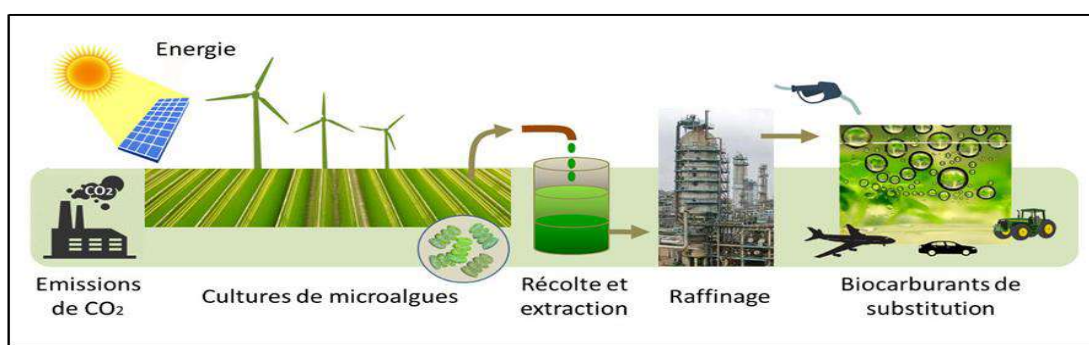
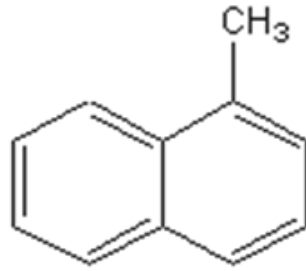


Figure I- 4 Processus de fabrication des biocarburants de 3<sup>ème</sup> génération [© Eric Maréchal]

**II) Indice de cétane :**

Le cétane ou hexadécane, est un hydrocarbure saturé de la famille des alcanes. Il contient 16 atomes de carbone (C) et 34 atomes d'hydrogène et son Formule chimique est C<sub>16</sub>H<sub>34</sub>. Il est une de la composition du diesel.

**Définition :** L'indice de cétane CN concerne les moteurs diesel. Il exprime la capacité du carburant diesel à s'enflammer et à brûler dans un moteur diesel (il favorise l'auto-inflammation dans la composition chimique du diesel). L'indice de cétane n'a aucun effet direct sur les performances du moteur. Plus l'indice de cétane est élevé, plus la vitesse d'allumage du carburant est élevée et plus le temps d'auto-inflammation du produit est court. L'échelle zéro de cet indice est donnée par la valeur du a-methylnaphthalene qui a une résistance élevée à l'inflammation et la valeur de 100 est donnée à l'hexadécane (C<sub>16</sub>H<sub>34</sub>) qui est encore appelé cétane.



Le rôle principal de l'indice de cétane diffère du rôle essentiel de l'octane dans l'amélioration des performances d'un moteur à carburant, de sorte que l'indice de cétane n'a pas d'effet direct sur les performances du moteur. Néanmoins, un indice de cétane plus élevé améliore plusieurs caractéristiques d'utilisation:

- Démarrage facile.
- Le bruit est moins intense surtout au ralenti froid.
- Réduire les émissions de fumées noires.

De plus, il est possible de réduire les émissions de polluants dans les gaz d'échappement grâce à une augmentation de l'indice de cétane. Si l'indicateur est trop faible, il augmentera le niveau de bruit et la difficulté d'un démarrage à froid, mais s'il est trop élevé (au-dessus de 60), il réduit le bruit mais les performances de la dynamique thermique seront très faibles. Ainsi, ne pas augmenter l'indice de cétane au-delà de 50 et 55 n'augmentera pas le coût de fonctionnement.

### **Moteur diesel:**

Le moteur diesel a été conçu par Rudolf Diesel à la fin du XIXe et au début du XXe siècle et est considéré comme un prototype pour le moteur qui porte son nom à ce jour. Il s'agissait d'un moteur à quatre temps qui injectait son carburant via un système d'injection directe à haute pression dans la chambre de combustion. Il se caractérise comme un moteur à combustion interne dont l'allumage n'est pas contrôlé automatiquement par le phénomène d'auto-allumage.

Ce type de moteur, qui a un taux de compression élevé, a connu une expansion rapide dans le domaine automobile depuis sa création.



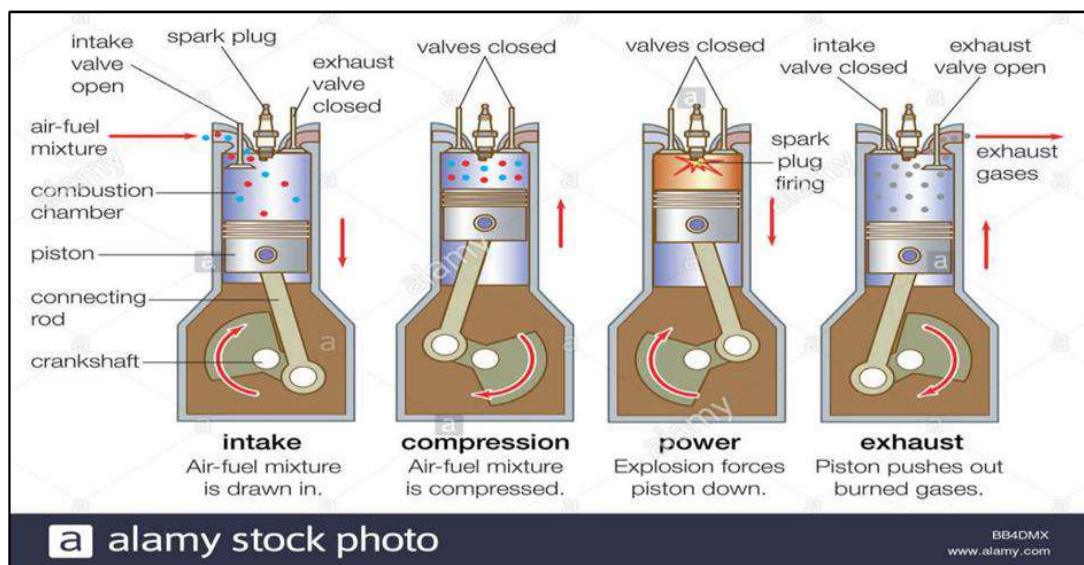


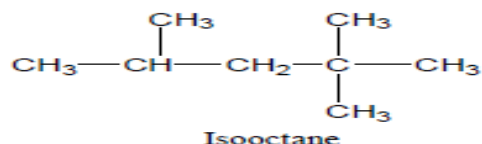
Figure I- 5 Fonctionnement des moteurs de diesel

Lorsque l'auto-inflammation se produit dans le moteur, elle provoque une explosion, et ce dernier provoque un phénomène appelé bruit ou clic. Cette explosion incontrôlée expose le moteur à des pressions mécaniques et thermiques nocives et dangereuses (blocage, émissions polluantes, fissures...) et la durabilité du moteur lui-même est menacée. Par conséquent, le moteur devait être équipé d'essence avec l'indice d'octane approprié spécifié par le constructeur du moteur.

### III) Indice d'octane :

L'octane est un composé hydrocarboné saturé de la famille des alcanes, de formule chimique  $C_8H_{18}$ . Il a 18 isomères et est considéré comme un composant essentiel des composants du carburant (essence).

**Définition :** l'indice d'octane (OI) est un nombre qui caractérise le délai d'inflammation du carburant (essence et GPL-C). Plus OI est élevé, plus le délai d'inflammation est élevé le carburant résiste bien à la détonation. <sup>5</sup>L'indice d'octane varie de zéro à 100, en comparant deux composés: n-heptane, qui reçoit un indice d'octane zéro car c'est le carburant le plus brûlant (favorise l'auto-inflammation du mélange), et 2, 2, 4-triméthylpentane (l'isooctane), qui retarde l'auto-inflammation du mélange et maintient le moteur (brûle doucement), reçoit un indice d'octane de 100.



### Types des indices d'octane :

– **Indice d'octane RON (indice recherche) :**

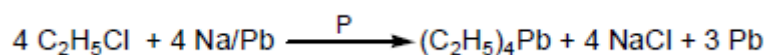
Qui est déterminé dans des conditions d'essais très éloignées de la réalité de fonctionnement des moteurs thermiques. [05]

– **Indice d'octane MON (indice moteur) :**

Qui est plus sévère et plus juste car il est déterminé dans des conditions plus proche de la réalité de fonctionnement des moteurs. L'indice moteur est inférieur environ de 10 points à l'indice recherche. [05]

Exemple : si RON = 95 , MON = 85 c'est le cas de super plomb.

L'essence quittant les raffineries a des indices modestes, il est donc nécessaire d'augmenter son indice d'octane. Pour cela, les additifs sont ajoutée une substance Tetra-éthyl plomb Pb (C<sub>2</sub>H<sub>5</sub>)<sub>4</sub>.



Mais il a été arrêté en 1970, et est maintenant interdit parce qu'il se propage et pollue l'atmosphère. Il a été remplacé par des hydrocarbures aliphatiques ramifiés ou leurs dérivés tels que le (méthyl tetra-butyl éther MTBE) ou (éther éthyle tetra-butyle ETBE).

### IV) Méthodes expérimentales et normes de détermination des indices de cétane et d'octane :

– **Indice de cétane :**

Dans le laboratoire d'ASG Analytik-Service GmbH à Neusaess, Allemagne. Le carburant a été testé à l'aide du test IQT par ASTM D6890, en ajustant d'abord la température de charge dans la chambre de combustion, et c'est l'une des normes ASTM, puis le point de consigne de l'unité de contrôle de la température est déterminé à  $818 \pm 30 \text{ K}$ , de sorte que Le n-heptane et le méthyl-cyclohexane sont deux substances de référence qui ont été déterminées auparavant

et ont un effet retardateur d'inflammation et la pression de charge est fixée à  $2,137 \pm 0,007$  MPa.

Le carburant liquide est injecté dans la chambre de combustion. Le délai d'allumage est défini comme le temps entre le début de l'injection de carburant (c'est-à-dire la montée de l'aiguille) et le début de la libération de chaleur indiquant de manière significative une augmentation significative de la pression. La valeur moyenne supérieure à 32 cycles de combustion est considérée comme le retard d'allumage final. Le nombre de cétane dérivé selon la norme ASTM D6890 peut être calculé comme suit :

- $DCN = 4.460 + 186.6/\tau$
- Si  $3.1 \text{ ms} \leq \tau \leq 6.5 \text{ ms}$ .
- Le délai d'allumage s'avère hors de cette plage à DCN selon :
- $DCN = 83.99 (\tau - 1.512)^{-0.658} + 3.547$

Une autre chambre de combustion à volume fixe appelée ASTM D7170 décrit le test d'allumage du fuel et quantifie le nombre de cétane dérivé ; les valeurs FIT-DCN peuvent différer des valeurs IQT-DCN par une grande différence en raison de conditions limites différentes.

### - **Indice d'octane :**

La principale cause de crépitement du moteur est l'allumage spontané du mélange d'essence. Bien que la modélisation cinétique détaillée ait considérablement amélioré la compréhension de la chimie de la combustion spontanée des divers composants du carburant, Mon et Ron sont encore largement utilisés comme références mondiales pour le comportement des antis clique les carburants.

Pour un segment ou une combinaison de carburant particulier, ceux-ci sont estimés dans des conditions de travail pré-caractérisées en augmentant régulièrement la proportion de pression d'un moteur CFR plus que possible. Il est classé comme le pourcentage d'iso-octane dans une combinaison avec le n-heptane avec une limite de bruit sourd telle que le carburant examiné. Pour des valeurs supérieures à 100, les valeurs RON et MON sont déterminées sur la base de la teneur en plomb tétraéthyle en combinaison avec un iso-octane ayant le même comportement antisalissure que la composition de carburant considérée.

La principale différence entre les deux essais est le régime moteur élevé et la température de l'air d'admission dans le test MON. De plus, la température d'entrée dans le test RON n'est déterminée que pour l'air, alors qu'elle est indiquée pour le mélange vaporisé dans le test MON. Ainsi, l'effet du refroidissement de la charge dû aux différentes températures de vaporisation des segments de carburant fait partie de la méthode donnée dans les estimations RON, alors qu'il n'affecte pas MON.

Notez que les taux d'incendie des fractions de carburant peuvent également affecter les indices d'octane, quels que soient les effets du comportement de démarrage automatique et de la chaleur de vaporisation.

Kalghatgi a développé le concept de l'indice d'octane (OI), Afin de mieux décrire la qualité antidétonante du carburant dans les conditions de fonctionnement réelles d'un moteur particulier, qui se caractérise par [06] :

$$OI = (1-K)*RON + K*MON \quad (1)$$

Où K est un facteur basé sur les conditions de fonctionnement du moteur et n'a rien à voir avec le carburant. K prend la valeur 1 ou 0 si les conditions de fonctionnement du moteur correspondent à celles des tests pour MON et RON, puis lors de la saisie de l'OS qui représente la différence entre RON et MON pour un carburant particulier, c'est :

$$OS = RON - MON \quad (2)$$

Donc l'équation (1) :

$$OI = RON - K*OS \quad (3)$$

Les conditions d'essai MON étaient initialement destinées à traiter les conditions de travail courantes des moteurs à carburant. Néanmoins, il y a eu des améliorations dans l'innovation des moteurs telles que l'injection directe, la turbocompression et le refroidissement intermédiaire à des températures d'admission nettement inférieures par rapport aux pressions d'amortissement des moteurs SI modernes.

### **V) Conclusion:**

La plupart des méthodes utilisées pour prédire la qualité de l'allumage du carburant consistent à étudier les indices de cétane et les indices d'octane en les testant dans le moteur (CN, MON,

RON), mais la plupart d'entre eux sont appliqués à des véhicules purs. Une autre méthode a donc été proposée pour prédire simultanément les indicateurs (IC, MON, RON) de mélanges de carburants purs, et cette méthode dépend de modèles d'apprentissage automatique.

Un modèle mathématique développé par machine Learning relie une matrice de structure moléculaire basée sur les propriétés structurelles des molécules de carburant à une matrice de qualité d'allumage.

Ce procédé est appliqué à un large éventail de composés tels que les alcanes, les alcènes, leurs dérivés, les alcools, les esters, les acides, les mélanges de carburants et d'autres composés.



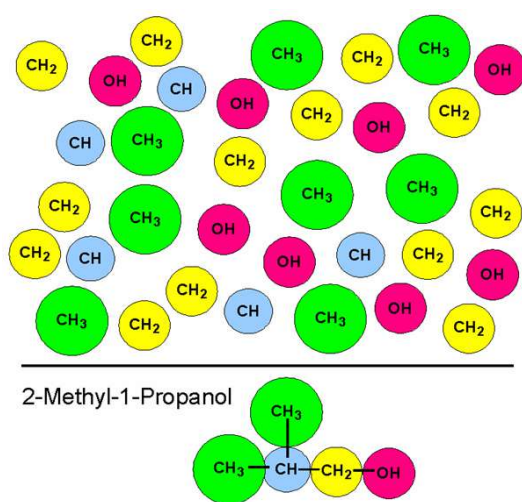


## I. Introduction :

Dans ce chapitre on va discuter le principe des techniques d'estimation des propriétés des molécules en se basant sur le principe d'additivité des groupes fonctionnels. Nous présentons par la suite la base de données ainsi la démarche de modélisation de l'estimation de l'indice de cétane basée sur les groupes de contribution en expliquant les étapes de la régression linéaire, le logiciel de XLSTAT et les hypothèses de la modélisation.

## II. Description de la méthode contribution des groupes et modélisation :

La méthode de contribution de groupes est une méthode qui utilise les informations structurales de base d'une molécule chimique, par exemple une liste de groupes fonctionnels simples ( $\text{CH}_3^-$ ,  $-\text{CH}_2^-$ ,  $>\text{CH}-$ ,  $>\text{C}<$ ,  $=\text{CH}_2$ , ...), associe des paramètres à ces groupes fonctionnels et calcule les propriétés physico-chimiques du composé en fonction de la somme des paramètres de groupes.



Cette méthode suppose qu'il n'y a pas d'interactions entre les groupes, et ne fait donc que sommer les contributions de chaque groupe sans ajouter de terme pour les interactions entre les groupes. L'utilisation de paramètres de groupes simples présente néanmoins l'avantage de ne nécessiter qu'un petit nombre de paramètres. : le nombre de paramètres d'interaction de groupe à prendre en compte devient en effet très élevé pour un nombre croissant de groupes (un seul pour deux groupes, trois pour trois groupes, six pour quatre groupes, 45 pour dix groupes et deux fois plus si les interactions ne sont pas symétriques).

Dans cette étude, la structure moléculaire d'un composé est considérée d'approximation: contribution du premier ordre. Le modèle développé est basé sur les principes dans ces points suivants:

1. En premier, d'un composé est considérée comme étant une fonction des paramètres de structure à charge qui sont ainsi déterminés en additionnant la fréquence du nombre de chaque groupe simple interaction se produisant dans la molécule.

## Chapitre 2 : Modélisation de la technique de prédiction des propriétés

2. La corrélation multiple linéaire de l'équation (1) est corrigée en utilisant l'équation ci-dessous:

$$\text{Propriété (P)} = \mathbf{F}(\text{structure}) \quad \text{Eq. (1)}$$

3. L'équation (1) est prédite des propriétés (P) en utilisant de multiples corrélations linéaires et les paramètres sont corrigés en utilisant l'équation suivante:

$$P = f\left(A + \sum_j n_j \Delta c_j\right) \quad \text{Eq. (2)}$$

Où  $n_j$  sont le nombre du premier ordre des groupes de type  $j$  dans la molécule et  $\Delta c_j$  sont des contributions pour le premier ordre et  $A$  est une constante.

4. La détermination des paramètres de réglage et les valeurs de groupe pour le modèle de corrélation multiple linéaire sont évalués par une méthode d'optimisation. La fonction d'objectif a été donnée par l'équation suivante:

$$F = \sum_{i=1}^n (P_i^{Exp} - P_i^{Cal})^2 \quad \text{Eq. (03)}$$

où les exposants Exp et Cal se réfèrent à des valeurs expérimentales et calculées respectivement. Tandis que  $i$  représente les molécules de la base de données.

5. Les groupes fonctionnels utilisés dans cette étude sont mentionnés sur le tableau suivant.

**Tableau II- 1** Groupes fonctionnelles de contribution

Composés	Groupes de contribution
Hydrocarbonés (non-ring)	CH <sub>3</sub> -, -CH <sub>2</sub> -, >CH-, >C<, =CH <sub>2</sub> , =CH-, =C<
Hydrocarbonés (ring)	-CH <sub>2</sub> -, >CH-, >C<, =CH-, =C<
Oxygène (non-ring)	-OH, -O-, >C=O, O=CH-, -COO-
Oxygène (ring)	O=CH-, -OH

### III. Base de données (base de calcul indice de cétane) :

La base de données utilisée dans cette étude a été compilée à partir de plusieurs sources, y compris les données publiées, qui se composent de 475 composés chimiques : hydrocarbures et oxyhydrocarbures.

Et nous avons eu une touche dans cette base de données, qui consistait à déterminer le nombre de groupes fonctionnels pour chaque composé. Par exemple, le pentane a deux groupes fonctionnels, 2 de CH<sub>3</sub> et 3 de CH<sub>2</sub>. Le tableau II-1 détaille cette base de données.

#### – Hydrocarbures :

Le groupe des hydrocarbures est constitué des alcanes linéaires (Propane,.....), isomères (2-méthylbutane, ...) et cycliques (cyclopentane,...) à et constitué aussi des alcènes linéaires (1-Pentène,.....) et cycliques (cyclopentène,.....).

– **Oxyhydrocarbures :**

Le groupe des hydrocarbures oxygénés comprend les alcools linéaires (methanol, ...) et cycliques (cyclohexanol, ...), les aldéhydes (butanal, ...), les esters saturés (methyl acetate, ...) et insaturés (methyl sorbate, ...), les éthers cycliques (dimethyl ether, ...), les composés aromatiques oxygénés (ethyl acetoacetate, ...) et non oxygénés( benzene, ...), les esters dicarboxyliques (Dimethyl malonate, ...), les acides carboxyliques (Decanoic Acid, ...) et les composés polyoxygénés (ethyl acetoacetate, ...) et multifonctionnels (1-butoxy-2-propanol, ...).

**Tableau II- 2 Données relatives à l'indice de cétones utilisés dans ce travail**

Types	Familles chimiques	Sous-familles chimiques	Nombre	Les sources
Hydrocarbures	Alcanes	Alcanes linéaires	18	[10],[11]
		Iso-alcanes	57	
		Cyclo-alcanes	50	
	Alcènes	Alcènes	32	
		Cyclo-alcènes	02	
Oxyhydrocarbures	Alcools	Alcools linéaires	42	
		Cyclo-alcools	10	
	Aldéhydes	Aldéhydes	18	
	Esters	Esters saturés	62	
		Esters insaturés	18	
		Esters dicarboxyliques	17	
	Ethers	Ethers cycliques	51	
	Furanes	Furanes	17	
	Aromatiques	Aromatique non oxygénés	48	
		Aromatique oxygénés	07	
	Acides	Acides carboxyliques	05	
	Autre composés	Composés polyoxygénés	03	
		Composés multifonctionnels	16	

**IV. Description de la technique et la méthode utilisées pour l'indice de Cétane :**

Pour modéliser la technique de prédiction du cétane, nous nous sommes appuyés sur le programme XLSTAT, et pour faciliter le processus de modélisation, nous avons choisi le modèle de régression linéaire.

– **Régression linéaire :**

La régression linéaire est sans aucun doute la méthode statistique la plus utilisée. On distingue habituellement la régression simple (une seule variable explicative) de la régression multiple (plusieurs variables explicatives) bien que le cadre conceptuel et les méthodes de calculs soient identiques. Le principe de la régression linéaire est de modéliser une variable dépendante quantitative Y, au travers d'une combinaison linéaire de p variables explicatives

## Chapitre 2 : Modélisation de la technique de prédiction des propriétés

quantitatives,  $X_1; X_2; X_3; \dots; X_p$ . Le modèle déterministe (ne prenant pas en compte d'aléa) s'écrit pour une observation  $i$ ,

$$y_i = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} + \epsilon_i$$

Où  $y_i$  est la valeur observée pour la variable dépendante pour l'observation  $i$ ,  $x_{ij}$  est la valeur prise par la variable  $j$  pour l'observation  $i$ , et  $\epsilon_i$  est l'erreur du modèle.

Le cadre statistique et les hypothèses qui l'accompagnent ne sont pas nécessaires pour ajuster ce modèle. Par ailleurs la minimisation par la méthode des moindres carrés (on minimise la somme des erreurs quadratiques  $\epsilon_i^2$  fournit une solution analytique exacte. Néanmoins si l'on veut pouvoir tester des hypothèses et mesurer le pouvoir explicatif des différentes variables explicatives dans le modèle, un cadre statistique est nécessaire. Les hypothèses de la régression linéaire sont les suivantes : les erreurs  $\epsilon_i$  suivent une même loi normale  $N(0, s)$  et sont indépendantes. L'écriture du modèle complétée par cette hypothèse a pour conséquence que, dans le cadre du modèle de régression linéaire, les  $y_i$  sont des réalisations de variables aléatoires de moyenne  $\mu_i$  et de variance  $s^2$ , avec

$$\mu_i = \beta_0 + \sum_{j=1}^p \beta_j x_{ij}$$

Les estimateurs des coefficients  $\beta$  et de leur variance sont donnés par :

$$\hat{\beta} = (X^t X)^{-1} X^t Y$$

Et

$$Var(\hat{\beta}) = \hat{\sigma}^2 (X^t X)^{-1}$$

Si l'on souhaite utiliser les différents tests proposés dans les résultats de la régression linéaire il est recommandé de vérifier a posteriori que les hypothèses sous-jacentes sont bien vérifiées. La normalité des résidus peut être vérifiée en analysant certains graphiques ou en utilisant un test de normalité. L'indépendance des résidus peut être vérifiée en analysant certains graphiques ou en utilisant le test de Durbin Watson disponible dans les outils d'analyse de séries chronologiques de XLSTAT.

## V. Description de l'outil logiciel –XLSTAT :

XLSTAT est développé depuis plus de dix ans et vise à fournir au plus grand nombre des outils d'analyse de données et statistiques puissants, complets et conviviaux. Il s'agit d'un outil d'analyse de données et de statistiques dans Microsoft Excel. Développé depuis plus de 10 ans dans le but de fournir au plus grand nombre des données puissantes, complètes et simples d'utilisation, il dispose de 240 fonctions statistiques allant des statistiques fonctionnelles à la machine learning.

Toutes les fonctions XLSTAT peuvent être ajoutées à partir de la barre de menu Excel. L'utilisation d'Excel comme interface rend le programme facile à utiliser et efficace. La qualité du calcul est la même que la qualité du programme. La puissance de XLSTAT vient à la fois du langage de programmation C++ et des algorithmes utilisés, qui sont le résultat des travaux de recherche de centaines de statisticiens, mathématiciens et informaticiens. A chaque fois qu'une nouvelle fonction XLSTAT sera développée, une phase de recherche bibliographique approfondie sera menée, voire des discussions avec des experts des méthodes associées seront menées. XLSTAT est l'outil statistique préféré des entreprises, des universités et de plus de 100 000 utilisateurs dans plus de 120 pays/régions.

## VI. Hypothèses de la modélisation :

Dans ce travail, nous avons choisi trois hypothèses de la modélisation ont été utilisés, à savoir, Stepwise (pour les H/C), Descendante (pour les HCO) et le meilleur modèle (pour tous les molécules) :

- **Stepwise** : le processus de sélection commence par l'ajout de la variable ayant la plus forte contribution au modèle (le critère utilisé est la statistique t de Student). Si une seconde variable est telle que la probabilité associée à son t est inférieure à la « **Probabilité pour l'entrée** », elle est ajoutée au modèle. De même pour une troisième variable. A partir de l'ajout de la troisième variable, après chaque ajout, on évalue pour toutes les variables présentes dans le modèle quel serait l'impact de son retrait (toujours au travers de la statistique t). Si la probabilité est supérieure à la « **Probabilité pour le retrait** », la variable est retirée. La procédure se poursuit jusqu'à ce que plus aucune variable ne puisse être ajoutée/retirée.
- **Descendante** : la procédure commence par l'ajout simultané de toutes les variables. Les variables sont ensuite retirées du modèle suivant la procédure utilisée pour la sélection progressive.
- **Meilleur modèle** : cette méthode permet de choisir le meilleur modèle parmi tous les modèles comprenant un nombre de variables variant de « Min variables » à « Max variables ». Par ailleurs le « critère » pour déterminer le meilleur modèle peut être choisi par l'utilisateur.

## *Chapitre 2 : Modélisation de la technique de prédiction des propriétés*

Le tableau ci-dessous détaille l'hypothèse de modélisation que nous avons utilisée avec les composés ; et la constante constante que nous avons définie ; Également on a proposé la Validation aléatoire

**Tableau II- 3 Hypothèses utilisés dans la modélisation**

<b>Hypothèse de modélisation</b>	<b>Composés</b>		<b>Constante fixe</b>	<b>Validation aléatoire (Nombre de validation)</b>
Stepwise	Hydrocarbures	Sans MW	-105.75	50
		Avec MW		
descendante	Oxyhydrocarbures	Sans MW	-105.75	100
		Avec MW		
Meilleur modèle	tous les composés (HC et HCO)	Sans MW	-145	150
		Avec MW		



## I. Introduction :

Nous avons parlé dans les chapitres précédant un petit peu de les carburants et les biocarburants et les méthode d'estimation d'indice de cétane et octane ,aussi on a faire un description pour les techniques et les méthodes utilisé pour calculer l'indice de cétane , Nous avons également indiqué l'outil logiciel XLSTAT et hypothèses de modélisation. Dans ce chapitre, on veut présentons les résultats de la régression linéaire pour certaines des données expérimentales obtenues par XLSTAT sur les méthodes proposés. Nous comparerons également les résultats que nous avons obtenus avec les résultats montrés dans les études précédentes. Nous utiliserons deux formes de représentation des résultats, qui sont les résultats tabulaires et les graphiques.

## II. Présentation des résultats :

### II.1.Régressions linéaire de HC sans et avec masse moléculaire :

Constante fixée : -105,75

Intervalle de confiance (%) : 95

Tolérance : 0,0001

Validation : Aléatoire

Nombre d'observations pour la validation : 50

Sélection de modèle : Stepwise

Probabilité pour l'entrée : 0,05 / Probabilité pour le retrait : 0,1

**Matrice de corrélation** : ce tableau est affiché afin de vous permettre d'avoir un aperçu des corrélations entre les différentes variables sélectionnées.

**Tableau III- 1**Matrice de corrélation de base de données sans masse moléculaire

	CH3-	-CH2-	>CH-	>C<	=CH2	=CH-	=C<	-CH2-	>CH-	>C<	=CH-	=C<	Measured _CN
CH3-	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
-CH2-	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
>CH-	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
>C<	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
=CH2	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
=CH-	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000
=C<	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000
-CH2-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000
>CH-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000
>C<	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000
=CH-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000
=C<	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000
CN	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>

**Tableau III- 2**Matrice de corrélation de base de données avec masse moléculaire

Masse moléculaire (g/mol)	CH3-	-CH2-	>CH-	>C<	=CH2	=CH-	=C<	-CH2-	>CH-	>C<	=CH-	=C<	Measured_CN	
Masse moléculaire (g/mol)	1	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
CH3-	0,000	1	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
-CH2-	0,000	0,000	1	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
>CH-	0,000	0,000	0,000	1	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
>C<	0,000	0,000	0,000	0,000	1	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
=CH2	0,000	0,000	0,000	0,000	0,000	1	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
=CH-	0,000	0,000	0,000	0,000	0,000	0,000	1	0,000	0,000	0,000	0,000	0,000	0,000	0,000
=C<	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1	0,000	0,000	0,000	0,000	0,000	0,000
-CH2-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1	0,000	0,000	0,000	0,000	0,000
>CH-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1	0,000	0,000	0,000	0,000
>C<	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1	0,000	0,000	0,000
=CH-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1	0,000	0,000
=C<	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1	0,000
Measured_CN	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1

**Coefficients d'ajustement (Measured\_CN) :** dans ce tableau sont affichées les statistiques relatives à l'ajustement du modèle de régression.

**Tableau III- 3** Statistiques relatives à l'ajustement du modèle de régression

	sans masse moléculaire	avec masse moléculaire
R <sup>2</sup>	0,960	0,964
R <sup>2</sup> ajusté	0,956	0,960

**R<sup>2</sup> :** le coefficient de détermination du modèle. La valeur de ce coefficient est comprise entre 0 et 1. XLSTAT le calcule comme suit :

$$R^2 = \frac{\sum_{i=1}^n \omega_i (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^n \omega_i (y_i - \bar{y}_i)^2}$$

Le R<sup>2</sup> s'interprète comme la proportion de la variabilité de la variable dépendante expliquée par le modèle. Plus le R<sup>2</sup> est proche de 1, meilleur est le modèle. L'inconvénient du R<sup>2</sup> est qu'il ne prend pas en compte le nombre de variables utilisées pour ajuster le modèle.

**R<sup>2</sup> ajusté :** le coefficient de détermination ajusté du modèle. Le R<sup>2</sup> ajusté peut être négatif si le R<sup>2</sup> est voisin de zéro. Ce coefficient n'est affiché que si la constante du modèle n'est pas fixée par l'utilisateur. Sa valeur est définie par

$$\hat{R}^2 = 1 - (1 - R^2) \frac{W-1}{W-p-1}$$

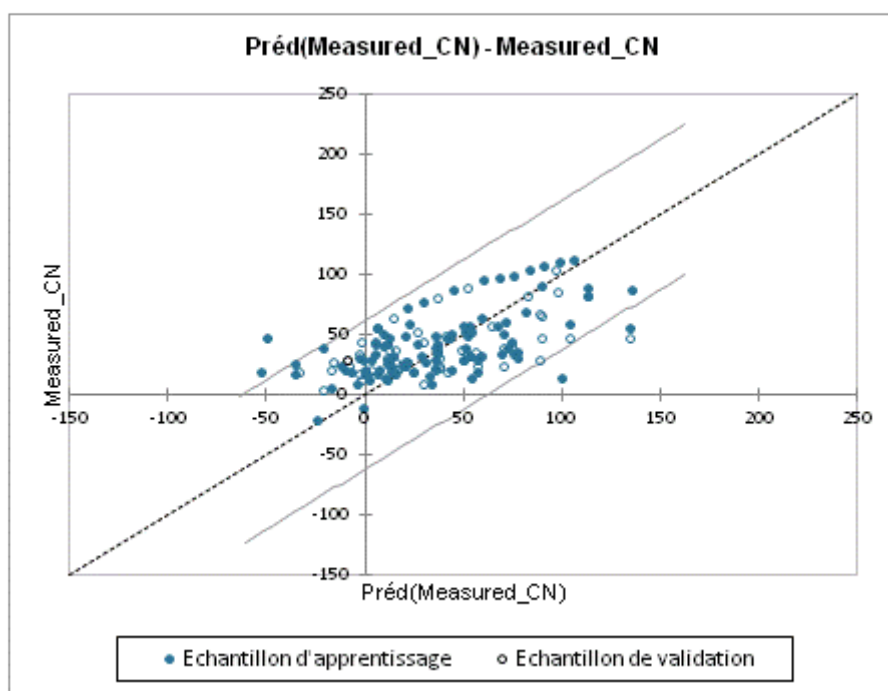
Le R<sup>2</sup> ajusté est une correction du R<sup>2</sup> qui permet de prendre en compte le nombre de variables utilisées dans le modèle.

Equation du modèle CN (sans masse moléculaire) :

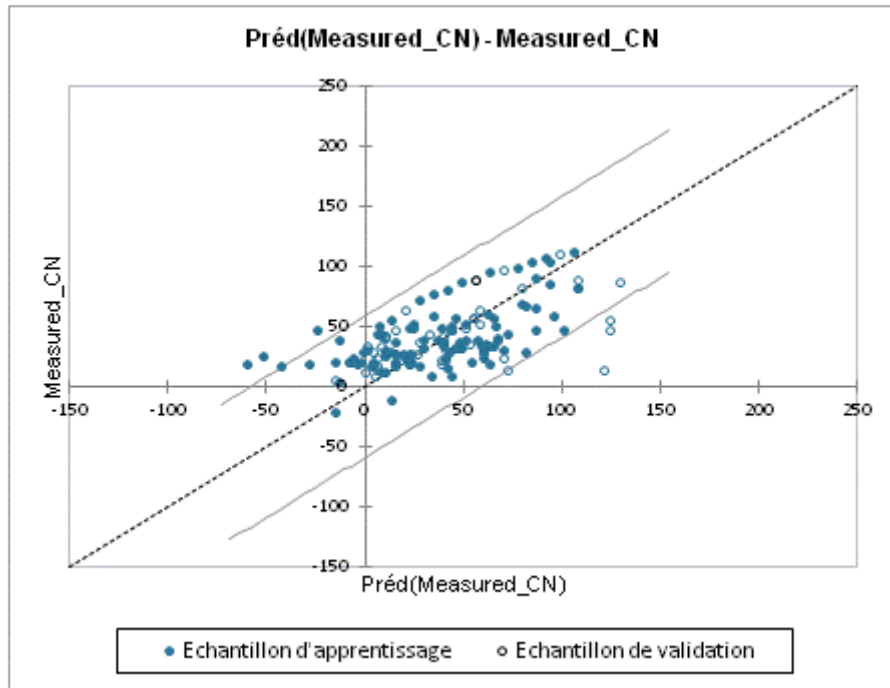
$$\text{CN} = -105,75 + 36,450 * (\text{CH}_3 -) + 7,697 * (-\text{CH}_2 -) - 22,391 * (> \text{CH} -) - 56,457 * (> \text{C} <) + 47,467 * (' = \text{CH}_2) + 19,215 * (' = \text{CH} -) - 43,080 * (' = \text{C} <) + 17,551 * (-\text{CH}_2 -) - 24,127 * (> \text{C} <)$$

1. Equation du modèle CN (avec masse moléculaire) :

$$\text{CN} = -105,75 + 1,287 * \text{Molecular weight (g/mol)} + 22,372 * (\text{CH}_3-) - 10,966 * (-\text{CH}_2-) - 49,296 * (> \text{CH}-) - 77,830 * (> \text{C} <) + 30,3782 * (' = \text{CH}_2) - 65,167 * (' = \text{C} <) - 21,3765 * (> \text{CH}-) - 42,771 * (> \text{C} <) - 36,0734 * (' = \text{C} <)$$



Le graphique III- 1 de la distance entre les prédictions et les observations de CN mesuré (sans MW)



**Le graphique III- 2 de la distance entre les prédictions et les observations de CN mesuré (avec MW)**

Le graphique III-3de la distance entre les prédictions et les observations de CN mesuré (sans MW). Les lignes parallèles du graphique représentent l'intervalle de confiance, où les points bleu foncé représentent l'indice de cétane expérimental et les autres points représentent le nombre de points calculés par la base de données dans cette étude. Le graphique de la distance entre les prédictions et les observations de CN mesuré (avec MW).

### **Discussions :**

Si les  $R$  et  $R$  du modèle sont proches de un et que les échantillons expérimentaux et calculés sont sur la même ligne dans l'intervalle de confiance du graphique correspondant à  $R$  et  $R$ , alors le modèle proposé est le meilleur modèle.

Pour les deux modèles que nous avons proposés dans l'étude HC, le meilleur modèle est celui dans lequel la masse moléculaire est introduite en tant que paramètre, **Tableau III-4** et **Les graphiques III-1 et III-2**, expliquent-ils.

### **Interprétation sans MW (Measured\_CN) :**

En utilisant la méthode de sélection Stepwise, 9 variables ont été retenues dans le modèle. Etant donné le  $R^2$ , 96% de la variabilité de la variable dépendante Measured\_CN sont expliqués par les 9 variables explicatives. Sur la base des sommes de carrés de type III, les variables suivantes apportent de l'information significative pour expliquer la variabilité de la variable dépendante Measured\_CN : CH3-, -CH2-, >CH-, >C<, '=CH2, '=CH-, '=C<, -CH2-, >C<.

Parmi les variables explicatives, sur la base des erreurs de type III, la variable -CH2- est la plus influente.

### **Interprétation (Measured\_CN) avec MW :**

En utilisant la méthode de sélection Stepwise, 10 variables ont été retenues dans le modèle. Etant donné le  $R^2$ , 96% de la variabilité de la variable dépendante Measured\_CN sont expliqués par les 10 variables explicatives. Sur la base des sommes de carrés de type III, les variables suivantes

apportent de l'information significative pour expliquer la variabilité de la variable dépendante Measured\_CN : Molecular weight (g/mol), CH<sub>3</sub>-, -CH<sub>2</sub>-, >CH-, >C<, =CH<sub>2</sub>, =C<, >CH-, >C<, =C<. Parmi les variables explicatives, sur la base des erreurs de type III, la variable Molecular weight (g/mol) est la plus influente.

**II.2. Régressions linéaire de HCO sans et avec masse moléculaire :**

Constante fixée : -105,75

Intervalle de confiance (%) : 95

Tolérance : 0,0001

Validation : Aléatoire

Nombre d'observations pour la validation : 100

Sélection de modèle : Descendante

Tableau III- 4 Matrice de corrélation de base de données sans masse moléculaire

	CH3-	-CH2-	>CH-	>C<	=CH2	=CH-	=C<	-CH2-	>CH-	>C<	=CH-	=C<	-OH	-O-	>C=O	O=CH-	-COO-	-O-	>C=O	Measured _CN	
CH3-	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
-CH2-	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
>CH-	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
>C<	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
=CH2	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
=CH-	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
=C<	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
-CH2-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
>CH-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
>C<	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
=CH-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
=C<	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
-OH	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
-O-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000
>C=O	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000
O=CH-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000
-COO-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000
-O-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000
>C=O	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000
Measur ed_CN	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>

Tableau III- 5 Matrice de corrélation de base de données avec masse moléculaire

	Molecular weight (g/mol)	CH3-	-CH2-	>CH-	>C<	=CH2	=CH-	=C<	-CH2-	>CH-	>C<	=CH-	=C<	-OH	-O-	>C=O	O=CH-	-COO-	-O-	>C=O	Measured _CN		
MW (g/mol)	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	
CH3-	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
-CH2-	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
>CH-	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
>C<	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
=CH2	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
=CH-	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
=C<	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
-CH2-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
>CH-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
>C<	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
=CH-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
=C<	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
-OH	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
-O-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
>C=O	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000
O=CH-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000
-COO-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000
-O-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000
>C=O	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000
Measured _CN	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000

**Coefficients d'ajustement (Measured\_CN) :**

Dans ce tableau sont affichées les statistiques relatives à l'ajustement du modèle de régression

**Tableau III- 6** Statistiques relatives à l'ajustement du modèle de régression

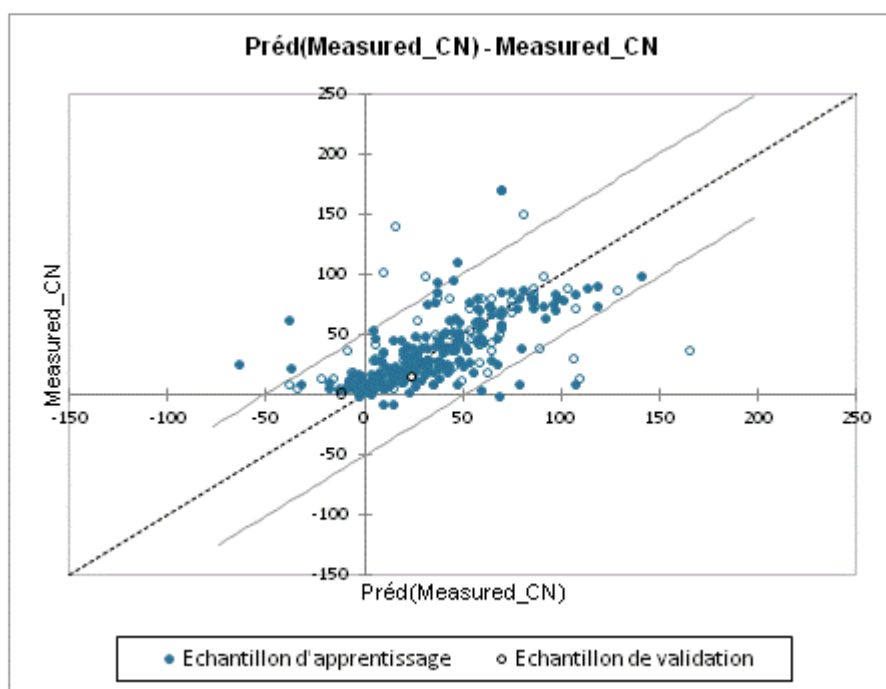
	sans masse moléculaire	avec masse moléculaire
R <sup>2</sup>	0,971	0,974
R <sup>2</sup> ajusté	0,969	0,972

**Equation du modèle CN (sans masse moléculaire) :**

$$\text{CN} = -105,75 + 43,743 * (\text{CH3 } -) + 5,387 * (-\text{CH2 } -) - 34,174 * (> \text{CH } -) - 60,210 * (> \text{C } <) + 31,213 * (' = \text{CH2}) - 37,967 * (' = \text{C } <) + 19,522 * (-\text{CH2 } -) - 16,904 * (> \text{CH } -) - 38,5817 * (> \text{C } <) + 14,902 * (' = \text{CH } -) - 25,684 * (' = \text{C } <) + 48,416 * (-\text{OH}) + 21,834 * (-\text{O } -) + 71,673 * \text{O} = \text{CH } - + 11,648 * (-\text{COO } -) + 51,824 * (-\text{O } -)$$

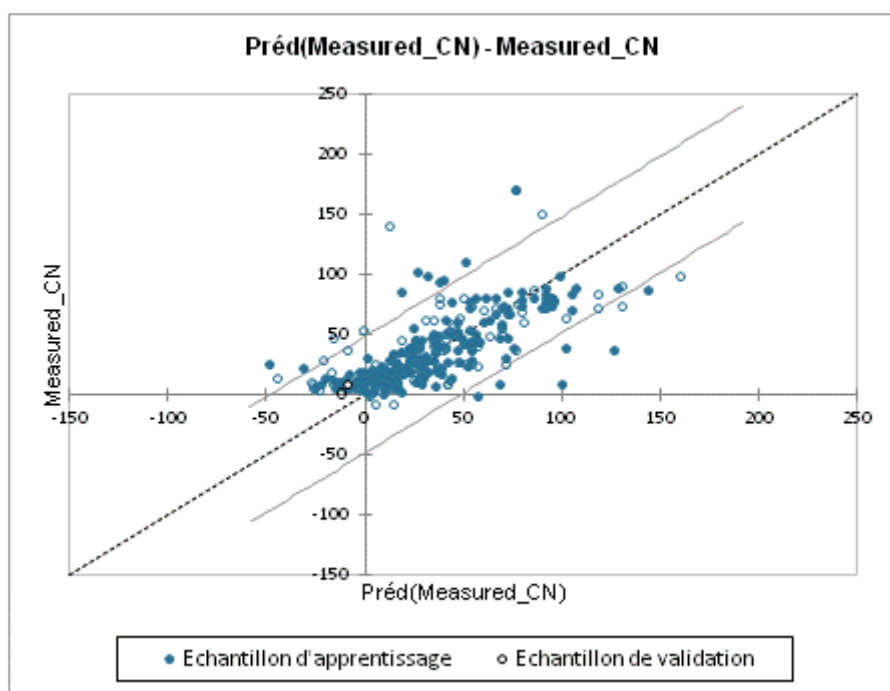
**Equation du modèle CN (avec masse moléculaire) :**

$$\text{CN} = -105,75 + 0,459 * \text{MW} \left( \frac{\text{g}}{\text{mol}} \right) + 32,586 * (\text{CH3 } -) - 38,278 * (> \text{CH } -) - 71,459 * (> \text{C } <) + 31,715 * (' = \text{CH2}) - 7,941 * (' = \text{CH } -) - 36,322 * (' = \text{C } <) + 15,362 * (-\text{CH2 } -) - 25,265 * (> \text{CH } -) - 53,136 * (> \text{C } <) + 8,156 * (' = \text{CH } -) - 25,244607028466 * (' = \text{C } <) + 34,959 * (-\text{OH}) + 18,316 * (-\text{O } -) + 64,032 * (\text{O} = \text{CH } -) - 11,450 * (-\text{COO } -) + 29,071 * (-\text{O } -)$$



**Le graphique III- 3 de la distance entre les prédictions et les observations de CN mesuré (sans MW)**





**Le graphique III- 4 de la distance entre les prédictions et les observations de CN mesuré (avec MW)**

### Discussion :

Pour les deux modèles que nous avons proposés dans l'étude HCO, le meilleur modèle est celui dans lequel la masse moléculaire est introduite en tant que paramètre, **Tableau III-7** et **Les graphiques III-3 et III-4**, expliquent-ils.

### Interprétation sans MW (Measured\_CN) :

En utilisant la méthode de sélection Descendante, 16 variables ont été retenues dans le modèle. Etant donné le  $R^2$ , 97% de la variabilité de la variable dépendante Measured\_CN sont expliqués par les 16 variables explicatives. Sur la base des sommes de carrés de type III, les variables suivantes apportent de l'information significative pour expliquer la variabilité de la variable dépendante Measured\_CN : CH3-, -CH2-, >CH-, >C<, '=CH2, '=C<, -CH2-, >CH-, >C<, '=CH-, '=C<, -OH, -O-, O=CH-, -COO-, -O-.

Parmi les variables explicatives, sur la base des erreurs de type III, la variable CH3- est la plus influente.

### Interprétation avec MW (Measured\_CN) :

En utilisant la méthode de sélection Descendante, 17 variables ont été retenues dans le modèle. Etant donné le  $R^2$ , 97% de la variabilité de la variable dépendante Measured\_CN sont expliqués par les 17 variables explicatives. Sur la base des sommes de carrés de type III, les variables suivantes apportent de l'information significative pour expliquer la variabilité de la variable

dépendante Measured\_CN: Molecular weight (g/mol),CH3-,>CH-,>C<,'=CH2,'=CH-,'=C<,-CH2-,>CH-,>C<,'=CH-,'=C<,-OH,-O-,O=CH-,-COO-,-O-.

Parmi les variables explicatives, sur la base des erreurs de type III, la variable Molecular weight (g/mol) est la plus influente.

### **II.3.Régressions linéaire de tous les composés sans et avec masse moléculaire :**

Constante fixée : -145

Intervalle de confiance (%) : 95

Tolérance : 0,0001

Validation : Aléatoire

Nombre d'observations pour la validation : 150

Sélection de modèle : Meilleur modèle / R<sup>2</sup> ajusté

Min variables : 1 / Max variables : 20

Graine (nombres aléatoires) : 1169121854

Tableau III- 7 Matrice de corrélation de base de données sans masse moléculaire

	CH3-	-CH2-	>CH-	>C<	=CH2	=CH-	=C<	-CH2-	>CH-	>C<	=CH-	=C<	-OH	-O-	>C=O	O=CH-	-COO-	-O-	>C=O	Measur ed _CN	
CH3-	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
-CH2-	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
>CH-	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
>C<	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
=CH2	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
=CH-	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
=C<	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
-CH2-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
>CH-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
>C<	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
=CH-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
=C<	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
-OH	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
-O-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000
>C=O	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000
O=CH-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000
-COO-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000
-O-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000
>C=O	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000
Measur ed _CN	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>

Tableau III- 8 Matrice de corrélation de base de données avec masse moléculaire

	MW(g/ mol)	CH3-	-CH2-	>CH-	>C<	=CH2	=CH-	=C<	-CH2-	>CH-	>C<	=CH-	=C<	-OH	-O-	>C=O	O=CH-	-COO-	-O-	>C=O	Measur ed_CN	
MW	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
CH3-	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
-CH2-	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
>CH-	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
>C<	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
=CH2	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
=CH-	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
=C<	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
-CH2-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
>CH-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
>C<	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
=CH-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
=C<	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
-OH	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
-O-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000	0,000
>C=O	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000	0,000
O=CH-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000	0,000
-COO-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000	0,000
-O-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000	0,000
C=O	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>	0,000	0,000
M_CN	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	<b>1</b>

**Coefficients d'ajustement (Measured\_CN) :** dans ce tableau sont affichées les statistiques relatives à l'ajustement du modèle de régression

**Tableau III- 9** Statistiques relatives à l'ajustement du modèle de régression

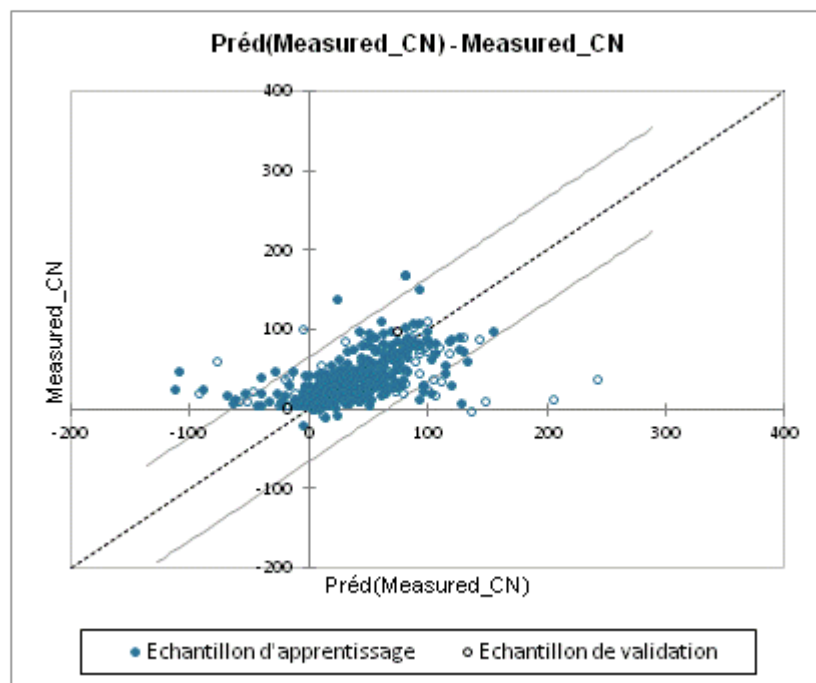
	sans masse moléculaire	avec masse moléculaire
R <sup>2</sup>	0,970	0,968
R <sup>2</sup> ajusté	0,968	0,966

**Equation du modèle CN (sans masse moléculaire) :**

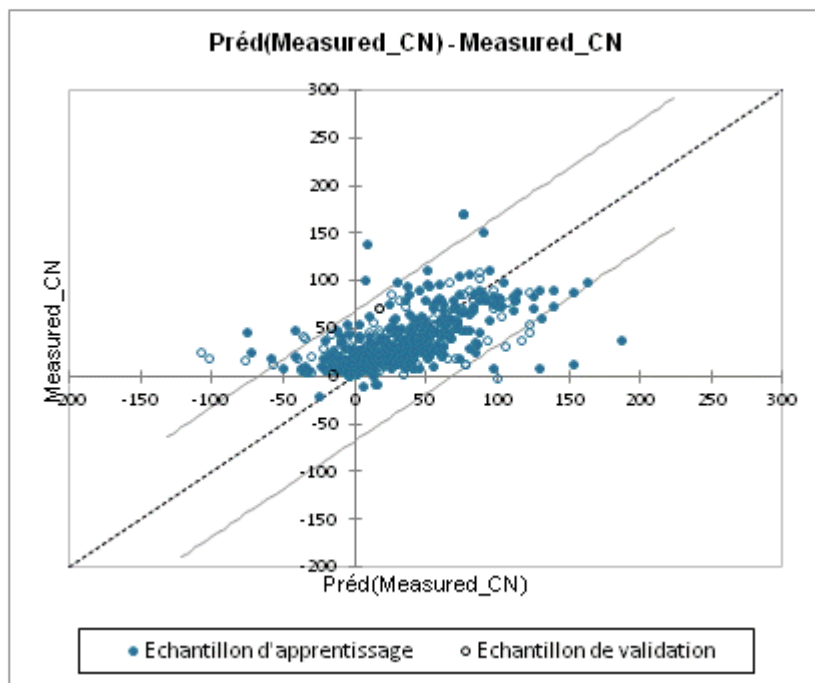
$$\text{CN} = -145 + 65,999 * (\text{CH3 } -) + 6,181 * (-\text{CH2 } -) - 58,598 * (> \text{CH } -) - 119,067 * (> \text{C } <) + 71,604 * (' = \text{CH2}) - 59,919 * (' = \text{C } <) + 24,118 * (-\text{CH2 } -) - 10,454 * (> \text{CH } -) - 67,486 * (> \text{C } <) + 21,999 * (' = \text{CH } -) - 40,009 * (' = \text{C } <) + 59,398 * (-\text{OH}) + 22,648 * (-\text{O } -) + 109,253 * (\text{O} = \text{CH}-) + 55,578 * (-\text{O } -)$$

**Equation du modèle CN (avec masse moléculaire) :**

$$\text{CN} = -145 + 0,681 * \text{MW} \left( \frac{\text{g}}{\text{mol}} \right) + 45,475 * (\text{CH3 } -) - 2,5097 * (-\text{CH2 } -) - 50,470 * (> \text{CH } -) - 94,828 * (> \text{C } <) + 62,965 * (' = \text{CH2}) - 11,182 * (' = \text{CH } -) - 54,560 * (' = \text{C } <) + 15,009 * (-\text{CH2 } -) - 24,525 * (> \text{CH } -) - 53,840 * (> \text{C } <) + 11,666 * (' = \text{CH } -) - 39,066 * (' = \text{C } <) + 45,825 * (-\text{OH}) + 15,951 * (-\text{O } -) + 75,458 * (\text{O} = \text{CH } -) - 20,099 * (-\text{COO } -) + 45,475 * (-\text{O } -)$$



**Le graphique III- 5 de la distance entre les prédictions et les observations de CN mesuré (sans MW).**



**Le graphique III- 6 de la distance entre les prédictions et les observations de CN mesuré (avec MW)**

### Discussion :

Pour les deux modèles que nous avons proposés dans l'étude tous les composés , le meilleur modèle est celui dans lequel la masse moléculaire est introduite en tant que paramètre, **Tableau III-10** et **Les graphiques III-5 et III-6**, expliquent-ils.

### Interprétation sans MW (Measured\_CN) :

En utilisant la méthode de sélection Meilleur modèle, 15 variables ont été retenues dans le modèle. Etant donné le  $R^2$ , 97% de la variabilité de la variable dépendante Measured\_CN sont expliqués par les 15 variables explicatives. Sur la base des sommes de carrés de type III, les variables suivantes apportent de l'information significative pour expliquer la variabilité de la variable dépendante Measured\_CN : CH<sub>3</sub>-, -CH<sub>2</sub>-, >CH-, >C<-, '=CH<sub>2</sub>-, '=C<-, -CH<sub>2</sub>-, >CH-, >C<-, '=CH-, '=C<-, -OH-, -O-, O=CH-, -O-.

Parmi les variables explicatives, sur la base des erreurs de type III, la variable CH<sub>3</sub>- est la plus influente.

### Interprétation avec MW (Measured\_CN) :

En utilisant la méthode de sélection Meilleur modèle, 18 variables ont été retenues dans le modèle. Etant donné le  $R^2$ , 97% de la variabilité de la variable dépendante Measured\_CN sont expliqués par les 18 variables explicatives. Sur la base des sommes de carrés de type III, les variables suivantes apportent de l'information significative pour expliquer la variabilité de la variable dépendante

Measured\_CN : Molecular weight (g/mol), CH<sub>3</sub>-, >CH-, >C<, '=CH<sub>2</sub>, '=CH-, '=C<, -CH<sub>2</sub>-, >CH-, >C<, '=CH-, '=C<, -OH-, -O-, O=CH-, -COO-, -O-.

Sur la base des sommes de carrés de type III, les variables suivantes n'apportent pas d'information significative pour expliquer la variabilité de la variable dépendante Measured\_CN : -CH<sub>2</sub>-. Vous pouvez essayer de les supprimer du modèle.

Parmi les variables explicatives, sur la base des erreurs de type III, la variable CH<sub>3</sub>- est la plus influente.

### **La comparaison avec des autres études :**

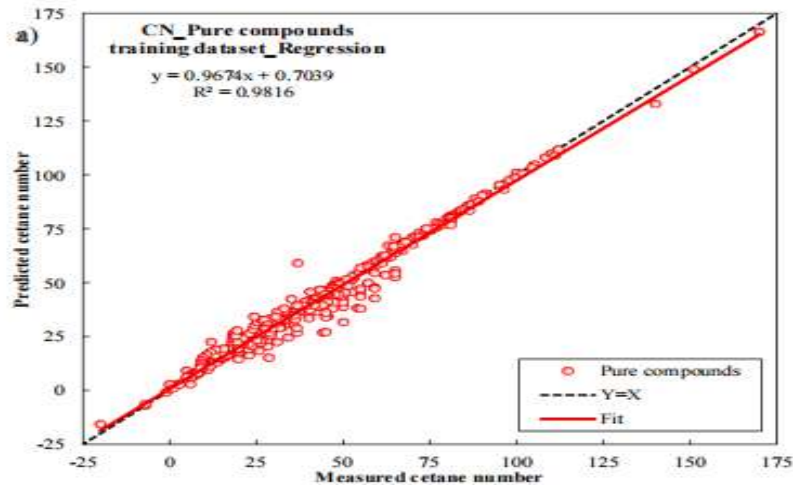
#### **Article [9] :**

Par rapport à l'étude décrite dans l'article {}, nous constatons que la base de données de cet article contient 483 composés du CN, où le R<sup>2</sup> de la méthode prédictive proposée pour les mélanges de carburants était de 0,98, tandis que pour notre études, la base de données collectée et modifiée par nous ; contient 475 et la valeur de R<sup>2</sup> était de 0,970 pour le modèle dans lequel le poids moléculaire n'a pas été entré comme variable et de 0,968 pour le modèle dans lequel le poids moléculaire a été entré comme variable.

**La figure III-7** et le **tableau III-11** ci-dessous dans cet article et **les figures III-5/6** et le **tableau III-10** ci-dessus dans notre étude clarifier la différence entre les études.

**Tableau III- 10** Statistiques relatives à l'ajustement du modèle de régression[9]

Property	Training dataset	R-squared	MAE	RMSE
CN	Pure compounds (483)	0.9816	1.891	3.580
	Pure compounds & Mixtures (603)	0.9911	1.460	2.526



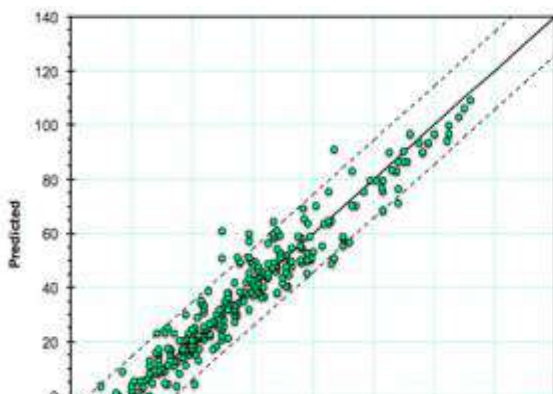
Le graphique III- 7 d'équivalence pour les valeurs mesurées et prédictives de CN par un modèle de régression d'auto-apprentissage pour l'ensemble de données de composés purs.

Article [10] :

Par rapport à l'étude décrite dans l'article {}, nous constatons que la base de données de cet article contient 449 composés CN, et la modélisation dans cet article était basée sur le réseau de neurones artificiels, où le R2 des hydrocarbures oxygénés était de 0,86, tandis que pour les hydrocarbures, il était supérieur à 0,9. Le tableau ci-dessous montre la différence entre les deux études.

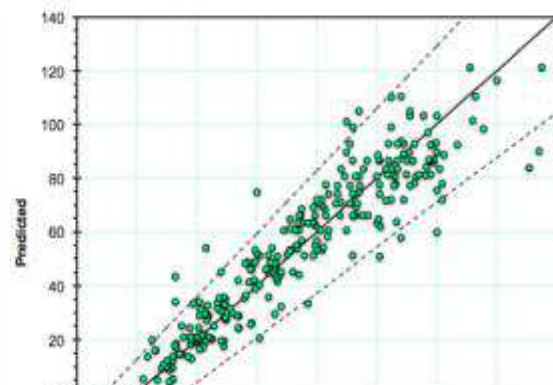
Tableau III- 11 la différence de R<sup>2</sup> entre les études

	R <sup>2</sup>	
	Notre étude	Les études d'article
CN - hydrocarbures	0,960	0.93
CN - hydrocarbures oxygénés	0,971	0.86



Le graphique III- 9 Valeurs prédites par rapport aux valeurs mesurées de CN pour les hydrocarbures [10] .

Fig. 17. Predicted versus measured values of CN for hydrocarbons.



Le graphique III- 8 Valeurs prédites par rapport aux valeurs mesurées de CN pour les hydrocarbures oxygénés [10] .

Fig. 18. Predicted versus measured values of CN for oxygenated hydrocarbons.



## **Conclusion :**

Un modèle de contribution du groupe de premier ordre de 474 molécules comme base de données a été développé pour prédictif / estimer la valeur de l'indice de cétane. La prise de décision au cours de la modélisation est de prendre des valeurs à la fois pour les hydrocarbures oxygénés ou non oxygénés. Cette base de données expérimentale a été documentée pour réaliser la modélisation sous XLSTAT.

Dans cette logique de prédiction, un chimiste souhaite cibler uniquement des nouvelles entités moléculaires en tant que composants potentiels de carburant dans la synthèse organique, qui sont prometteurs bien qu'ils n'aient pas encore été synthétisés, peut utiliser cette technique de prédiction intégrée à un modèle de conception assistée par ordinateur du combustible constitue pour la recherche de ces carburants.

L'épuisement des réserves pétrolières de l'Algérie nous oblige à utiliser des ressources vitales pour couvrir les besoins en ces matériaux, et le raffinage de ces sources est une mesure audacieuse pour l'Algérie, car il est temps de compter sur les biocarburants dans notre pays.

Le but de cette étude était de trouver ou de créer une nouvelle façon de contribuer à la prédiction de la caractéristique CN qui détermine la qualité du carburant (combustible pur dans cette étude), où cette méthode était d'étudier la contribution de chaque groupe individuellement, c'est-à-dire en négligeant l'interaction qui existe entre les groupes, et ces informations peuvent ne pas être exactes, mais nous attendons et suggérons que l'interaction qui existent entre les groupes soient incluses dans les études à venir, et nous suggérons également que cette base de données soit modélisée via la plateforme en ligne OCHEM afin de contribuer à trouver une nouvelle méthode fiable pour la prévision des indices CN, RON et MON.

## ANNEXES

### Les résultats de H/C :

#### Statistiques descriptives :

Les tableaux de statistiques descriptives présentent pour toutes les variables sélectionnées des statistiques simples. Pour les variables dépendantes et les variables explicatives quantitatives, sont affichés le nombre d'observations, le nombre de données manquantes, le nombre de données non manquantes, la moyenne, et l'écart-type (non biaisé).

Le tableau III-1 et III-2 présentent les statistiques descriptives de la base de données qui présente les minimum, les maximum, la moyenne des valeurs qui ont été utilisés dans la modélisation.

**Tableau III-1** Statistique descriptive de base de données sans la masse moléculaire

Variable	Observations	Obs. avec données manquantes	Obs. sans données manquantes	Minimum	Maximum	Moyenne	Ecart-type
Measured_CN	108	0	108	-20,000	112,000	40,923	25,910
CH3-	108	0	108	0,000	9,000	2,444	1,549
-CH2-	108	0	108	0,000	20,000	4,315	5,173
>CH-	108	0	108	0,000	3,000	0,463	0,766
>C<	108	0	108	0,000	3,000	0,111	0,395
=CH2	108	0	108	0,000	2,000	0,185	0,414
=CH-	108	0	108	0,000	4,000	0,296	0,687
=C<	108	0	108	0,000	3,000	0,176	0,577
-CH2-	108	0	108	0,000	10,000	1,741	2,559
>CH-	108	0	108	0,000	5,000	0,602	1,076
>C<	108	0	108	0,000	2,000	0,083	0,310
=CH-	108	0	108	0,000	2,000	0,056	0,268
=C<	108	0	108	0,000	1,000	0,046	0,211
-OH	108	0	108	0,000	0,000	0,000	0,000
-O-	108	0	108	0,000	0,000	0,000	0,000
>C=O	108	0	108	0,000	0,000	0,000	0,000
O=CH-	108	0	108	0,000	0,000	0,000	0,000
-COO-	108	0	108	0,000	0,000	0,000	0,000
-O-	108	0	108	0,000	0,000	0,000	0,000
>C=O	108	0	108	0,000	0,000	0,000	0,000

**Tableau III-2** Statistique descriptives de base de données avec la masse moléculaire

Variable	Observations	Obs. avec données manquantes	Obs. sans données manquantes	Minimum	Maximum	Moyenne	Ecart-type
Measured_CN	108	0	108	-20,000	112,000	40,320	25,403
Molecular	108	0	108	44,094	296,562	149,446	60,492

## ANNEXES

weight (g/mol)							
CH3-	108	0	108	0,000	7,000	2,556	1,481
-CH2-	108	0	108	0,000	18,000	4,519	5,138
>CH-	108	0	108	0,000	3,000	0,509	0,755
>C<	108	0	108	0,000	2,000	0,139	0,373
=CH2	108	0	108	0,000	2,000	0,157	0,391
=CH-	108	0	108	0,000	4,000	0,204	0,576
=C<	108	0	108	0,000	3,000	0,139	0,520
-CH2-	108	0	108	0,000	10,000	1,583	2,495
>CH-	108	0	108	0,000	4,000	0,611	1,066
>C<	108	0	108	0,000	2,000	0,074	0,297
=CH-	108	0	108	0,000	2,000	0,065	0,283
=C<	108	0	108	0,000	2,000	0,065	0,283
-OH	108	0	108	0,000	0,000	0,000	0,000
-O-	108	0	108	0,000	0,000	0,000	0,000
>C=O	108	0	108	0,000	0,000	0,000	0,000
O=CH-	108	0	108	0,000	0,000	0,000	0,000
-COO-	108	0	108	0,000	0,000	0,000	0,000
-O-	108	0	108	0,000	0,000	0,000	0,000
>C=O	108	0	108	0,000	0,000	0,000	0,000

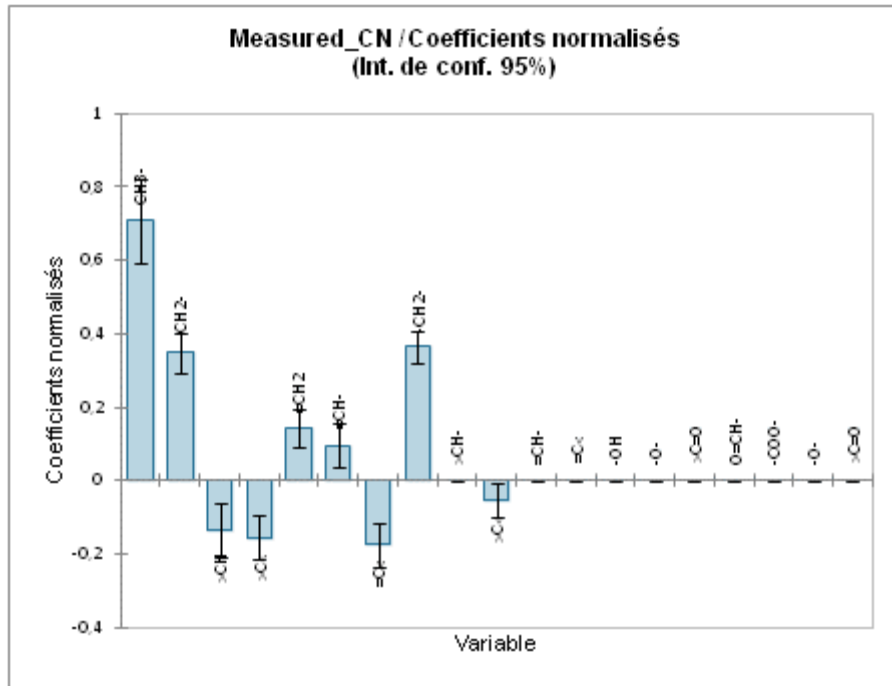
**Le tableau des paramètres du modèle :** affiche l'estimation des paramètres, l'écart-type correspondant, le t de Student, la probabilité correspondante, ainsi que l'intervalle de confiance. Si l'intervalle de confiance comprend 0, alors le poids d'une variable dans le modèle n'est pas significatif.

**Tableau III-6** Paramètres du modèle (Measured\_CN) avec et sans MW

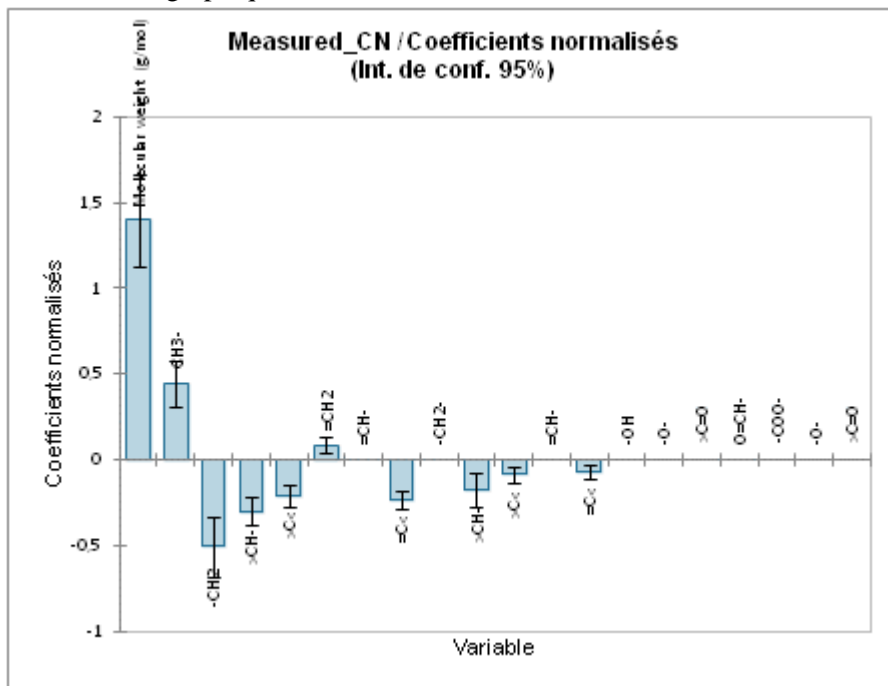
Source	avec MW	sans MW
	Valeur	Valeur
Constante	-105,750	-105,750
Molecular weight (g/mol)	1,287	0
CH3-	22,372	36,451
-CH2-	-10,967	7,698
>CH-	-49,297	-22,392
>C<	-77,831	-56,457
=CH2	30,378	47,468
=CH-	0,000	19,216
=C<	-65,167	-43,081
-CH2-	0,000	17,552
>CH-	-21,377	0,000
>C<	-42,772	-24,128
=CH-	0,000	0,000
=C<	-36,073	0,000
-OH	0,000	0,000
-O-	0,000	0,000
>C=O	0,000	0,000

## ANNEXES

O=CH-	0,000	0,000
-COO-	0,000	0,000
-O-	0,000	0,000
>C=O	0,000	0,000



Le graphique III-1 des coefficients normalisés sans MW



Le graphique III-1 des coefficients normalisés avec MW

## ANNEXES

### Les résultats de H/C/O :

**Tableau III-8** Statistique descriptives de base de données sans la masse moléculaire

Variable	Observations	Obs. avec données manquantes	Obs. sans données manquantes	Minimum	Maximum	Moyenne	Ecart-type
<b>Measured</b>							
_CN	212	0	212	-7,000	170,000	36,477	29,215
CH3-	212	0	212	0,000	6,000	1,797	1,031
-CH2-	212	0	212	0,000	23,000	5,175	5,319
>CH-	212	0	212	0,000	3,000	0,245	0,581
>C<	212	0	212	0,000	1,000	0,038	0,191
=CH2	212	0	212	0,000	2,000	0,042	0,224
=CH-	212	0	212	0,000	6,000	0,392	1,193
=C<	212	0	212	0,000	3,000	0,057	0,318
-CH2-	212	0	212	0,000	6,000	0,316	1,016
>CH-	212	0	212	0,000	3,000	0,071	0,337
>C<	212	0	212	0,000	2,000	0,014	0,153
=CH-	212	0	212	0,000	10,000	1,288	2,205
=C<	212	0	212	0,000	4,000	0,491	0,926
-OH	212	0	212	0,000	2,000	0,292	0,466
-O-	212	0	212	0,000	6,000	0,392	0,883
>C=O	212	0	212	0,000	1,000	0,033	0,179
O=CH-	212	0	212	0,000	1,000	0,033	0,179
-COO-	212	0	212	0,000	2,000	0,382	0,616
-O-	212	0	212	0,000	1,000	0,075	0,265
>C=O	212	0	212	0,000	1,000	0,028	0,166

**Tableau III-9** Statistique descriptives de base de données avec la masse moléculaire

Variable	Observations	Obs. avec données manquantes	Obs. sans données manquantes	Minimum	Maximum	Moyenne	Ecart-type
<b>Measured</b>							
_CN	212	0	212	-1,000	170,000	36,810	28,137
<b>MW</b>							
(g/mol)	212	0	212	46,068	424,728	172,345	77,476
CH3-	212	0	212	0,000	6,000	1,807	0,971
-CH2-	212	0	212	0,000	25,000	5,594	5,437
>CH-	212	0	212	0,000	3,000	0,193	0,520
>C<	212	0	212	0,000	1,000	0,033	0,179
=CH2	212	0	212	0,000	1,000	0,024	0,152
=CH-	212	0	212	0,000	6,000	0,396	1,252
=C<	212	0	212	0,000	3,000	0,047	0,288
-CH2-	212	0	212	0,000	6,000	0,344	1,071
>CH-	212	0	212	0,000	3,000	0,066	0,316

## ANNEXES

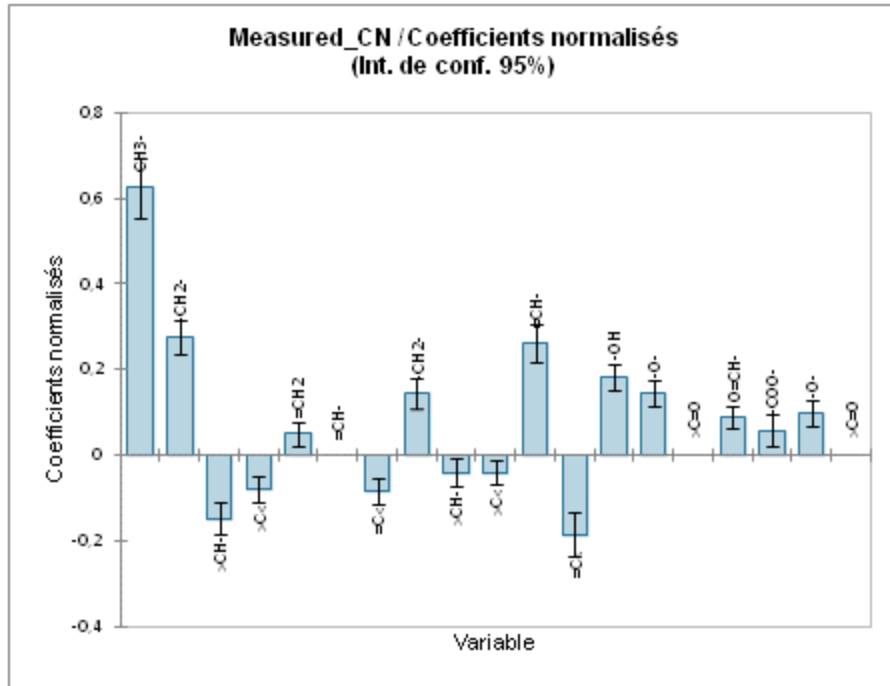
>C<	212	0	212	0,000	2,000	0,024	0,205
=CH-	212	0	212	0,000	10,000	1,094	2,072
=C<	212	0	212	0,000	4,000	0,387	0,827
-OH	212	0	212	0,000	2,000	0,250	0,445
-O-	212	0	212	0,000	5,000	0,321	0,742
>C=O	212	0	212	0,000	1,000	0,075	0,265
O=CH-	212	0	212	0,000	1,000	0,019	0,136
-COO-	212	0	212	0,000	2,000	0,434	0,624
-O-	212	0	212	0,000	2,000	0,104	0,321
>C=O	212	0	212	0,000	1,000	0,028	0,166

## ANNEXES

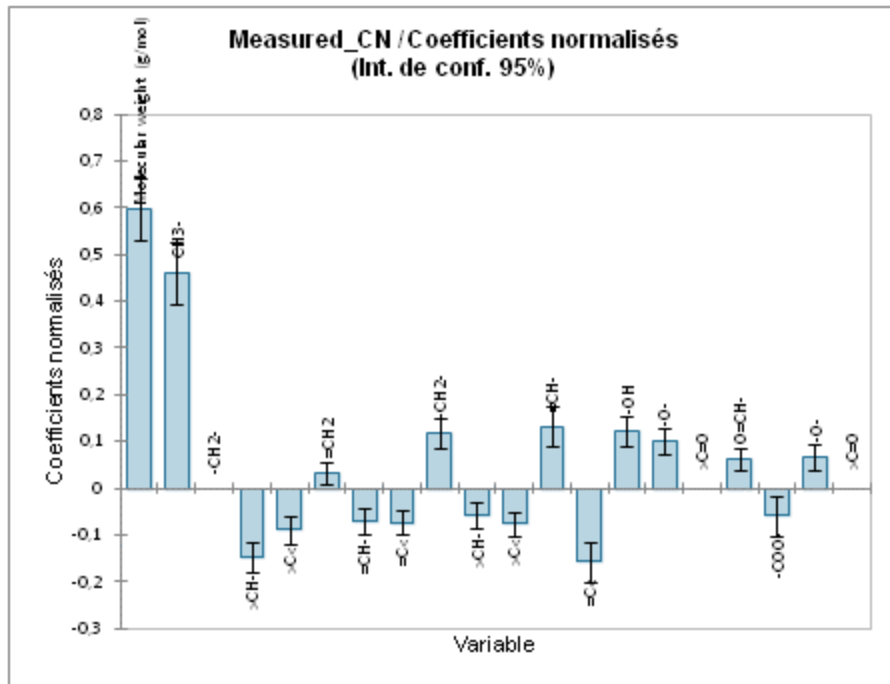
Tableau III-11 Paramètres du modèle (Measured\_CN) avec et sans MW

Source	Valeur	
	avec MW	sans MW
Constante	-105,750	-105,750
Molecular weight (g/mol)	0,459	-
CH3-	32,587	36,451
-CH2-	0,000	7,698
>CH-	-38,279	-22,392
>C<	-71,460	-56,457
=CH2	31,716	47,468
=CH-	-7,941	19,216
=C<	-36,323	-43,081
-CH2-	15,363	17,552
>CH-	-25,265	0,000
>C<	-53,137	-24,128
=CH-	8,157	0,000
=C<	-25,245	0,000
-OH	34,959	0,000
-O-	18,316	0,000
>C=O	0,000	0,000
O=CH-	64,033	0,000
-COO-	-11,451	0,000
-O-	29,072	0,000
>C=O	0,000	0,000

## ANNEXES



Le graphique III-1 des coefficients normalisés sans MW



Le graphique III-1 des coefficients normalisés avec MW



## ANNEXES

### Les résultats de tous les composés :

Tableau III-1 Statistique descriptives de base de données sans la masse moléculaire

Variable	Observations	Obs. avec données manquantes	Obs. sans données manquantes	Minimum	Maximum	Moyenne	Ecart-type
Measured_C							
N	320	0	320	-20,000	170,000	38,860	29,150
CH3-	320	0	320	0,000	9,000	2,084	1,275
-CH2-	320	0	320	0,000	23,000	4,966	5,343
>CH-	320	0	320	0,000	3,000	0,313	0,655
>C<	320	0	320	0,000	3,000	0,081	0,316
=CH2	320	0	320	0,000	2,000	0,069	0,266
=CH-	320	0	320	0,000	6,000	0,366	1,083
=C<	320	0	320	0,000	3,000	0,084	0,407
-CH2-	320	0	320	0,000	10,000	0,791	1,782
>CH-	320	0	320	0,000	5,000	0,259	0,742
>C<	320	0	320	0,000	2,000	0,044	0,247
=CH-	320	0	320	0,000	10,000	0,659	1,619
=C<	320	0	320	0,000	4,000	0,269	0,715
-OH	320	0	320	0,000	2,000	0,172	0,394
-O-	320	0	320	0,000	6,000	0,256	0,766
>C=O	320	0	320	0,000	1,000	0,028	0,166
O=CH-	320	0	320	0,000	1,000	0,019	0,136
-COO-	320	0	320	0,000	2,000	0,275	0,548
-O-	320	0	320	0,000	2,000	0,066	0,260
>C=O	320	0	320	0,000	1,000	0,013	0,111

Tableau III-1 Statistique descriptives de base de données sans la masse moléculaire

Variable	Observations	Obs. avec données manquantes	Obs. sans données manquantes	Minimum	Maximum	Moyenne	Ecart-type
Measured_CN	320	0	320	-20,000	170,000	38,610	29,584
MW (g/mol)	320	0	320	44,094	424,728	161,823	73,082
CH3-	320	0	320	0,000	7,000	2,028	1,238
-CH2-	320	0	320	0,000	25,000	4,953	5,435
>CH-	320	0	320	0,000	3,000	0,281	0,640
>C<	320	0	320	0,000	2,000	0,069	0,266
=CH2	320	0	320	0,000	2,000	0,091	0,319
=CH-	320	0	320	0,000	6,000	0,319	0,975
=C<	320	0	320	0,000	3,000	0,078	0,367
-CH2-	320	0	320	0,000	10,000	0,803	1,818
>CH-	320	0	320	0,000	4,000	0,253	0,691
>C<	320	0	320	0,000	2,000	0,034	0,199
=CH-	320	0	320	0,000	10,000	0,766	1,781

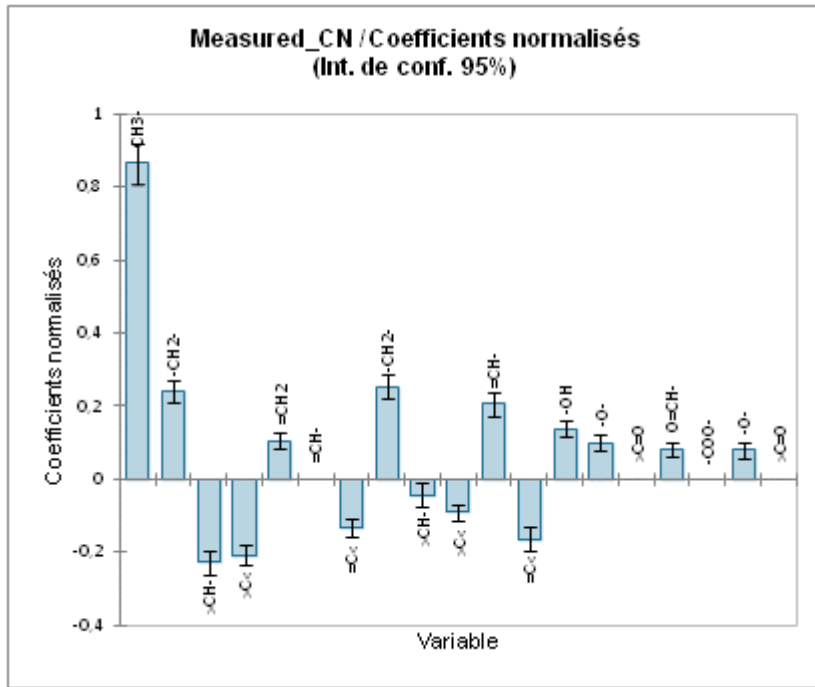
## ANNEXES

=C<	320	0	320	0,000	4,000	0,309	0,768
-OH	320	0	320	0,000	2,000	0,178	0,391
-O-	320	0	320	0,000	6,000	0,250	0,726
>C=O	320	0	320	0,000	1,000	0,034	0,182
O=CH-	320	0	320	0,000	1,000	0,019	0,136
-COO-	320	0	320	0,000	2,000	0,275	0,542
-O-	320	0	320	0,000	1,000	0,053	0,225
>C=O	320	0	320	0,000	1,000	0,019	0,136

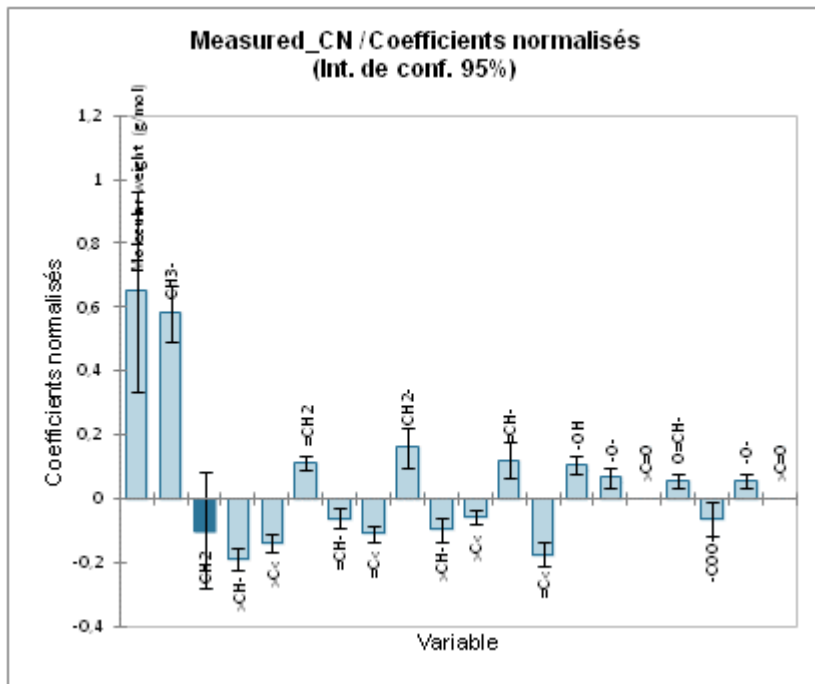
**Tableau III-11** Paramètres du modèle (Measured\_CN) avec et sans MW

Source	Valeur	
	avec MW	sans MW
Constante	-145,000	-145,000
Molecular weight (g/mol)	0,681	-
CH3-	45,475	65,999
-CH2-	-2,510	6,181
>CH-	-50,470	-58,599
>C<	-94,829	-119,068
=CH2	62,966	71,604
=CH-	-11,183	0,000
=C<	-54,561	-59,920
-CH2-	15,010	24,119
>CH-	-24,526	-10,454
>C<	-53,840	-67,487
=CH-	11,667	21,999
=C<	-39,066	-40,010
-OH	45,825	59,398
-O-	15,952	22,648
>C=O	0,000	0,000
O=CH-	75,459	109,253
-COO-	-20,099	0,000
-O-	45,475	55,579
>C=O	0,000	0,000

## ANNEXES



Le graphique III-1 des coefficients normalisés sans MW



Le graphique III-1 des coefficients normalisés avec MW

## ANNEXES

### الملخص:

لتقييم جودة الوقود قبل استعماله لتحريك المركبات (تشغيل المحركات) ولتفادي خسارة المحرك, وجب علينا دراسة الخصائص الكيميائية والفيزيائية لتحسين جودة الوقود وذلك بتحديد عدد اسيتان (المحركات الديزل) وعدد الألكانات (في محركات البنزين).

ولأجل التقييم السريع والغير مكلف اعتمدنا في دراستنا على النمذجة باستعمال برنامج XLSTAT حيث قدمنا مجموعة من ارقام السيتان المستعملة في المركبات الكيميائية النقية, وقد بلغ عدد المركبات النقية المندرجة في قاعدة البيانات المجمعة و المعدل عليهم 474 مركب نقي, ومن هذا المنبر نذكر ان الهدف من هذا العمل هو تطوير تقنية التنبؤ برقم السيتان على اساس المجموعات المساهمة والخصائص الفيزيائية/الكيميائية.

### Résumé :

Pour évaluer la qualité du carburant avant de l'utiliser pour déplacer des véhicules (roulant le moteur de la voiture) et pour éviter les pertes de moteur, il faut étudier les propriétés chimiques et physiques pour améliorer la qualité du carburant en déterminant le nombre de cétane (moteurs diesel) et le nombre d'octane (dans les moteurs à essence).

Pour une évaluation rapide et peu coûteuse, nous nous sommes appuyés sur notre étude de modélisation à l'aide du programme XLSTAT, où nous avons fourni un ensemble d'indices de cétane utilisés dans les composés chimiques purs, et le nombre de composés purs inclus dans la base de données collectée a atteint 474 composés purs, et à partir de cette plate-forme, nous mentionnons que le but de ce travail est de développer une technique de prédiction de l'indice de cétane basée sur les groupes contributeurs et les propriétés physiques et chimiques.

### Abstract:

To assess the quality of fuel before using it to move vehicles (running engines) and to avoid engine loss, We have to study the chemical and physical properties to improve fuel quality by determining the number of cetane (diesel engines) and the number of octane (in gasoline engines).

For a quick and inexpensive evaluation, we relied in our study on modeling through using the XLSTAT program, where we provided a set of cetane numbers used in pure chemical compounds, the number of pure compounds included in the collected and modified database reached 474 pure compounds. From this platform, we mention that the aim of this work is to develop a technique for predicting the cetane number based on the contributing groups and physical/chemical properties.