



جامعة بجاية
Tasdawit n'Bgayet
Université de Béjaïa

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Abderrahmane Mira-Béjaïa

Faculté de Technologie

Département de Génie des Procédés

Mémoire Présenté par :

M^r MANSOUR Khaled

Pour l'obtention du diplôme de MAGISTER

Filière : Génie des Procédés

Option : Génie chimique

Thème :

Etude des Techniques de Prédiction des
Propriétés Physiques et Thermodynamiques des
Systèmes Organiques. Application aux problèmes
d'intérêt pétrolier.

Soutenu le 20/12/2014

Devant le jury composé de :

M ^r BELHAMEL Kamel	Prof, U.A.M.Bejaia	Président.
M ^r SIFAOUI Hocine	MCA, U.A.M.Bejaia	Examineur.
M ^r BELABBAS Imad	MCA, U.A.M.Bejaia	Examineur.
M ^r KORICHI Mourad	MCA, U.K.M.Ouargla	Rapporteur.

Année Universitaire 2014/2015



Remerciements

Mes remerciements s'adressent premièrement et avant tout au Dieu tout puissant, qui m'a aidé à réaliser ce travail, et pour sa grâce tout le long de ma vie.

*Je tiens à exprimer ma profonde reconnaissance et mes sincères remerciements à mon encadreur **Mr M. Korichi** pour son aide, ces conseils, sa disponibilité et son sérieux.*

*Je tiens à remercier profondément **Mr K. BELHAMEL**, Professeur à l'Université de Béjaïa pour le plaisir qu'il me fait en acceptant de présider mon jury de soutenance.*

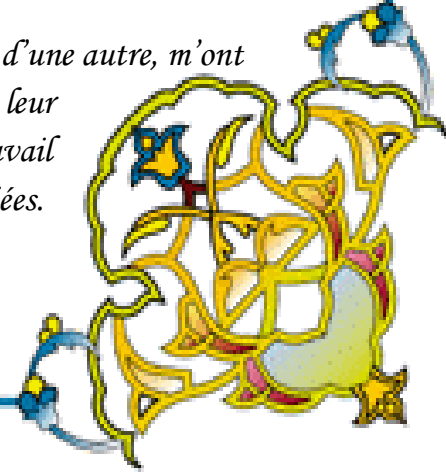

*Je remercie vivement **Mr H.SIFAOU** et **Mr I. BELABBAS** qui ont bien voulu examiner ce travail, malgré leurs nombreuses occupations.*

*Je tiens à exprimer ma profonde reconnaissance et mes sincères remerciements à **Mr M.BERKANI** pour son aide.*

Je remercie également tout le personnel du laboratoire de génie des procédés à l'université de Kasdi Merbehi Ouargla.

A mes chers parents pour leurs sacrifices et encouragements, durant toute la période de mes études, et à toute ma famille.

Que toutes les personnes qui, d'une façon ou d'une autre, m'ont aidé par leurs conseils, leur soutien et leur présence tout au long de ce travail en soient ici vivement remerciées.





Dédicaces



J'ai le grand honneur de dédier ce modeste travail à :

A mon cher père ;

A ma chère mère ;

A mes très chers sœurs et frères ;

*A toutes ma famille, mes tantes et oncles, mes cousins et mes
cousines ;*

A mon encadreur Mr M. Korichi ;

*A mes collègues de la promotion 2011/2012 de génie des
procédés*

A tous les enseignants du département de génie des procédés

A tous ceux qui pensent à Khaled

Table des matières

Introduction générale.....	1
----------------------------	---

Chapitre 01 : Etude des méthodes de contribution de groupes

1.1. Introduction	3
1.2. Méthodes de contribution de groupes d'ordre simples	4
1.2.1. Méthode de <i>Lydersen (1955)</i>	4
1.2.2. Méthode de <i>Joback et Reid (1987)</i>	5
1.2.3. Méthode de <i>Marrero et Pardillo (1999)</i>	8
1.2.4. Méthode de <i>Wen et Quiang (2001)</i>	11
1.3. Méthodes de contribution de groupes multi-ordre	15
1.3.1. Méthode de <i>Constantinou et Gani (1994)</i>	15
1.3.2. Méthode de <i>Marrero et Gani (2001)</i>	20
1.4. Méthode de contribution de la position de groupes	24
1.4.1. Méthode de <i>Wang, et al., (2008)</i>	24
1.5. Méthodes basées sur la combinaison	27
1.5.1. Méthode de <i>Wilson et Jasperson (1996)</i>	27
1.5.2. Méthode de <i>Cordes et Rarey (2002)</i>	29
1.5.3. Méthode de <i>Nannoolal, et al., (2004)</i>	31
1.6. Autres méthodes	34
1.7. Discussion des méthodes d'estimation.....	34

Chapitre 02 : Méthodologie du développement de la méthode proposée

2.1. Introduction	37
2.2. Approche choisie.....	38
2.3. Définition des groupes structurels.....	40
2.3.1. Groupes principaux.....	40
2.3.2. Groupes de correction	42

2.4.	Base de données	43
2.4.1.	Collection des données	43
2.4.2.	Structure de la base de données	45
2.4.3.	Vérification de la base de données.....	46
2.4.4.	Propriétés de la base de données.....	46
2.5.	Développement des modèles	47
2.5.1.	Température d'ébullition.....	47
2.5.2.	Modèles proposés des autres propriétés.....	51
2.6.	Ajustement du modèle.....	51
2.6.1.	Méthodes de régression.....	52
2.6.1.1.	Régression linéaire	52
2.6.1.2.	Régression non-linéaire	53
2.6.2.	Outils informatiques.....	53
2.6.2.1.	Transformation de la base de données.....	54
2.6.2.2.	Création d'un fichier en R	55
2.6.2.3.	Obtention des contributions de groupes	56
2.6.3.	Validation des modèles	57
2.6.3.1.	R^2 et R^2 ajusté.....	57
2.6.3.2.	Analyse des résidus	58
a.	Analyse graphique des résidus	58
b.	Analyse quantitative des résidus.....	58

Chapitre 03 : Résultats et discussion

3.1.	Introduction	59
3.2.	Température d'ébullition.....	60
3.3.	Température critique	64
3.4.	Pression critique	67
3.5.	Volume critique.....	70

3.6.	Température de fusion.....	73
3.7.	Enthalpie de fusion.....	76
3.8.	Enthalpie standard de formation (phase gazeuse)	79
	Conclusion générale	82
	Bibliographie	83
	Annexe 01 : Contributions des interactions de groupes	86
	Annexe 02 : Exemples.....	92
	Annexe 03 : Programme sur R	103

Liste des figures

Figure 1.1 : Conjugué dominant, conjugué récessif généré et l'opérateur de conjugaison correspondant.	16
Figure 1.2 : Détermination de la position d'un groupe dans la molécule.	25
Figure 1.3 : Schéma montrant la distribution des méthodes d'estimation en fonction du type et des années.	36
Figure 2.1 : Schéma résumant la démarche du développement des modèles proposés.	38
Figure 2.2 : Page de site internet de la base de données NIST. (19/09/2014).....	44
Figure 2.3 : Type de la recherche dans la base de données NIST. (19/09/2014).....	44
Figure 2.4 : Schéma illustrant le processus de la régression.	53
Figure 2.5: Transformation de la base de données d'une feuille XLXS à une feuille CSV.....	54
Figure 2.6 : Importation de la base de données dans le logiciel R.	55
Figure 3.1 : Comparaison entre les valeurs expérimentales et estimées de (T_{eb}).	61
Figure 3.2 : Dispersion des résidus de la (T_{eb}) en fonction des (T_{eb}) expérimentales.	61
Figure 3.3 : Dispersion des résidus de la (T_{eb}) en fonction des (T_{eb}) estimées.	62
Figure 3.4 : Comparaison entre les valeurs expérimentales et estimées de (T_c).	65
Figure 3.5 : Dispersion des résidus de la (T_c) en fonction des (T_c) expérimentales.	65
Figure 3.6 : Dispersion des résidus de la (T_c) en fonction des (T_c) estimées.	65
Figure 3.7 : Comparaison entre les valeurs expérimentales et estimées de (P_c).	68
Figure 3.8 : Dispersion des résidus de la (P_c) en fonction des (P_c) expérimentales.	68
Figure 3.9 : Dispersion des résidus de la (P_c) en fonction des (P_c) estimées.	68

Figure 3.10 : Comparaison entre les valeurs expérimentales et estimées du (V_c).	71
Figure 3.11 : Dispersion des résidus du (V_c) en fonction des (V_c) expérimentales.	71
Figure 3.12 : Dispersion des résidus du (V_c) en fonction des (V_c) estimées.....	71
Figure 3.13 : Comparaison entre les valeurs expérimentales et estimées de (T_f).	74
Figure 3.14 : Dispersion des résidus de la (T_f) en fonction des (T_f) expérimentales.	74
Figure 3.15 : Dispersion des résidus de la (T_f) en fonction des (T_f) estimées.	74
Figure 3.16 : Comparaison entre les valeurs expérimentales et estimées de (ΔH_{fus}).	77
Figure 3.17 : Dispersion des résidus de la (ΔH_{fus}) en fonction des (ΔH_{fus}) expérimentales.....	77
Figure 3.18 : Dispersion des résidus de la (ΔH_{fus}) en fonction des (ΔH_{fus}) estimées.	77
Figure 3.19 : Comparaison entre les valeurs expérimentales et estimées de ($\Delta_f H^\circ_{gaz}$).	80
Figure 3.20 : Dispersion des résidus de la ($\Delta_f H^\circ_{gaz}$) en fonction des ($\Delta_f H^\circ_{gaz}$) expérimentales.	80
Figure 3.21 : Dispersion des résidus de la ($\Delta_f H^\circ_{gaz}$) en fonction des ($\Delta_f H^\circ_{gaz}$) estimées.	80

Liste des tableaux

Tableau 1.1 : Application de la méthode de <i>Lydersen (1955)</i>	5
Tableau 1.2 : Nombres des valeurs expérimentales des propriétés utilisées dans la méthode de <i>Joback et Reid (1987)</i>	6
Tableau 1.3 : Application de la méthode de <i>Joback et Reid (1987)</i>	7
Tableau 1.4 : Application de la méthode de <i>Marrero et Pardillo (1999)</i>	10
Tableau 1.5 : Application de la méthode de <i>Wen et Quiang (2001)</i>	13
Tableau 1.6 : Application du deuxième modèle de la méthode de <i>Wen et Quiang (2001)</i>	14
Tableau 1.7 : Application de la méthode de <i>Constantinou et Gani (1994)</i>	19
Tableau 1.8 : Nombre de propriétés utilisées dans la méthode de <i>Marrero et Gani (2001)</i>	21
Tableau 1.9 : Application de la méthode de <i>Marrero et Gani (2001)</i>	23
Tableau 1.10 : Application de la méthode de <i>Wang, et al., (2008)</i>	26
Tableau 1.11 : Application de la méthode de <i>Wilson et Jasperson (1996)</i>	28
Tableau 1.12 : Application de la méthode de <i>Cordes et Rarey (2002)</i>	30
Tableau 1.13 : Application de la méthode de <i>Nannoolal, et al., (2004)</i>	33
Tableau 2.1: Comparaison de la température d'ébullition des composés en fonction de leur type de liaison.	38
Tableau 2.2 : Température d'ébullition d'un aldéhyde, acétone, alcool et un éther.	40
Tableau 2.3: Groupes utilisés pour la génération des groupes structuraux.	42
Tableau 2.4 : Nombre des valeurs expérimentales des propriétés dans la base de données. ...	46
Tableau 2.5: Résultats de la régression des trois modèles.	50

Tableau 3.1 : Performances du modèle de la CIG pour la (T_{eb}) pour les différentes classes...	60
Tableau 3.2 : Comparaison entre les performances du modèle de la CIG et des modèles précédents pour la (T_{eb}).	63
Tableau 3.3 : Performances du modèle de la CIG pour la (T_c) pour les différentes classes.	64
Tableau 3.4 : Comparaison entre les performances du modèle de la CIG et des modèles précédents pour la (T_c).	66
Tableau 3.5 : Performances du modèle de la CIG pour la (P_c) pour les différentes classes.	67
Tableau 3.6 : Comparaison entre les performances du modèle de la CIG et des modèles précédents pour la (P_c).	69
Tableau 3.7 : Performances du modèle de la CIG pour le (V_c) pour les différentes classes....	70
Tableau 3.8 : Comparaison entre les performances du modèle de la CIG et des modèles précédents pour le (V_c).	72
Tableau 3.9 : Performances du modèle de la CIG pour la (T_f) pour des différentes classes....	73
Tableau 3.10 : Comparaison entre les performances du modèle de la CIG et des modèles précédents pour la (T_f).	75
Tableau 3.11 : Performances du modèle de la CIG pour la (ΔH_{fus}) pour les différentes classes.	76
Tableau 3.12 : Comparaison entre les performances du modèle de la CIG et le modèle de la CG de 3 ^{eme} ordre pour la (ΔH_{fus}).	78
Tableau 3.13 : Performances du modèle de la CIG pour la ($\Delta_f H^\circ_{gaz}$) pour les différentes classes.....	79
Tableau 3.14 : Comparaison entre les performances du modèle de la CIG et le modèle de la CG de 3 ^{eme} ordre pour la ($\Delta_f H^\circ_{gaz}$).	81

Liste des abréviations

Propriétés

T_{eb}	température d'ébullition
T_c	température critique
P_c	pression critique
V_c	volume critique
T_f	température de fusion
PE	point éclair
ΔH°_{for}	enthalpie standard de formation à 298 K
ΔH_v	enthalpie de vaporisation à 298 K
ΔG_{for}	énergie standard de Gibbs à 298 K
ΔH_{fus}	enthalpie de fusion
$\Delta_f H^\circ_{(gaz)}$	enthalpie standard de formation (phase gazeuse)
ΔS_v	entropie de vaporisation
M	masse molaire
N	nombre d'atome dans la molécule sans l'hydrogène
n_A	nombre d'atome total dans la molécule
exp	expérimentale
est	estimée

Groupes structurels

C_i	contribution de groupe de type i
N_i	nombre d'apparition de groupe de type i
D_k	contribution de groupe de type k
M_k	nombre d'apparition de groupe de type k
D_j	contribution de groupe de type j
M_j	nombre d'apparition de groupe de type j
E_k	contribution d'un groupe de type k
O_k	nombre d'apparition de groupe de type k
A_i	contribution de groupes de type i
A_j	contribution de groupes de type j
n_r	nombre de cycle présent dans la molécule
CI	contribution de l'interaction
CIG	contribution des interactions de groupes
a,b,c,...etc	paramètres d'ajustement

Base de données

Cas	the Chemical Abstracts Service Registry Number
SDF	Standard Data Format
CAPEC	Common Attack Pattern Enumeration and Classification
DDB	Dortmund Data Bank
InChI	International Chemical Identification
SMILES	Simplified Molecular-Input Line-Entry Specification

Paramètres statistiques

PEMA	pourcentage d'erreur moyenne absolu
EMA	erreur moyenne absolue
DS	déviation standard
R^2	coefficient de la détermination
R^2 ajusté	coefficient de la détermination ajusté

Introduction générale

Introduction générale

Grâce aux nouvelles techniques de synthèse, la gamme des composés organiques ne cesse d'augmenter (plus de 91¹ millions composés organiques et inorganiques rapportés par CAS number).

Pour l'étude, la conception, la simulation, l'optimisation des procédés (chimiques, biochimiques, ou environnementaux), ainsi que la conception des produits assistée par ordinateur (CPAO), il est indispensable de connaître les propriétés (physiques, thermodynamiques, etc.) des corps purs et complexes.

Cependant, les méthodes expérimentales permettant d'obtenir les propriétés peuvent être consommatrices du temps, très coûteuses et parfois dangereuses.

Depuis longtemps *des méthodes d'estimation des propriétés* sont utilisées comme une alternative aux méthodes expérimentales, ces dernières années, les méthodes d'estimation ont connu un développement très rapide grâce à l'évolution des outils informatiques.

Les méthodes d'estimation donnent une idée approximative du comportement des produits chimiques et peuvent être suffisantes pour sélectionner des produits chimiques pour l'étude.

En se basant sur la structure moléculaire, les méthodes d'estimation des propriétés peuvent être s'appuyer sur ;

- La contribution de groupes ;
- La RQSP (relation quantitative structure-propriétés) ;
- La combinaison de ces deux.

Les méthodes basées sur les contributions de groupes sont les plus anciennes, *Lydersen (1955)* a été le premier à établir une technique permettant de prédire la température et la pression critiques des composés organiques via la contribution de groupes.

L'objectif principal de ce travail, est de développer des modèles fiables et précis qui permettent la prédiction des propriétés physiques et thermodynamiques des composés organiques, pour résoudre des problèmes dans le domaine du pétrole et du gaz naturel.

¹ <http://www.cas.org>

Pour atteindre cet objectif, on a adopté la démarche suivante :

- Choisir une approche qui représente la relation structure-propriété, et avec cette approche qu'on choisit les groupes structuraux,
- Collecter les données expérimentales des propriétés à estimer, ceci forme une base de données,
- Sélectionner des modèles pour chaque propriété, ces modèles représentent le lien entre la propriété et les groupes structuraux,
- Pour faciliter l'obtention des résultats, on doit choisir un outil informatique performant pour minimiser le temps de calcul et augmenter la précision, et la dernière étape est la validation des modèles, afin de juger les performances de nos modèles et les comparer avec celles des modèles déjà publiés.

Le présent travail, s'articule autour de trois (3) chapitres :

Dans le chapitre 01, on a décrit quelques méthodes de prédiction des propriétés physiques et thermodynamiques. L'objectif de cette description est de présenter l'approche de chaque méthode et son contexte théorique s'il existe, la base de données utilisée dans le développement, le type du modèle, le type de la méthode de régression avec laquelle ils ont obtenu les valeurs des contributions de groupes, l'algorithme d'optimisation, et à la fin de ce chapitre une petite discussion comparative de toutes ces méthodes est aussi réalisée.

Dans le chapitre 02, on a décrit la méthodologie du développement de la méthode proposée. Premièrement avec l'approche sélectionnée, cette approche est la base du choix des groupes structuraux, ces groupes sont choisis pour représenter la structure de la molécule, avec la contribution de ces groupes qu'on obtient la valeur de la propriété. Ensuite, la base de données, sa collection, sa structure et tout ce qui concerne cette base. L'étape suivante, est le développement des modèles de prédiction, en détaillant les méthodes, les outils et tous les moyens nécessaires pour l'obtention de ces modèles, et à la fin de ce chapitre on a cité les méthodes et les moyens de validation pour confirmer que ces modèles sont les meilleurs.

Dans le chapitre 03, on a illustré les résultats obtenus par les modèles proposés sous forme des tableaux et des figures, une discussion des résultats en comparant les résultats obtenus avec les résultats des autres modèles déjà publiés.

Chapitre 01 :

Etude des méthodes de contribution de groupes

1.1. Introduction

Les méthodes des contributions de groupes sont les méthodes les plus anciennes, elles permettent d'évaluer certaines propriétés des composés purs ou complexes à partir de la structure moléculaire. Les propriétés physiques, thermodynamiques, etc, en utilisant le modèle mathématique qui caractérise une méthode, égales à la somme des multiplications des contributions de groupes structuraux (atomes, interactions, groupes fonctionnels, etc.) constituant la molécule, par leur nombre de fois d'apparition dans la molécule.

Historiquement, *Guldberg (1890)* a été le premier qui a constaté que la température normale d'ébullition des liquides non-associatifs (eau-éthylène glycol) est reliée à la température critique par l'équation 1.1, connue aussi par *la règle de Guldberg* :

$$T_{eb} = \frac{2}{3} T_c \quad (1.1)$$

Donc, cette relation a été la première équation permettant l'obtention de la température d'ébullition (*Eagleson, 1993*).

Riedel (1949), *Vowles (1951)* et *Lydersen (1955)* ont proposé une modification sur l'équation 1.1 pour prédire la température critique, l'équation 1.2 :

$$T_c = \frac{T_{eb}}{\theta} \quad (1.2)$$

La valeur de θ varie d'une gamme de composés à une autre. Sa valeur peut être obtenue par la sommation des contributions de groupes structuraux formant la molécule.

Riedel (1949) a proposé 22 groupes structuraux simples, l'équation 1.3 a été utilisée pour calculer θ afin d'estimer la température critique en utilisant l'équation 1.2. Pour la pression critique l'équation 1.4 a été proposée par *Riedel (1949)* en combinant la contribution des atomes et de groupes :

$$\theta = 0,574 + \sum_i N_i C_i \quad (1.3)$$

$$P_c = \frac{M}{(0,33 - \sum_i N_i C_i)^2} \quad (1.4)$$

Où

C_i est la contribution de groupe de type i , N_i est le nombre d'apparition de ce groupe dans la molécule et M est la masse molaire.

Vowles (1951) a proposé que la valeur de θ peut être calculée en sommant les contributions des atomes (contribution atomique, ordre zéro, contribution élémentaire) (*Nannoolal, 2006*).

1.2. Méthodes de contribution de groupes d'ordre simples

1.2.1. Méthode de *Lydersen (1955)*

Cette méthode est comptée la première méthode des contributions de groupes, *Lydersen (1955)* a étendu la méthode de *Riedel (1949)* en intégrant un ensemble plus vaste de groupes via une base de données plus large, il a choisi 35 groupes structuraux pour obtenir les propriétés du point critique, ainsi, il a proposé un modèle quadratique pour l'estimation de θ , l'équation 1.5. Pour l'estimation de la température critique il a proposé l'équation 1.2, pour obtenir la pression et le volume critiques il a proposé les équations 1.6, et 1.7 respectivement :

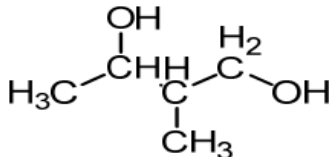
$$\theta = 0,567 + \sum_i N_i C_i - (\sum_i N_i C_i)^2 \quad (1.5)$$

$$P_c = \frac{M}{(0,34 - \sum_i N_i C_i)^2} \quad (1.6)$$

$$V_c = 0,04 + \sum_i N_i C_i \quad (1.7)$$

Lydersen (1995) a utilisé 557, 474 et 327 valeurs expérimentales de la température, la pression et le volume critiques respectivement des composés organiques pour le développement du modèle ainsi que les contributions de groupes (*Nannoolal, 2006*).

Tableau 1.1 : Application de méthode de *Lydersen (1955)*.

Température critique T _c			
<u>Nom</u> : 2-méthyl-1,3-butanediol <u>Cas Number</u> : 684-84-4 <u>Formule moléculaire</u> : C ₅ H ₁₂ O ₂ <u>La masse moléculaire</u> =104,14758 g/mol T _c (exp) = 615,15 K T _{eb} (exp) = 473,15 K		<u>Structure moléculaire</u> : 	
Méthode des contributions de groupes			
Groupe i	Nombre d'apparition N _i	Contribution C _i	N _i * C _i
-CH3	2	0,02	0,04
>CH2	1	0,02	0,02
>CH-	2	0,012	0,024
-OH	2	0,082	0,164
Σ _i N _i C _i			0,248
$T_c = \frac{T_{eb}}{0,567 + \sum_i N_i C_i - (\sum_i N_i C_i)^2} \Rightarrow T_c = \frac{473,15}{(0,567 + 0,248 - 0,061504)} = 627,939K$			
T _c (exp)= 615,15K, erreur= 627,939-615,15 = 12,789K, 2,079%			

1.2.2. Méthode de *Joback et Reid (1987)*

Joback et Reid (1987) ont réévalué le système des contributions de groupes de la méthode de *Lydersen (1955)*, ils ont ajouté des nouveaux groupes fonctionnels comme le groupe =N- (cycle) et ignorant d'autres comme les groupes (Silicium) >Si< et (Bore) >B-. 41 groupes structuraux ont été sélectionnés pour décrire la structure des molécules. La méthode peut prédire un ensemble très important des propriétés, physiques, thermodynamique, etc, le tableau 1.2 illustre ces propriétés avec le nombre des valeurs expérimentales utilisé pour le développement de cette méthode.

Tableau 1.2 : Nombres des valeurs expérimentales des propriétés utilisées dans la méthode de *Joback et Reid (1987)*.

Propriétés	Valeurs expérimentales
Température d'ébullition	438
Température critique	409
Pression critique	392
Volume critique	310

Les équations choisies pour estimer les propriétés via la méthode de *Joback et Reid (1987)* sont les suivantes :

$$T_{eb} = 198,2 + \sum_i N_i C_i \quad (1.8)$$

$$T_f = 122 + \sum_i N_i C_i \quad (1.9)$$

$$T_c = \frac{T_{eb}}{(0,584 + 0,965 \sum_i N_i C_i - (\sum_i N_i C_i)^2)} \quad (1.10)$$

$$P_c = \frac{1}{(0,113 + 0,0032n - \sum_i N_i C_i)^2} \quad (1.11)$$

$$V_c = 17,5 + \sum_i N_i C_i \quad (1.12)$$

$$\Delta G_{f, 298,15K}^\circ = 53,88 + \sum_i N_i C_i \text{ (kJ/mol)} \quad (1.13)$$

$$\Delta H_{f, 298,15K}^\circ = 68,29 + \sum_i N_i C_i \text{ (kJ/mol)} \quad (1.14)$$

$$C_p^\circ(T) = \left[\sum_i N_{ai} C_{ai} - 37,93 \right] + \left[\sum_i N_{bi} C_{bi} + 0,210 \right] T + \left[\sum_i N_{ci} C_{ci} - 3,91 \times 10^{-4} \right] T^2 + \left[\sum_i N_{di} C_{di} + 2,06 \times 10^{-7} \right] T^3 \quad (1.15)$$

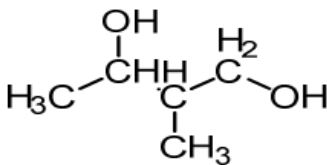
(J/mol.K)

Pour le modèle de la température critique, la température normal d'ébullition est nécessaire, on utilise des valeurs expérimentales s'ils sont disponibles, si non, on peut les obtenir à partir d'une méthode de prédiction, par exemple, équation 1.8 (*Poling, et al., 2001*).

Joback et Reid (1987) ont employé la régression linéaire, et leur fonction objective à été la minimisation de la somme des carrée des erreurs absolues, malgré que cette méthode montre une erreur très importante. Cette méthode de régression n'est pas utile, surtout avec les constantes critiques.

Le nombre bas de groupes structuraux utilisé (41) offre la simplicité à la méthode d'une part, et de l'autre part la simplicité de la représentation structurale des molécules, alors, la distinction entre les isomères sera irréaliste, par exemple (*2-méthylhexane et 3-méthylhexane*) ont la même fragmentation, donc, ils y auront les mêmes valeurs des propriétés); c'est ce qui conduit à une mauvaise description des effets structuraux de la molécule et ceci explique la faible précision, ce qu'il a amené à sa disparition.

Tableau 1.3 : Application de la méthode *Joback et Reid (1987)*.

Température normale d'ébullition T _{eb}			
<u>Nom</u> : 2-méthyl-1,3-butanediol <u>Cas Number</u> : 684-84-4 <u>Formule moléculaire</u> : C ₅ H ₁₂ O ₂ <u>La masse moléculaire</u> =104,14758 g/mol T _{eb} (exp) = 473,15 K		<u>Structure moléculaire</u> : 	
Méthode des contributions d'atomes et de groupes			
Groupe i	Nombre d'apparition N _i	Contribution C _i	N _i * C _i
-CH3	2	23,58	47,16
>CH2	1	22,88	22,88
>CH-	2	21,74	43,48
-OH	2	92,88	185,76
Σ _i N _i C _i			299,28
T _{eb} = 198,2 + Σ _i N _i C _i => T _{eb} = 198,2 + 299,28 = 497,48K			
T _{eb} (exp)= 421,15K, erreur = 497,48-421,15 = 76,33K, 18,12%			

1.2.3. Méthode de Marrero et Pardillo (1999)

En se basant sur la nouvelle idée de *Pardillo-Fontdevila et Gonzalez-Rubio (1997)* la *contribution des interactions de groupes*, qui propose la contribution des interactions entre les groupes de la molécule au lieu de la contribution des groupes simples qui constituent la molécule, *Marrero et Pardillo (1999)* ont proposé une nouvelle technique pour l'estimation de certaines propriétés des composés organiques purs en utilisant la structure de la molécule.

Les propriétés selon *Marrero et Pardillo (1999)* d'un composé, sont considérées comme des fonctions des paramètres structuraux dépendants, qui sont de ce fait, déterminées par la sommation des contributions des interactions de groupes multipliées par son nombre d'apparition dans la molécule. Ils ont sélectionné un ensemble de 39 groupes simples pour générer les groupes structuraux pour permettre le traitement d'une vaste gamme des composés organiques, ces groupes sont les mêmes déjà utilisés par *Joback et Reid (1987)* et par *Lydersen (1955)* auparavant, avec l'omission de =NH et =N- non-cyclique.

Pour comparer la méthode des contributions de groupes et des contributions des interactions de groupes, en utilisant le même modèle employé par *Joback et Reid (1987)* et via un ensemble de 507 valeurs expérimentales de la température d'ébullition, ils ont trouvé des nouvelles valeurs aux contributions de groupes, l'équation 1.16 caractérise le même modèle utilisé par *Joback et Reid (1987)* mais cette fois la constante égale à 218,13 au lieu de 198,2. Les équations utilisées pour la prédiction de la température d'ébullition et les constants critiques sont les suivantes :

$$T_{eb} = 218,13 + \sum_i N_i C_i \quad (1.16)$$

$$T_{eb} = \frac{\sum_i N_i C_i}{M^{0,404}} + 156 \quad (1.17)$$

$$T_c = \frac{T_{eb}}{(0,5851 - 0,9286 \sum_i N_i C_i - (\sum_i N_i C_i)^2)} \quad (1.18)$$

$$P_c = \frac{1}{(0,1285 - 0,0059 n_A - \sum_i N_i C_i)^2} \quad (1.19)$$

$$V_c = 25,1 + \sum_i N_i C_i \quad (1.20)$$

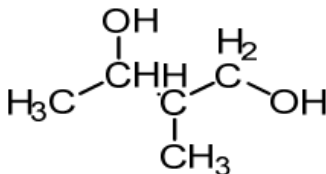
Marrero et Pardillo (1999) ont employé l'analyse des moindres carrées pour déterminer les contributions de groupes simples et les contributions des interactions de groupes, ils ont utilisés 491, 445 et 289 valeurs pour la température, la pression et le volume critiques respectivement.

Concernant la régression linéaire, ils ont choisi la procédure de la décomposition en valeurs singulières (*Forsythe, et al., 1977*) comme algorithme d'optimisation, et pour la régression non linéaire ils ont choisi celui de *Levenberg-Marquardt*, et la fonction objective est de minimiser la somme des carrés des différences entre les valeurs expérimentales et estimées (*Marrero-Morejon, et al., 1999*).

La méthode a offert des bons résultats concernant la précision et la distinction entre les isomères en comparaison avec la méthode de *Joback et Reid (1987)*, mais il existe quelques molécules la méthode ne fait pas la distinction entre eux, comme (*3-méthylheptane* et *4-méthylheptane*). Le nombre de groupes structuels (174) est très important en comparaison avec le nombre de valeurs expérimentales utilisé (507 de T_{eb}), ce qui diminue la précision.

Un autre résultat au petit nombre de composés organiques utilisé pour générer les groupes structuels est que la méthode n'arrive pas à faire la fragmentation de quelques molécule en raison de l'absence des groupes qui forme la molécule, comme l'interaction entre $>CH-$ et $\equiv C-$ via une seule liaison, donc le manque de cette interaction ne permet pas de fragmenter (*3-Méthyle-1-butyle*), (*Nannoolal, 2006*) a rapporté que la méthode arrive à fragmenter 1665 composés parmi 2800 composés disponible dans la base de données de Dortmund (*Nannoolal, 2006*).

Tableau 1.4 : Application de la méthode de Marrero et Pardillo (1999).

Température normale d'ébullition T _{eb}			
<u>Nom</u> : 2-méthyl-1,3-butanediol <u>Cas Number</u> : 684-84-4 <u>Formule moléculaire</u> : C ₅ H ₁₂ O ₂ <u>La masse moléculaire</u> =104,14758 g/mol T _{eb} (exp) = 473,15 K		<u>Structure moléculaire</u> : 	
Méthode des contributions des interactions de groupes			
Groupe i	Nombre d'apparition N _i	Contribution C _i	N _i * C _i
-CH3 et >CH-	2	194,27	388,54
>CH2 et >CH-	1	244,14	244,14
>CH2 et -OH	1	673,24	673,24
>CH- et >CH-	1	291,41	291,41
>CH- et -OH	1	585,99	585,99
Σ _i N _i C _i			2183,32
$T_{eb} = \frac{\sum_i N_i C_i}{M^{0,404}} + 156 \Rightarrow T_{eb} = \frac{2183,32}{104,147^{0,404}} + 156 = 490,185K$			
T _{eb} (exp)= 473,15 K, erreur = 490,185-473,15 = 17,035 K, 3,6%			

1.2.4. Méthode de *Wen et Quiang (2001)*

Le but de cette méthode a été de développer une nouvelle technique des contributions de groupes qui peut fournir une bonne précision avec un nombre de groupes structuraux un peu bas.

Pour atteindre ce but, une nouvelle approche a été proposée, au lieu de compter l'interaction de deux groupes fonctionnels, on compte l'interaction entre un groupe et l'atome adjacent, c'est-à-dire que le groupe structural est constitué d'un pair de groupe-atome adjacent.

Dans la méthode de CIG celle de *Marrero et Pardillo (1999)*, pour le groupe $-\text{CH}_3$, si on compte les groupes adjacents de carbone on aura, $-\text{CH}_3$, $>\text{CH}_2$, $>\text{CH}-$ et $>\text{C}<$, mais si on compte les atomes adjacents seulement, on aura un atome, c'est le carbone $>\text{C}<$, donc, au lieu de quatre 4 interactions on aura une. Ainsi, cette méthode contient des groupes pour la correction des paires groupe-atome adjacent, ces groupes constitués de l'oxygène, l'azote et du soufre et ne contiennent pas de carbone.

Le modèle général de cette méthode est le suivant :

$$\mathbf{F} = \mathbf{a} + \sum_{ij} \mathbf{N}_{ij} \mathbf{C}_{ij} + \sum_k \mathbf{M}_k \mathbf{D}_k$$

Où :

- \mathbf{C}_{ij} est la contribution de groupe structural contient un pair groupe-atome adjacent et \mathbf{N}_{ij} est le nombre d'apparition de ce groupe dans la molécule.
- \mathbf{D}_k est la contribution d'un groupe qui ne contient pas de carbone (groupe de correction) et \mathbf{M}_k le nombre d'apparition de ce groupe dans la molécule.

Dans cette méthode, deux modèles ont été proposés pour la prédiction de la température critique, l'un utilise la température d'ébullition comme entrée, et l'autre n'utilise pas la température d'ébullition, les équations caractéristiques des modèles de la température, la pression et le volume critiques sont les suivantes :

$$T_c^{2,747} \times 10^{-6} = 4,72 + \sum_{ij} N_{ij} C_{ij} + \sum_k M_k D_k \quad (1.21)$$

$$\left(\frac{T_{eb}}{T_c - T_{eb}}\right)^{-0,5} \times 10^2 = 127,754 + \sum_{ij} N_{ij} C'_{ij} + \sum_k M_k D'_k \quad (1.22)$$

$$P_c^{-0,5} \times 10^2 = 37,293 + \sum_{ij} N_{ij} C_{ij} + \sum_k M_k D_k \quad (1.23)$$

$$V_c = -27,04 + \sum_{ij} N_{ij} C_{ij} + \sum_k M_k D_k \quad (1.24)$$

✚ C'ij et D'ij réfèrent aux contributions de groupes de l'équation de la température critique avec température d'ébullition comme entrée.

Malgré que le nombre de paramètres de cette méthode a diminué et la précision a augmenté en comparaison avec la méthode de *Marrero et Pardillo (1999)*, mais le nombre de paramètres est resté un peu grand, 127 paramètres structuraux groupe-atome adjacent contenant le carbone, et les paramètres structuraux sans carbone sont 13, ainsi, le nombre de composés utilisés dans le développement de la méthode est un peu petit, 471, 408 et 275 et valeur expérimentales de la température, pression et volume critiques qui ont été employées (*Wen, et al., 2001*).

Tableau 1.5 : Application de la méthode de *Wen et Quiang (2001)*.

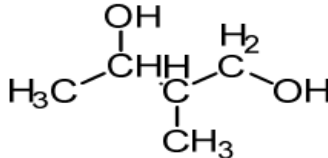
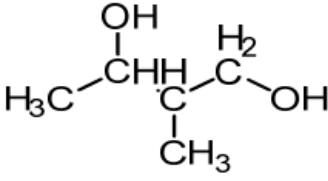
Température critique T _c			
<u>Nom</u> : 2-méthyl-1,3-butanediol <u>Cas Number</u> : 684-84-4 <u>Formule moléculaire</u> : C ₅ H ₁₂ O ₂ <u>La masse moléculaire</u> =104,14758 g/mol <u>T_c (exp)</u> = 615,15 K <u>T_{éb} (exp)</u> = 473,15 K		<u>Structure moléculaire</u> : 	
Méthode des contributions des paires groupe-atome adjacent			
Groupe i	Nombre d'apparition N _i	Contribution C _i	N _i * C _i
-CH ₃ et >C<	2	-1,1863	-2,3726
>CH ₂ et >C<	1	2,2116	2,2116
>CH ₂ et -O-	1	-0,0296	-0,0296
>CH- et >C<	5	2,76	13,8
>CH- et -O-	1	0,9548	0,9548
Σ _{ij} N _{ij} C _{ij}			14,5642
-OH	2	12,1779	24,3558
Σ _k M _k D _k			24,3558
$\left(\frac{T_{éb}}{T_c - T_{éb}}\right)^{-0,5} \times 10^2 = 127,754 + \sum_{ij} N_{ij} C'_{ij} + \sum_k M_k D'_k \Rightarrow$ $\left(\frac{473.15}{T_c - 473.15}\right)^{-0.5} \times 10^2 = 127,754 + 14,5642 + 24,3558 = 643,469K$ $T_c(\text{exp})= 615,15K, \text{ erreur} = 643,469-615,15 = 28,319 K, \quad 4,6\%$			

Tableau 1.6 : Application du deuxième modèle de la méthode de *Wen et Quiang (2001)*.

Température critique T _c			
<u>Nom</u> : 2-méthyl-1,3-butanediol <u>Cas Number</u> : 684-84-4 <u>Formule moléculaire</u> : C ₅ H ₁₂ O ₂ <u>La masse moléculaire</u> =104,14758 g/mol <u>T_c (exp)</u> = 615,15 K <u>T_{éb} (exp)</u> = 473,15 K		<u>Structure moléculaire</u> : 	
Méthode des contributions des paires groupe-atome adjacent			
Groupe i	Nombre d'apparition N _i	Contribution C _i	N _i * C _i
-CH ₃ et >C<	2	2,424	4,848
>CH ₂ et >C<	1	2,124	2,124
>CH ₂ et -O-	1	-0,03	-0,03
>CH- et >C<	5	1,934	9,67
>CH- et -O-	1	-3,163	-3,163
Σ _{ij} N _{ij} C _{ij}			13,449
-OH	2	18,847	37,694
Σ _k M _k D _k			37,694
$T_c^{2,747} \times 10^{-6} = 4,72 + \sum_{ij} N_{ij}C_{ij} + \sum_k M_kD_k \Rightarrow$			
$T_c^{2,747} \times 10^{-6} = 4,72 + 13,449 + 37,694 = 661,017 \text{ K}$			
T _c (exp)= 615,15 K, erreur = 661,017-615,15 = 45,867 K, 7,45%			

1.3. Méthodes de contribution de groupes multi-ordre

1.3.1. Méthode de *Constantinou et Gani (1994)*

Une nouvelle technique d'estimation a été proposée par *Constantinou et Gani (1994)* pour la prédiction des propriétés thermophysiques des composés organiques purs en se basant seulement sur la structure moléculaire. L'estimation des propriétés est effectuée à deux niveaux, le premier, en utilisant des groupes de premier ordre, et le deuxième en utilisant des groupes de deuxième ordre qui utilisent les groupes de premier ordre comme blocs constitutifs.

Il existe deux types de groupes utilisés pour l'estimation, le premier pour les propriétés des corps purs déjà connus (*Reid, et al., 1987* et *Lyman, et al., 1990*), et le deuxième pour les propriétés des corps complexes ceux utilisés par (*Fredenslund, et al., 1977*; *Dear et Deal, 1969*).

Pour contourner cet inconvénient, ils ont proposé les groupes utilisés par (*Fredenslund, et al., 1977*; *Dear et Deal, 1969*) comme des groupes de premier ordre, ainsi, ils ont considéré les groupes apparus dans des cycles aliphatiques équivalents à ceux identiques apparus dans les non-cycliques.

Afin de fournir plus d'informations sur la structure moléculaire, ils ont introduit le concept de deuxième ordre, visant l'amélioration de la précision, la fiabilité et la gamme d'application d'un part, et de surmonter les inconvénients des groupes de premier ordre, comme la description partielle des effets de proximité et la distinction entre les isomères. Contrairement au cas des groupes de premier ordre, il peut y avoir des structures moléculaire qui n'ont besoin pas des groupes de deuxième ordre. La définition et l'identification des groupes de second ordre ont une base théorique, ils ont proposé *le principe de conjugaison*.

En basant sur les travaux de *Mavrovouniotis (1990)* et *Constantinou, et al., (1993)* comme plateforme, ils ont fait le choix des groupes de deuxième ordre. La base théorique du principe de conjugaison est que les composants sont présentés comme des hybrides de nombreux conjugués.

Chaque forme conjuguée est une structure idéalisée avec la localisation de la liaison à un ordre entier et une charge entière sur l'atome. La forme conjuguée purement covalente est la conjuguée dominante, et les formes ioniques sont les conjuguées récessives, qui peuvent être obtenues à partir d'un réarrangement des électrons paires de la forme dominante. Un opérateur de conjugaison définit un modèle particulier d'arrangement électronique. Lorsqu'il est appliqué au conjugué dominant, un opérateur donne toute une classe de conjugués récessifs. Les opérateurs de conjugaison sont représentés par une sous chaîne distincte avec deux ou trois liaisons, comme, C-C-C-H et O=C-C, la figure 1.1 représente un conjugué dominant, un conjugué récessif généré par l'opérateur de conjugaison correspondants.

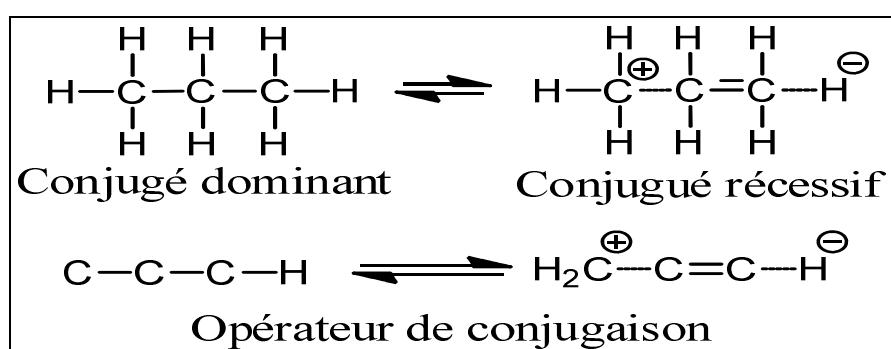


Figure 1.1 : Conjugué dominant, conjugué récessif généré et l'opérateur de conjugaison correspondant.

Dans ce cadre, la propriété d'un composé est estimée en déterminant et en combinant les propriétés de ces formes conjuguées. Les propriétés des formes conjuguées sont estimées par les opérateurs de conjugaison (une contribution fixe est associée à chaque opérateur par régression qui reflète la contribution de la classe correspondante à ces formes conjuguées).

Pour l'identification des groupes de deuxième ordre ils ont mis en œuvre les principes suivants :

- ♣ La structure d'un groupe de deuxième ordre doit inclure une sous chaîne distincte contient au moins un opérateur important de conjugaison.
- ♣ La structure d'un groupe de deuxième ordre fonctionnel devrait avoir les groupes voisins de premier ordre comme des blocs constitutifs, et devrait être aussi petite que possible.

- ♣ Les groupes de deuxième ordre basés sur l'opérateur (les opérateurs) commun (s) doivent être traités également dans la méthode.
- ♣ Les performances des groupes de deuxième ordre sont indépendantes de la molécule dans laquelle ils ont lieu, satisfaisant l'hypothèse fondamentale de la contribution de groupes.

Le modèle général proposé pour la prédiction des propriétés est le suivant :

$$f(X) = \sum_i N_i C_i + W \sum_j M_j D_j$$

Où :

f est la fonction d'une propriété X , C_i la contribution d'un groupe de premier ordre de type i qui apparaît N_i fois dans la molécule, D_j la contribution d'un groupe de deuxième ordre de type j qui apparaît M_j fois dans la molécule. Le constant W a deux valeurs, 1 avec la présence des groupes de deuxième ordre et 0 avec leur absence.

Pour la détermination des contributions de groupes de deuxième ordre ils ont employé la régression pas à pas (*stepwise*), en divisant la régression en deux étapes :

1. Une première régression est réalisée pour déterminer les contributions de premier ordre C_i , où ($W=0$), (c'est-à-dire que les contributions de deuxième ordre ne sont pas considérés dans cette étape) ;
2. En utilisant les contributions de premier ordre C_i , les groupes de deuxième ordre sont activés ($W=1$) et les contributions de deuxième ordre sont estimées par la régression.

Ce type de régression assure l'indépendance des contributions de premier ordre à ceux de deuxième ordre et les groupes de deuxième ordre servent à la correction de l'approximation des groupes de premier ordre.

Pour la sélection de la fonction f , ils ont fait une gamme des fonctions, et pour chaque propriété ils ont choisi la meilleure qui peut s'adapter le plus possible à la propriété, ainsi, l'expression doit être en mesure de fournir une extrapolation suffisante du comportement de la propriété.

Concernant la régression, ils ont utilisé la méthode des moindres carrés afin de déterminer la contribution des groupes (premier et deuxième ordre). L'algorithme d'optimisation utilisé pour la régression est l'approche modifiée par *Levenberg*, et la fonction objective est de minimiser la somme des carrés des différences entre les valeurs expérimentales et prédites (*Constantinou, et al., 1994*).

Pour la détermination les contributions de groupes de chaque propriété, ils ont utilisé, 285, 269, 251, 392, 312, 373, 225 et 333 valeurs expérimentales pour (T_c), (P_c), (V_c), (T_{eb}), (T_f), (ΔH°_{for}), (ΔH_v) et (ΔG_{for}) respectivement, les équations employées pour prédire ces propriétés sont les suivantes :

$$T_c = 181,128 \ln(\sum_i N_i C_i + W \sum_j M_j D_j) \quad (1.25)$$

$$P_c = \frac{1}{(0,10022 + \sum_i N_i C_i + W \sum_j M_j D_j)^2} + 1,3705 \quad (1.26)$$

$$V_c = -0,00435 + \sum_i N_i C_i + W \sum_j M_j D_j \quad (1.27)$$

$$T_{eb} = 204,359 \ln(\sum_i N_i C_i + W \sum_j M_j D_j) \quad (1.28)$$

$$T_f = 102,425 \ln(\sum_i N_i C_i + W \sum_j M_j D_j) \quad (1.29)$$

$$\Delta H^\circ_{for} = 10,835 + \sum_i N_i C_i + W \sum_j M_j D_j \quad (1.30)$$

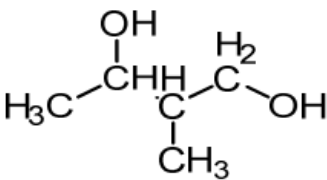
$$\Delta H^\circ_v = 6,829 + \sum_i N_i C_i + W \sum_j M_j D_j \quad (1.31)$$

$$\Delta G^\circ_{for} = -14,828 + \sum_i N_i C_i + W \sum_j M_j D_j \quad (1.32)$$

Le choix des groupes de premier ordre n'a aucune base théorique et ces groupes n'ont aucune relation avec le type de composé, ainsi, ces groupes ne sont pas capable de capter les effets de proximité et de faire la distinction entre les isomères. Leur considération que les groupes apparus dans des cycles aliphatiques équivalents à ceux identiques apparus dans les non-cycliques a l'inconvénient de ne pouvoir pas faire la distinction entre des isomères spéciaux.

Cependant, les groupes proposés de second ordre ne sont pas en mesure de fournir une bonne représentation de composés contenant plus d'un cycle ainsi que, dans certains cas, à chaîne ouverte composés polyfonctionnels avec plus de quatre atomes de carbone dans la chaîne principale (*Marrero, et al., 2001*).

Tableau 1.7 : Application de la méthode de *Constantinou et Gani (1994)*.

Température normale d'ébullition T_{eb}			
<u>Nom</u> : 2-méthyl-1,3-butanediol <u>Cas Number</u> : 684-84-4 <u>Formule moléculaire</u> : $C_5H_{12}O_2$ <u>La masse moléculaire</u> =104,14758 g/mol $T_{eb}(exp) = 473,15\text{ K}$		<u>Structure moléculaire</u> : 	
Méthode des contributions de groupes de deuxième ordre			
Groupe i	Nombre d'apparition N_i	Contribution C_i	$N_i * C_i$
-CH ₃	2	0,8894	1,7788
>CH ₂	1	0,9225	0,9225
>CH-	2	0,6033	1,2066
-OH	2	3,2152	6,4304
$\sum_i N_i C_i$			10,3383
CH(CH ₃)CH(CH ₃)	1	0,1798	0,1798
>CHOH	1	-0,5385	-0,5385
$W \sum_j M_j D_j$			-0,3587
$T_{eb} = T_{eb0} \ln (\sum_i N_i C_i + W \sum_j M_j D_j) \Rightarrow$ $T_{eb} = 204,359 \ln(10,3383 - 0,3587) = 470,136K$ $T_{eb}(exp)= 473,15\text{ K}$, erreur = $ 470,136-473,15 = 3,014K$, 0,63%			

Stefanis, et al. 2004 ont utilisé cette méthode pour l'estimation du point éclair, en ajoutant 26 groupes aux groupes de premier ordre, ainsi que 12 groupes de deuxième ordre. Le modèle général pour l'estimation de cette propriété est le même utilisé par *Constantinou et Gani (1994)*. L'analyse par la méthode des moindres carrés a été utilisée pour déterminer les contributions de premier et de deuxième ordre. L'approche modifiée de *Levenberg* a été utilisée afin de minimiser la somme totale des erreurs quadratiques entre les valeurs expérimentales et estimées (*Stefanis, et al., 2004*).

L'équation caractéristique de cette propriété est la suivante :

$$PE = 216 + \sum_i N_i C_i + W \sum_j M_j D_j \quad (1.33)$$

1.3.2. Méthode de *Marrero et Gani (2001)*

Marrero et Gani (2001) ont proposé une nouvelle technique pour la prédiction des propriétés de composés organiques purs effectuée sur trois niveaux, le niveau primaire utilise les contributions de groupes simples, ce niveau permet de décrire une grande variété de composés organiques, néanmoins, ce niveau est incapable de distinguer entre les isomères.

Alors que les groupes de deuxième ordre permettent une meilleure description des effets de proximité et la différenciation entre les isomères. Ces groupes font face avec les composés multifonctionnels, polaires et apolaires de taille moyenne, C=3-60, et les composés aromatiques ou cycloaliphatiques avec un seul cycle et plusieurs substituants.

Les groupes de troisième ordre offrent plus d'informations structurales sur les fragments moléculaires des composés dont la description via les groupes de premier et deuxième ordre est impossible, tels que les hétérocycliques complexes et les grands composés acycliques polyfonctionnels (C=7-60), dans le but est d'améliorer la précision, la fiabilité et le domaine d'application pour un nombre important de propriétés de corps purs.

Leur recherche de couvrir tous les types de composés par les groupes de premier ordre, elle leur a conduit à un nombre très importants de groupe 182 groupes, et pour la correction ils ont sélectionné 122 et 66 groupes structurels de deuxième et de troisième ordre respectivement. Pour la sélection des groupes de premier ordre, il a resté toujours le même problème, le choix de ces groupes n'a aucune base théorique, ainsi que la méthode est très compliquée.

Le nombre de données utilisées dans la génération de ces groupes structuraux est résumé dans le tableau suivant :





Tableau 1.8 : Nombre de propriétés utilisées dans la méthode de *Marrero et Gani (2001)*.

Propriétés	T _c	P _c	V _c	T _{eb}	T _f	ΔH ^o _{for}	ΔH ^o _v	ΔG ^o _{for}	ΔH ^o _{fus}
N° de données	783	775	762	1794	1547	686	437	679	711

Le modèle général de cette méthode est le suivant :

$$f(X) = \sum_i N_i C_i + W \sum_j M_j D_j + Z \sum_k O_k E_k$$

Où :

-  f est la fonction d'une propriété X ;
-  C_i est la contribution d'un groupe de premier ordre de type i qui apparaît N_i fois dans la molécule ;
-  D_j est la contribution d'un groupe de deuxième ordre de type j qui apparaît M_j fois dans la molécule. Le constant W a deux valeurs, 0 avec l'absence des groupes de deuxième ordre, et 1 avec leur présence ;
-  E_k est la contribution d'un groupe de troisième ordre de type k qui apparaît O_k fois dans la molécule, le constant Z aussi a deux valeurs, 0 avec l'absence des groupes de deuxième ordre et 1 avec leur présence.

Les fonctions choisies pour chaque propriétés sont les mêmes utilisées par *Constantinou et Gani (1994)*, et la régression cette fois ci a été réalisée en trois étapes :

1. La régression est réalisée pour déterminer les contributions C_i des groupes de premier ordre où les constants W et Z sont maintenus à zéro ;
2. En utilisant les contributions C_i résultantes de la première régression, et en fixant le constant W=1 et le Z=0, la régression est réalisée pour déterminer les contributions D_j des groupes de deuxième ordre ;
3. Les constants W et Z sont maintenus à 1, et en utilisant les contributions obtenues par les deux régressions précédentes C_i et D_j, la troisième régression est réalisée pour déterminer les contributions E_k des groupes de troisième ordre.

Les équations caractéristiques des propriétés de ce travail sont les suivantes :

$$T_c = 231,239 \ln(\sum_i N_i C_i + W \sum_j M_j D_j + Z \sum_k O_k E_k) \quad (1.34)$$

$$P_c = \frac{1}{(0,108998 + \sum_i N_i C_i + W \sum_j M_j D_j + Z \sum_k O_k E_k)^2} + 5,9827 \quad (1.35)$$

$$V_c = 7,95 + \sum_i N_i C_i + W \sum_j M_j D_j + Z \sum_k O_k E_k \quad (1.36)$$

$$T_{eb} = 222,543 \ln(\sum_i N_i C_i + W \sum_j M_j D_j + Z \sum_k O_k E_k) \quad (1.37)$$

$$T_f = 147,450 \ln(\sum_i N_i C_i + W \sum_j M_j D_j + Z \sum_k O_k E_k) \quad (1.38)$$

$$\Delta H_{for}^\circ = 5,549 + \sum_i N_i C_i + W \sum_j M_j D_j + Z \sum_k O_k E_k \quad (1.39)$$

$$\Delta H_v^\circ = 11,733 + \sum_i N_i C_i + W \sum_j M_j D_j + Z \sum_k O_k E_k \quad (1.40)$$

$$\Delta G_{for}^\circ = -34,967 + \sum_i N_i C_i + W \sum_j M_j D_j + Z \sum_k O_k E_k \quad (1.41)$$

$$\Delta H_{fus}^\circ = -2,806 + \sum_i N_i C_i + W \sum_j M_j D_j + Z \sum_k O_k E_k \quad (1.42)$$

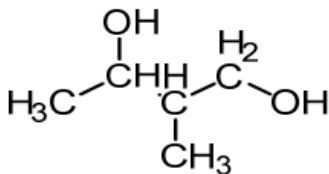
L'algorithme d'optimisation utilisé pour l'ajustement des données a été la technique de *Levenberg-Marquardt*, et la fonction objective est de minimiser la somme des carrés des différences entre les valeurs expérimentales et prédites (*Marrero, et al., 2001*).

Hukkerika, et al., 2013 ont étendu la prédiction des propriétés thermodynamiques via la contribution des groupes de trois ordres pour estimer l'enthalpie de formation dans la phase gazeuse $\Delta_f H^\circ_{(gaz)}$. 861 valeurs expérimentales à partir la base de données *CAPEC* ont été utilisées pour développer les nouveaux paramètres.

La même procédure de régression a été employée, mais, les groupes structuraux sont pas les mêmes, parce que la sélection de ces groupes dépend de la base de données. Donc, des nouveaux paramètres ont été introduits avec des nouvelles valeurs. Le nombre de paramètres est 220, 130, 74 groupes de premier, deuxième et de troisième ordre respectivement, l'équation caractéristique de la propriété est la suivante :

$$\Delta_f H^\circ_{(gaz)} = 42,2361 + \sum_i N_i C_i + W \sum_j M_j D_j + Z \sum_k O_k E_k \quad (1.43)$$

Tableau 1.9 : Application de la méthode de Marrero et Gani (2001).

Température normale d'ébullition T_{eb}			
<u>Nom</u> : 2-méthyl-1,3-butanediol <u>Cas Number</u> : 684-84-4 <u>Formule moléculaire</u> : $C_5H_{12}O_2$ <u>La masse moléculaire</u> =104,14758 g/mol $T_{eb}(\text{exp}) = 473,15 \text{ K}$		<u>Structure moléculaire</u> : 	
Méthode des contributions de groupes de deuxième ordre			
Groupe i	Nombre d'apparition N_i	Contribution C_i	$N_i * C_i$
-CH ₃	2	0,8491	1,6982
>CH ₂	1	0,7141	0,7141
>CH-	2	0,2925	0,585
-OH	2	2,5670	5,134
$\sum_i N_i C_i$			8,1313
CH(CH ₃)CH(CH ₃)	1	0,6130	0,6130
>CHOH	1	-0,2825	-0,2825
$W \sum_j M_j D_j$			0,3305
HO-(CH _n) _m -OH	1	0,7193	0,7193
$Z \sum_k O_k E_k$			0,7193
$T_{eb} = T_{eb0} \ln (\sum_i N_i C_i + W \sum_j M_j D_j + Z \sum_k O_k E_k) =>$ $T_{eb} = 222,543 \ln(8,1313 + 0,3305 + 0,7193) = 493,41K$ $T_{eb}(\text{exp})= 473,15 \text{ K}$, erreur = $ 493,41-473,15 = 20,26K$, 4,28%			

1.4. Méthode de contribution de la position de groupes

1.4.1. Méthode de Wang, et al., (2008)

Dans cette méthode, l'estimation de la température critique des composés organiques inclus des chaîne de carbone de C₂ jusqu'à C₁₈, avec des composés contenant (O, N, Cl, Br et S) est proposée via la contribution de groupes. Ils ont aussi introduit la correction via la position de groupe afin de prendre en compte l'interaction à longue distance ce qui ainsi pourrait faire la distinction entre les isomères (*cis*- et *trans*-).

Un ensemble de 467 composés ont été utilisés dans le développement de la méthode, où le nombre de paramètres est 73, la représentation des ces paramètres est presque la même employée par (Benson, et al., 1969), C-(CH₃)(H)₃ où le premier carbone représente l'élément qui forme le centre de groupe, les groupes ou atomes entre parenthèse représentent les éléments auquel il est lié.

L'équation proposée pour la prédiction de la température critique est la suivante :

$$T_c = \sum_i N_i A_i + \sum_j A_j \tanh \left(\frac{N_j}{N} \right) + \sum_k A_k P_k - 6900,452 \exp \left(\frac{1}{M} \right) + 7,723 \exp \left(\frac{1}{N} \right) - 7409,2$$

$$N = \sum_i N_i + \sum_j N_j \quad (1.44)$$

Où :

- ✚ A_i et A_j sont les contributions de groupes de type i et j ;
- ✚ N_i est le nombre de groupe dans lequel l'atome de carbone forme le centre de groupe dans la molécule, N_j est le nombre de groupe dans lequel un autre atome que le carbone forme le centre et N est le nombre total de groupes ;
- ✚ P_k est le facteur de la corrélation de position, ce facteur est déterminé selon la nomenclature de UICPA, par exemple dans 2,2,4,5-tetraméthylhexane, selon UICPA, la nomenclature de ce composé débute de droite à gauche, donc, le facteur de position pour le groupe (>CH-) égal à 4 et 5, la figure 1.2 montre la procédure avec laquelle on obtient P_k où M est la masse moléculaire (Wang, et al., 2008).

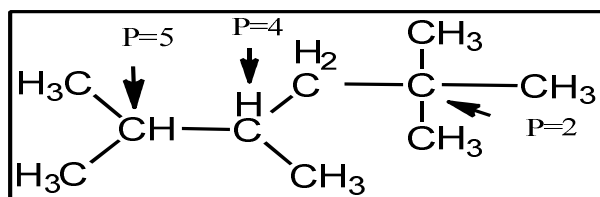


Figure 1.2 : Détermination de la position d'un groupe dans la molécule.

En 2009 *Wang et al.* ont proposé une équation pour la prédiction de la température normale d'ébullition T_{eb} , une base de donnée constituée de 417 composés organiques a été utilisée pour le développement de cette équation.

La température normale d'ébullition est une fonction de toutes les contributions de groupes, ainsi, la correction de la position. Dans cette méthode, les groupes de deuxième ordre sont les mêmes employés par (*Benson, et al., 1969*) et quelques groupes de troisième ordre ont été utilisées (*Wang, et al., 2009*).

L'équation utilisée pour la prédiction de cette propriété est la suivante :

$$T_{eb} = \sum_i N_i A_i + \sum_j A_j \tanh \tanh \left(\frac{N_j}{N} \right) + \sum_k A_k P_k - 4491,3432 \exp \exp \left(\frac{1}{M} \right) + 17,3615 \exp \exp \left(\frac{1}{N} \right) + 4775,5907$$

$$N = \sum_i N_i + \sum_j N_j \quad (1.45)$$

Une équation pour l'estimation de la température de fusion a été proposée par *Wang, et al., (2009)*, 730 composés organiques ont été employés dans le développement de cette équation :

$$T_f = \sum_i N_i A_i + \sum_j A_j \tanh \tanh \left(\frac{N_j}{N} \right) + \sum_k A_k P_k - 5758,997 \exp \exp \left(\frac{1}{M} \right) + 51,127 \exp \exp \left(\frac{1}{N} \right) + 5963,486$$

$$N = \sum_i N_i + \sum_j N_j \quad (1.46)$$

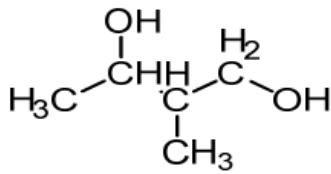
(*Wang, et al., 2009*).

Une somme de 287 composés organiques ont été employés pour prédire le point d'éclair PE (*Jia, et al., 2012*).

$$PE = \sum_i N_i A_i + \sum_j A_j \tanh \tanh \left(\frac{N_j}{N} \right) + \sum_k A_k P_k - 11818,047575 \exp \exp \left(\frac{1}{M} \right) + 184,520799 \exp \exp \left(\frac{1}{N} \right) + 11962,9332$$

$$N = \sum_i N_i + \sum_j N_j \quad (1.47)$$

Tableau 1.10 : Application de la méthode de Wang, *et al.*, (2008).

Température normale d'ébullition T _{eb}			
<u>Nom</u> : 2-méthyl-1,3-butanediol <u>Cas Number</u> : 684-84-4 <u>Formule moléculaire</u> : C ₅ H ₁₂ O ₂ <u>La masse moléculaire</u> =104,14758 g/mol <u>Nombre d'atomes</u> : N=7 T _{eb} (exp) = 473,15 K		<u>Structure moléculaire</u> : 	
Méthode des contributions de position des groupes			
Groupe i	Nombre d'apparition N _i	Contribution C _i	N _i * C _i
C-(CH)(H) ₃	2	0,2688	0,5376
C-(C)(O)(H) ₂	1	39,1793	39,1793
C-(C) ₂ (O)(H)	1	49,1420	49,1420
C-(C) ₃ (H)	1	38,4077	38,4077
Σ _i N _i A _i			127,267
O-(C)(H)	2	340,1722	680,344
Σ _j A _j tanh(N _j /N) = 680.344 * tanh($\frac{2}{7}$)			189,261
>CH- position	2	-4,1845	-8,369
-OH position	1+3=4	-0,3567	-1,4268
Σ _k A _k P _k			-9,796
T _{eb} = Σ _i N _i A _i + Σ _j A _j tanh(N _j /N) + Σ _k A _k P _k - 4491,3432 exp (1/M) + 17,3615 exp(1/N) + 4775,5907 T _{eb} = 127,267 + 189,261 - 9,796 - 4491,3432 exp (1/104,147) + 17,3615 exp(1/7) + 4775,5907 = 567,674K T _{eb} (exp)= 473,15 K, erreur = 567,674-473,15 = 94,513K, 19,97%			

1.5. Méthodes basées sur la combinaison

1.5.1. Méthode de *Wilson et Jasperson (1996)*

Cette méthode permet d'estimer la température et la pression critiques des espèces organique et inorganiques. L'estimation de ces propriétés en utilisant cette méthode est basée sur deux niveaux, le premier emploie les contributions des atomes constituant la molécule, et le deuxième utilise les contributions de groupes fonctionnels constituant la molécule. Le modèle de la température critique utilise le nombre de cycle présent dans la molécule et la température d'ébullition comme entrées, et le modèle de la pression critique utilise la température critique et le nombre de cycle présents dans la molécule comme entrées, les équations caractérisantes de ce modèle sont les suivantes :

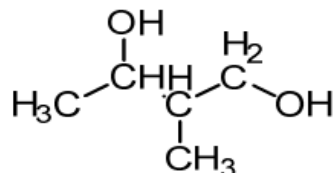
$$T_c = T_{eb} / \left[0,048271 - 0,019846n_r + \sum_k M_k A_k + \sum_i N_i C_i \right]^{0,2} \quad (1.48)$$

$$P_c = 0,0186233T_c \times [-0,96601 + \exp(y)]^{-1} \quad (1.49)$$

$$\text{Où : } y = -0,00922295 - 0,0290403n_r + 0,041 \left[\sum_k M_k A_k + \sum_i N_i C_i \right]$$

- n_r est le nombre de cycle présent dans la molécule ;
- A_k est la contribution d'un atome de type k dans la molécule et M_k le nombre d'apparition de cet atome dans la molécule ;
- C_i est la contribution d'un groupe de type i dans la molécule et N_i son nombre d'apparition dans la molécule.

Tableau 1.11 : Application de la méthode de *Wilson et Jasperson (1996)*.

Température critique T _c			
<u>Nom</u> : 2-méthyl-1,3-butanediol <u>Cas Number</u> : 684-84-4 <u>Formule moléculaire</u> : C ₅ H ₁₂ O ₂ <u>La masse moléculaire</u> =104,14758 g/mol T _c (exp) = 615,15 K T _{eb} (exp) = 473,15 K		<u>Structure moléculaire</u> : 	
Méthode des contributions des atomes et de groupes			
Atome i	Nombre d'apparition N _i	Contribution C _i	N _i * C _i
C	5	0,008532	0,04266
H	12	0,002793	0,033516
O	2	0,020341	0,040682
Σ _k M _k A _k			0,116858
-OH, C ₅ ou plus	2	0,01	0,02
Σ _i N _i C _i			0,02
T _c = T _{eb} /[0,048271 – 0,019846n _r + Σ _k M _k A _k + Σ _i N _i C _i] ^{0,2} =>			
T _c = 473,15 /[0,048271 + 0,116858 + 0,02] ^{0,2} = 662,985K			
T _c (exp)= 615,15 K, erreur = 662,985-615,15 = 47,835K, 7,77%			

1.5.2. Méthode de Cordes et Rarey (2002)

Cordes et Rarey (2002) ont proposé une nouvelle approche pour la prédiction du point normal d'ébullition. 2500 composés organiques ont été utilisées, ces composés et leurs valeurs expérimentales des propriétés ont été fournis par la banque de donnée Dortmund DDB. La prédiction est basée exclusivement sur la structure moléculaire du composé. Au lieu d'améliorer l'estimation en introduisant différents niveaux d'approximation, ils ont suggéré une définition plus scientifique aux groupes de premier ordre, qui constitue la base de la contribution de groupe.

A partir de l'équation de *Clausius–Clapeyron*, la température normale d'ébullition peut être écrite par le rapport entre ΔH_v et ΔS_v :

$$T_{\text{éb}} = \frac{\Delta H_v}{\Delta S_v} \quad (1.50)$$

A partir de cette équation et l'analyse de quelques méthodes d'estimation disponibles, ils ont proposé l'équation suivante pour prédire la température normale d'ébullition :

$$T_{\text{éb}} = \frac{\sum_i N_i C_i}{n^{0,6713} + 1,4442} + 59,344 \quad (1.51)$$

Où :

- C_i est la contribution de groupe de type i ;
- N_i est le nombre d'apparition de ce groupe dans la molécule ;
- n est le nombre d'atomes dans la molécule (sans compter l'hydrogène).

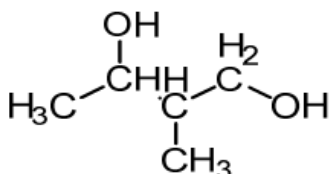
La définition des groupes a été réalisée en fonction de leurs voisins. Donc, pour cette méthode ils ont considéré que :

- ✚ Il n'est pas nécessaire de faire la distinction entre le carbone et la silicone comme atome voisin ;
- ✚ Une électronégativité très élevée (N, O, F et Cl), ou voisins aromatiques souvent influent significativement la contribution des groupes structurels ;
- ✚ Il est souvent très important si un groupe fait partie d'une chaîne, un cycle ou système aromatique.

Un ensemble de groupes de second ordre a été employé pour la correction.

La régression dans cette méthode a été effectuée à deux niveaux. La première régression a été utilisée pour optimiser les paramètres d'ajustement (a, b et c) en utilisant une régression non-linéaire en employant un algorithme simple. La deuxième régression a été de type multilinéaire, et la méthode des moindres carrées a été employée. La fonction objective est de minimiser l'écart quadratique entre les valeurs expérimentales et estimées (*Cordes, et al., 2002*).

Tableau 1.12 : Application de la méthode de *Cordes et Rarey (2002)*.

Température normale d'ébullition T _{eb}			
<u>Nom</u> : 2-méthyl-1,3-butanediol <u>Cas Number</u> : 684-84-4 <u>Formule moléculaire</u> : C ₅ H ₁₂ O ₂ <u>La masse moléculaire</u> =104,14758 g/mol <u>Nombre d'atomes</u> : N=7 T _{eb} (exp) = 473,15 K		<u>Structure moléculaire</u> : 	
Méthode des contributions de groupes			
Groupe i	Nombre d'apparition N _i	Contribution C _i	N _i * C _i
-CH ₃	2	188,555	377,11
>CH ₂	1	250,119	250,119
>CH-	2	260,938	521,876
-OH	2	411,08	822,16
Σ _i N _i C _i			1971,265
$T_{eb} = \frac{\sum_i N_i C_i}{n^{0,6713} + 1,4442} + 59,344 \Rightarrow T_{eb} = \frac{1971,265}{7^{0,6713} + 1,4442} + 59,344 = 443,108K$			
T _{eb} (exp)= 473,15 K, erreur = 443,108-473,15 = 30,042K, 6,34%			

1.5.3. Méthode de Nannoolal, et al., (2004)

En se basant sur le travail de *Cordes et Rarey (2002)*, cette méthode a été étendue par l'élargissement de la base de données (2850 valeurs expérimentales de la température d'ébullition), l'inclure des paramètres stériques, corrections supplémentaires, paramètres d'interaction de groupes ainsi que plusieurs groupes additionnels afin d'augmenter la gamme d'application. Le formalisme mathématique a été modifié pour permettre un examen séparé et une régression des contributions individuelles.

Dans cette méthode 117 groupes de premier ordre ont été sélectionnés, 16 groupes de deuxième ordre pour la correction, ainsi que 79 interactions des groupes pour les composés multifonctionnels, en total 212 paramètres structurels.

L'équation proposée dans ce travail pour estimer T_{eb} est la suivante:

$$T_{eb} = \frac{\sum_i N_i C_i}{n^{0,6583} + 1,6868} + 84,3395 \quad (1.52)$$

La procédure de la régression est la même utilisée dans la méthode de *Cordes et Rarey (2002)*. Le développement de cette méthode a été réalisé selon les familles chimiques, c'est-à-dire le choix des paramètres structurels et la régression.

Le deuxième modèle employé pour l'interaction de groupes est le suivant :

$$IG = \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^m \frac{C_{i-j}}{m-1} \text{ où } (C_{i-j} = C_{j-i}) \quad (1.53)$$

Où le C_{i-j} est la contribution de l'interaction entre deux groupes i et j , le n est le nombre des atomes dans la molécule sauf l'hydrogène et m le nombre de groupe i et j (*Nannoolal, et al., 2004*).

En utilisant la même procédure employée dans le développement du modèle de la température normale d'ébullition, ainsi, 588, 486 et 348 valeurs expérimentales de la température, la pression et le volume critiques ils ont développé des modèles permettent de prédire les constants critiques, sachant que la température normale d'ébullition est nécessaire pour l'estimation de la température critique, les équations caractéristiques de ces propriétés sont les suivantes :

$$T_c = T_{eb} \left(0,6990 + \frac{1}{0,9889 + \left(\sum_i N_i C_i + CI \right)^{0,8607}} \right) \quad (1.54)$$

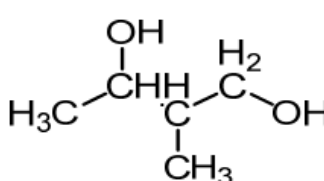
$$\frac{P_c}{\text{kPa}} = \frac{M^{-0,14041} \left(\frac{\text{g}}{\text{mol}} \right)}{\left(0,00939 + \sum_i N_i C_i + CI \right)^2} \quad (1.55)$$

$$\frac{V_c}{10^{-6} \text{ m}^3 \text{ mol}^{-1}} = \frac{\sum_i N_i C_i + CI}{n^{-0,2266}} + 86,1539 \quad (1.56)$$

Où :

- C_i est la contribution de groupe de type i ;
- N_i est le nombre d'apparition de groupe i ;
- M est la masse molaire (g/mol) ;
- n est le nombre des atomes dans la molécule (sans compter l'hydrogène) ;
- CI est la contribution des interactions des groupes (*Nannoolal, et al., 2007*).

Tableau 1.13 : Application de la méthode de *Nannoolal, et al., (2004)*.

Température normale d'ébullition T _{eb}			
<u>Nom</u> : 2-méthyl-1,3-butanediol <u>Cas Number</u> : 684-84-4 <u>Formule moléculaire</u> : C ₅ H ₁₂ O ₂ <u>La masse moléculaire</u> =104,14758 g/mol <u>Nombre d'atomes</u> : N=7 T _{eb} (exp) = 473,15 K		<u>Structure moléculaire</u> : 	
Méthode des contributions de groupes et d'interaction de groupes			
Groupe i	Nombre d'apparition N _i	Contribution C _i	N _i * C _i
-CH ₃	2	177,3066	354,6132
>CH ₂	1	239,4531	239,4531
>CH-	2	240,6785	481,357
-OH	2	390,2446	780,4892
Interaction i	Nombre d'apparition N _i	Contribution C _i	N _i * C _i
HO-OH	1	291,7985	291,7985
$IG = \frac{C_{HO-OH} + C_{HO-OH}}{n(m-1)} = \frac{2C_{HO-OH}}{7(2-1)} = \frac{2}{7} * C_{HO-OH}$			83,371
$\sum_i N_i C_i$			1939,2835
$T_{eb} = \frac{\sum_i N_i C_i}{n^{0,6583} + 1,6868} + 84,3395 \Rightarrow T_{eb} = \frac{1939,2835}{7^{0,6583} + 1,6868} + 84,3395 = 451,14K$			
T _{eb} (exp)=473,15 K, erreur = 451,14-473,15 = 22,01K, 4,65%			

1.6. Autres méthodes

Plusieurs méthodes d'estimation des propriétés des composés organique sont rapportées dans la littérature (*Poling, et al., 2001*), (*Boethling, et al., 2000*), etc.

La méthode proposée par *Stein et Brown (1994)* pour la prédiction de la température normale d'ébullition est une extension de la méthode de *Joback et Reid (1987)*. Cette extension est clairement observée dans l'augmentation du nombre de paramètres structurels, de 41 groupes jusqu'à 85 groupes. Cette méthode a été développée par 4426 composés organiques et ils ont proposé une correction au modèle de *Joback et Reid (1987)* (*Reinhard, et al., 1999*).

La méthode des groupes espace vectoriel, cette méthode a été proposée par *Wen et Qiang (2002)*, une quarantaine (40) de groupes ont été sélectionnés dans cette méthode pour décrire les composés organiques, ces groupes sont les mêmes utilisés par *Joback et Reid (1987)*. Dans cette méthode la molécule est considérée comme étant dans un espace donné, et chaque groupe de cette molécule n'est qu'un point dans l'espace. La méthode consiste à transformer la molécule à une matrice via des graphes topologiques et en utilisant cette matrice qui représente la molécule on peut prédire la propriété via leur modèle (*Nannoolal, 2004*).

1.7. Discussion des méthodes d'estimation

Le but principal des méthodes d'estimation citées dans ce chapitre est de prédire les propriétés des composés organiques purs et complexes visant les critères suivants :

- ✚ La vaste gamme d'application ;
- ✚ L'exigence minimale des données d'entrées ;
- ✚ L'approche choisie doit fournir une précision fiable des propriétés (le minimum d'erreurs) ;
- ✚ Offre une bonne distinction entre les isomères ;
- ✚ Le temps de calcul doit être le minimum possible ;

On ne peut pas trouver une seule méthode avec tous les critères cités ci-dessus, parce que ces critères sont compétitifs, la vaste gamme d'application ainsi que la distinction entre les isomères nécessite une vaste base de données, ce qui augmente le nombre de groupes structuraux proposés pour décrire la structure des molécules, en conséquent, le temps de calcul est plus grand et la méthode devient insupportable.

Si on veut une méthode avec une bonne précision, les méthodes qui basent sur un modèle non-linéaire offrent cette précision mais ces méthodes sont très compliquées d'un part, et de l'autre part, les méthodes avec un modèle qui vient d'une régression linéaire sont simples à utiliser mais leur précision est minimale.

Concernant le critère des entrées, généralement les méthodes développées jusqu'à maintenant utilisent la masse molaire ou le nombre d'atomes comme entrées, sauf quelques méthodes qui nécessitent la température d'ébullition comme entrée pour la prédiction de la température critique.

Il existe des méthodes qui fournissent une estimation simple et rapide comme la méthode de *Joback et Reid(1987)*, mais avec une gamme d'application limitées et un intervalle d'erreurs considérable, malgré tous les inconvénients qu'on peut trouver dans la littérature sur cette méthode, mais elle a été toujours la base du développement de plusieurs d'autres méthodes.

Après cette méthode, d'autres techniques ont été développées afin de surmonter leur inconvénients, comme la méthode de *Marrero et Pardillo (1999)* ainsi que *Wen et Quiang (2001)*, qui ont utilisé les contributions des interactions de groupes au lieu des contributions de groupes simples, ces méthodes ont fourni une meilleure distinction entre les isomères, plus de précision, mais la gamme d'application a restée toujours limitée due au manque de données.

La disponibilité des propriétés expérimentales a permis de développer des méthodes avec une vaste gamme d'application, une précision plus grande, comme la méthode de *Marrero et Gani (2001)*, qui a été effectuée sur trois niveaux, mais cette méthode est plus complexe, c'est pour cette raison que ces méthodes ont été intégrées dans des logiciels pour faciliter l'estimation.

D'autres méthodes comme de *Nannoolal, et al., (2004)* et la méthode des contributions de la position de groupes ont été développées, en essayant d'améliorer l'estimation de propriétés, en élargissant la gamme d'application, améliorant la distinction entre les isomères, augmentant la précision, etc, mais il reste toujours des inconvénients, soit concernant le nombre de paramètres ou la complexité de la méthode.

Donc, un ingénieur ou un chimiste qui doit utiliser une technique d'estimation doit sacrifier en choisissant la simplicité et la rapidité, au lieu de la vaste gamme d'application, la distinction entre les isomères, la fiabilité et la précision, ou d'orienter vers les méthodes les plus complexes, pour profiter leurs avantages.

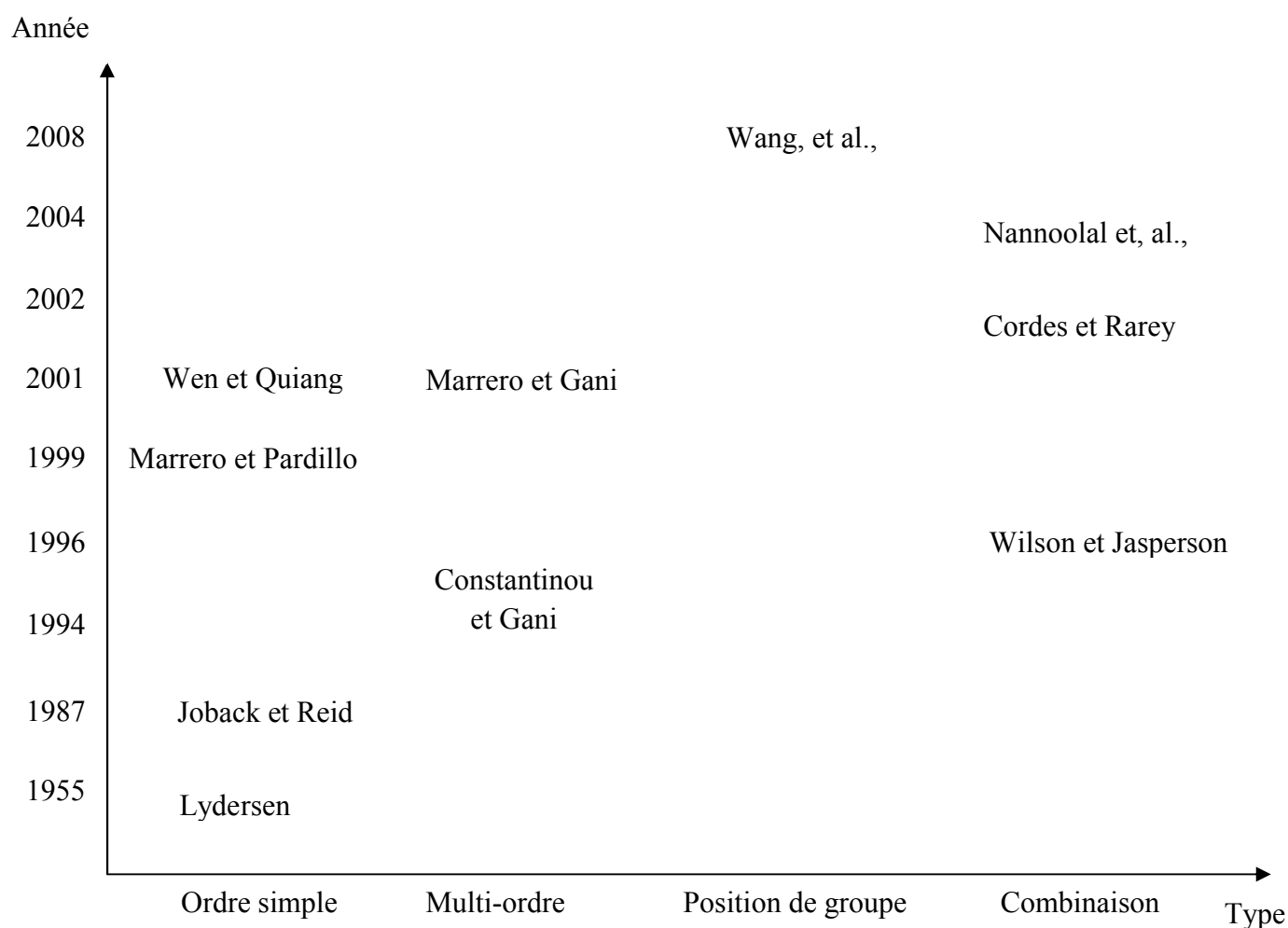


Figure 1.3 : Schéma montrant la distribution des méthodes d'estimation en fonction du type et des années.

Chapitre 02 :

Méthodologie du développement de la méthode proposée

2.1. Introduction

Les méthodes d'estimation des propriétés, telles que la méthode de *Joback et Reid (1987)*, *Constantinou et Gani (1994)*, *Marrero et Gani (2001)*, *Nannoolal, et al., (2004)*, sont très utilisées dans la prédiction de propriétés des composés organiques purs ou complexes.

Ces méthodes fournissent une estimation simple et rapide en considérant que la propriété est obtenue en sommant les contributions des atomes, des groupes fonctionnels, des interactions de groupes, etc, qui composent la molécule multipliant par leur nombre de fois d'apparition dans la molécule, c'est-à-dire que la propriété est obtenue seulement par la structure de la molécule, en se basant sur le principe que les propriétés sont en fonction des paramètres structuraux dépendants.

Le développement de ces méthodes nécessite les étapes suivantes :

- ❖ Comprendre la relation structure-propriété et choisir l'approche qui représente mieux cette relation ;
- ❖ Le choix des groupes structuraux (variable indépendants) et une procédure de fragmentation pour la description des molécules;
- ❖ La collection d'une base de données contenant une large gamme des composés organiques avec leurs valeurs expérimentales des propriétés;
- ❖ La sélection d'un modèle mathématique qui décrit la propriété en fonction des groupes structuraux;
- ❖ Le choix d'une procédure de régression qui peut traiter correctement un grand nombre des propriétés et de paramètres à la fois;
- ❖ Le choix des outils informatiques les plus performants pour minimiser le temps de calculs et augmenter la précision;
- ❖ Le choix des outils d'analyse et de validation (R^2 , R^2 ajusté, PEMA, EMA, DS, etc), avec lesquels on optimise les modèles développés et en même temps on compare ces modèles avec ceux déjà cités dans la littérature.

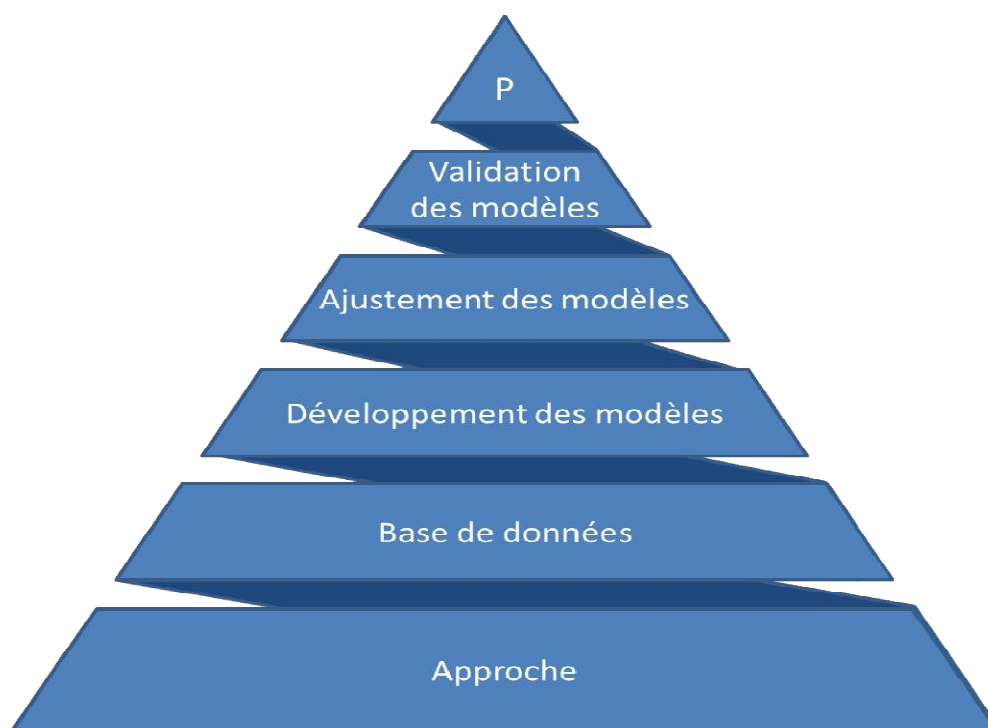


Figure 2.1 : Schéma résumant la démarche du développement des modèles proposés.

2.2. Approche choisie

Le type des liaisons entre les atomes de la molécule appelées aussi « forces intramoléculaires » (*Myers, 2003*) ont une influence très importante sur les propriétés de ces molécules ainsi que leur structure (*Johnson, 2002*). Une petite comparaison entre les propriétés de quelques composés en fonction des liaisons peut montrer clairement l'influence de ces liaisons. Le tableau 2.1 donne des exemples :

Tableau 2.1: Comparaison de la température d'ébullition des composés en fonction de leur type de liaison.

Liaison simple	T _{eb} (K)	Liaison double	T _{eb} (K)	Liaison triple	T _{eb} (K)
Ethane	184,55	Ethylène	169,47	Acétylène	189,35
Propane	231,11	Propène	225,46	Propyne	249,94
Butane	272,6	Butène	266,9	Butyne	281,22

On peut voir clairement l'influence du type de la liaison (simple, double ou triple) sur la valeur de la température d'ébullition des composés qui ont le même nombre de carbone.

Marrero et Pardillo (1999) ont proposé la contribution des interactions de groupes comme approche pour leur méthode au lieu de la contribution de groupes simples qui constituent la molécule. La méthode a présenté une amélioration remarquable concernant la précision et la distinction entre les isomères en comparaison avec les méthodes déjà disponibles dans la littérature surtout la méthode de *Joback et Reid (1987)*.

Le manque de données expérimentales a influencé la gamme d'application et la précision de la méthode, qu'on peut utiliser avec les petites molécules. Cependant, avec les molécules de grosses tailles tels que les composés polycycliques, polyfonctionnels, etc, la méthode échoue.

Dans ce travail, on propose la même approche que celle utilisée par *Marrero et Pardillo (1999)*, mais avec des modifications, visant l'amélioration de la méthode.

- Eliminer les groupes non nécessaire, comme par exemple (-CH₃ et -CH₃), (=CH₂ et =CH₂), (-CH₃ et -Br) qui peuvent se trouver seulement dans l'éthane, éthylène, bromométhane respectivement, ces groupes structuraux n'ont aucune utilité, car ils ne vont pas apparaître dans d'autre composés, et on a toutes les propriétés de ces composés ;
- Modifier quelques groupes structuraux pour minimiser leur nombre au minimum possible, comme les groupes (-CH₃ et -CN) et (-CH₂ et -CN), donc on les remplace par un seul groupe (-C≡ et ≡N) ;
- Ajouter des nouveaux groupes selon les besoins et la disponibilité de données, comme par exemple le groupes structuraux (>CH- et -Br), (>CH- et -C≡), etc.
- Faire la distinction entre la liaison aromatique et la liaison double, parce que ces deux liaisons ne sont pas identiques, l'enthalpie moyenne de la liaison double égale 612 kJ/mol et celle de la liaison aromatique égale 518 kJ/mol, donc, on doit distinguer entre ces deux liaisons par des groupes structuraux différents (*Atkins, et al., 1998*).

2.3. Définition des groupes structuraux

2.3.1. Groupes principaux

Les propriétés dans ce travail sont obtenues en fonction des groupes, ces derniers sont obtenus par la structure moléculaire. La définition de ces groupes est réalisée en fonction des interactions entre les groupes simples, ces groupes sont les mêmes que ceux définis par *Joback et Reid (1987)* avec quelques additions.

Les critères de choix des groupes :

1. L'idée principale sur laquelle cette méthode est basée, c'est l'interaction entre les groupes, c'est à dire « la liaison », donc, le choix des groupes est obéi au type de la liaison entre les groupes structuraux, soit simple, double, triple ou aromatique. Malgré que ce choix peut augmenter le nombre de groupes structuraux, mais il a plusieurs avantages ; il améliore la distinction entre les isomères, offre une meilleure description de la structure moléculaire, augmente la fiabilité et la précision des modèles.
2. Les familles chimiques ont une structure molécule divers (le groupe fonctionnel, les liaisons, les atomes, etc), chaque famille a une gamme distincte concernant les propriétés physiques, thermodynamiques ou autres, le tableau 2.2 illustre la différence entre la température d'ébullition des composés qui ont la même formule brute mais ils appartiennent à des familles diverse.

Tableau 2.2 : Température d'ébullition d'un aldéhyde, acétone, alcool et un éther.

Les composés	C ₃ H ₆ O propanal	C ₃ H ₆ O Acétone	C ₃ H ₈ O 1-Propanol	C ₃ H ₈ O méthyl éthyle éther
T _{eb} (K)	321,15	329,44	370,35	280,5

A partir de ce tableau, on peut noter la différence entre la température d'ébullition de ces composés. Dans ce travail, on a choisi des groupes structuraux qui font la distinction entre ces familles pour obtenir des valeurs estimées proches aux valeurs expérimentales, comme par exemple le groupe (-CH₃ et >CO) est utilisé pour les acétones, le groupe (-CH₃ et -CHO) est utilisé pour les aldéhydes et le groupe (-CH₃ et -COOH) est utilisé pour les acides carboxyliques, etc.

3. Des groupes structurels ont été sélectionnés pour assurer la distinction entre la structure des composés cycliques et acycliques. L'expérience acquise par les autres méthodes montre que cette distinction entre ces deux types de structure assure plus de fiabilité à la méthode, *Constantinou et Gani (1994)* ont considéré que les groupes apparus dans des cycles aliphatiques sont équivalents en contribution à ceux identiques en structure apparus dans les non-cycliques, tels que le groupe $>\text{CH}_2$, ils ont considéré que la contribution de ce groupe est la même, soit dans le propane ou cyclopropane, cette considération n'a pas le pouvoir de faire la distinction entre ces deux structures.
4. Le manque de données expérimentales a obligé *Marrero et Pardillo (1999)* de considérer que les groupes structurels qui ont une double liaison tels que $(-\text{CH}=\text{R})$ et $(=\text{CH}-\text{R})$ qui apparaissent dans des composés cycles aliphatiques sont identiques aux groupes qui apparaissent dans des cycles aromatiques. Dans ce travail, on a ajouté des nouveaux groupes structurels pour assurer la distinction entre ces deux types de structures.
5. Dans ce travail, l'interaction est entre deux groupe voisins seulement, ce n'est pas entre deux atomes. Dans le groupe structurel $(-\text{CH}_3 \text{ et } >\text{CO})$ par exemple, on compte une interaction entre ces deux groupes sans compter l'interaction entre le carbone et l'oxygène dans le groupe $>\text{CO}$.
6. La priorité dans ce niveau est aux familles chimiques dans la distinction entre les groupes structurels, pour les esters, on compte l'interaction entre $-\text{CH}_3$ et $-\text{COO}-$, parce que ce groupe est défini pour les esters, donc, on ne doit pas compter l'interaction entre $-\text{CH}_3$ et $>\text{CO}$ un groupe structurels et l'interaction entre $>\text{CO}$ et $-\text{O}-$ un autre groupe structurels. Le tableau 2.3 résume les groupes utilisés pour la génération des groupes structurels :

Tableau 2.3: Groupes utilisés pour la génération des groupes structuraux.

Groupes avec liaison simple					
-CH ₃	>CH ₂		>CH-	>C<	
Groupes avec liaison double					
=CH ₂	=CH- s	-CH= d	=C< s	>C= d	=C=
Groupes avec liaison triple					
≡CH			≡C-		
Groupes dans des cycles avec liaison simple					
>CH ₂ ⊙		>CH-⊙		>C<⊙	
Groupes dans des cycles avec liaison double					
=CH- s ⊙		-CH= d⊙		>C= d⊙	
Groupes dans des cycles aromatiques					
=CH- @			=C< @		
Groupe d'acétone dans un cycle					
>CO⊙					
Groupe d'acétone avec liaison simple					
>CO					
Groupe d'oxygène avec liaison simple					
-O-					
Groupe amine avec liaison simple					
>NH s					
Groupe soufre avec liaison simple					
-S- s					
Groupe d'un isocyanate avec double liaison					
=N- s			-N= d		
Groupe d'amine dans un cycle avec double liaison					
-N= ⊙					

Ces groupes sont utilisés dans la génération des groupes structuraux, l'interaction entre le groupe -CH₃ et >CH₂ forme un groupe structurel, entre -CH₃ et >CH- un autre groupe structurel, etc.

2.3.2. Groupes de correction

Afin d'augmenter la fiabilité, la précision et la distinction entre les isomères spéciaux, on a introduit quelques groupes de correction, *cis*, *trans*, *ortho*, *meta* et *para*. On a fait la distinction entre ces groupes et les groupes de tableau 2.3 parce que la différence entre *cis* et *trans* est assurée par quatre groupes, c'est-à-dire trois interactions et pour éviter la complexité, on choisi seulement des groupes sans interactions.

2.4. Base de données

L'objectif principal de cette base de données est de servir au développement des modèles.

2.4.1. Collection des données

Un ensemble très large de base de données est disponible sur le marché, sous plusieurs formes (livres, fichier Excel, fichier SDF, sites internet, etc). Ces bases de données contiennent des données expérimentales et estimées des composés purs et complexes.

Dans le développement d'une technique de prédiction, on utilise seulement les données expérimentales pour l'obtention des contributions de groupes.

Dans les livres, les composés généralement sont classés par l'ordre alphabétique, comme la collection de (*Yaws, 2008*), le livre de (*Lide, 2010*), (*Dean, 1999*), etc. Ces livres contiennent, le nom du composé, formule, Cas number (The Chemical Abstracts Service Registry Number), masse molaire et la valeur de la propriété. Les propriétés disponibles dans ces livres sont généralement les propriétés physiques, thermodynamiques, du transport, etc.

La base de données sous forme SDF, Excel, etc sont aussi disponibles et offrent une large gamme des composés avec leurs propriétés, on peut citer celle de CAPEC (Common Attack Pattern Enumeration and Classification), DDB (dortmund data bank), etc, la classification se distingue d'une base à une autre, ainsi que la disponibilité des propriétés, etc.

Une autre forme de base de données, c'est les sites internet, comme par exemple NIST (National Institute of Standards and Technology), BioChemFinder, ChemSpider, etc. Chaque site offre un type de recherche qui permet l'accès rapide aux propriétés par rapport aux livres ou fichier SDF, la recherche se déroule via le nom, formule, Cas Number, InChI, SMILES, etc. A chaque fois, on sélectionne la propriété désirée, la figure 2.2 montre la page internet de la base de données NIST :

WebBook de Chimie NIST

Base de données standard de référence NIST numéro 69

Aller voir: [Options de recherche](#), [Modèles et outils](#), [Documentation](#), [Notes](#)

Show Crédits

Le NIST se réserve le droit à l'avenir de facturer l'accès à cette base de données.

Options de recherche [remonter](#)

Recherches générales	Recherches basées sur les propriétés physiques
<ul style="list-style-type: none"> Formule Nom Numéro d'enregistrement de CAS Réaction Auteur Structure 	<ul style="list-style-type: none"> Propriétés d'énergétique d'ion Énergies vibrationnelles et électroniques Poids moléculaire

Figure 2.2 : Page de site internet de la base de données NIST. (19/09/2014)

Recherche des données d'espèce par le nom chimique

Veuillez suivre les étapes ci-dessous pour effectuer votre recherche ([Aide](#)):

- Entrer un nom d'espèce chimique ou de modèle (par exemple, méthane, *2-héxène):
- Sélectionner l'unité désirée pour les données thermodynamiques:
 - ☒ SI ☐ basé-calorie
- Sélectionner le type(s) désiré des données:

Données Thermodynamiques	Autres Données
<input type="checkbox"/> Phase Gazeuse	<input type="checkbox"/> Spectre IR
<input type="checkbox"/> Phase condensée	<input type="checkbox"/> Spectre THz IR
<input type="checkbox"/> Changement de phase	<input type="checkbox"/> Spectre de Masse
<input type="checkbox"/> Réaction	<input type="checkbox"/> Spectre UV/Vis
<input type="checkbox"/> Energies ioniques	<input type="checkbox"/> Chromatographie en phase gazeuse
<input type="checkbox"/> Groupes ioniques	<input type="checkbox"/> Niveaux d'énergie vibrationnelle & électronique
	<input type="checkbox"/> Constantes des molécules diatomiques
	<input type="checkbox"/> Loi d'Henry
- Cliquer ici pour lancer la recherche:

Figure 2.3 : Type de la recherche dans la base de données NIST. (19/09/2014)

2.4.2. Structure de la base de données

Notre base de données est constituée principalement par : les hydrocarbures, les halogènes, les composés contenant l'oxygène, l'azote, le soufre, et composés multifonctionnels.

Notre base de données est sous forme d'un fichier Excel, dans lequel existe :

- ♣ ID (Nombre d'identification du composé dans notre base de données) ;
- ♣ Formule chimique du composé ;
- ♣ Nom du composé ;
- ♣ CAS number (The Chemical Abstracts Service Registry Number), quelque fois on l'utilise comme un outil de recherche d'un composé quelconque dans notre base de données ou bien dans les autres bases de données, soit dans l'internet ou dans les livres ;
- ♣ Masse molaire ;
- ♣ Nombre d'atomes ;
- ♣ InChI (International Chemical Identification) ;
- ♣ SMILES (Simplified Molecular-Input Line-Entry Specification) ;
- ♣ Classe du composé ;
- ♣ Valeurs expérimentales des propriétés (T_{eb} , T_c , P_c , V_c , T_f , etc) ;
- ♣ Groupes structuraux et les groupes de correction avec lesquels les composés chimiques vont être fragmentés.

Les composés chimiques dans notre base de données sont classés en grand catégorie en hydrocarbures, halogènes, composés contenant l'oxygène, l'azote, le soufre, et des composés multifonctionnels, ensuite, les composés chimiques dans chaque catégorie sont classés en famille chimique (alcane, alcène, acétone, aldéhyde, etc).

Ces familles chimiques sont ensuite classées selon :

1. Le type de liaison (simple, double, triple ou aromatique) ;
2. La structure du composé soit cyclique ou acyclique.

Cette classification ne permet pas seulement de faciliter la fragmentation, mais elle facilite aussi la vérification des données et de résultats.

2.4.3. Vérification de la base de données

Avant d'utiliser cette base de données, on doit vérifier que les valeurs collectées sont les valeurs expérimentales, cette étape est très délicate et cruciale surtout avec le grand nombre de données. La source principale de nos données est les livres de (*Yaws, 2008*), toutes les autres sources citées auparavant (les livres, les sites internet, etc), servent à la vérification.

2.4.4. Propriétés de la base de données

Le tableau 2.4 résume les propriétés disponibles dans notre base de données et leurs valeurs expérimentales en fonction des familles chimiques :

Tableau 2.4 : Nombre des valeurs expérimentales des propriétés dans la base de données.

	Hydrocarbure	Oxygène	Azote	Soufre	Halogènes	multi	Total
T_{eb}	838	702	207	101	379	85	2312
T_c	335	188	95	13	134	46	811
P_c	333	163	91	6	125	38	756
V_c	331	159	76	5	120	39	730
T_{fus}	419	299	123	60	192	29	1124
ΔH_{fus}	264	95	37	33	62	9	501
$\Delta_f H^\circ_{(gaz)}$	143	101	38	14	42	13	351

2.5. Développement des modèles

2.5.1. Température d'ébullition

La recherche d'un type de fonction ainsi qu'un modèle mathématique pour la prédiction des propriétés représente une étape très importante, parce que tout le travail est basé sur le modèle choisi (la régression, les corrections, l'analyse de résultats, validation, etc).

Les critères du choix du type de la fonction sont les suivants :

1. La fonction doit réaliser l'additivité des contributions ;
2. Elle doit présenter le meilleur ajustement possible des données expérimentales ;
3. Elle devrait fournir bonne aptitude à l'extrapolation et par conséquent, une large gamme d'application (*Marrero, et al., 2001*).

Dans la littérature plusieurs modèles ont été développés pour la prédiction des propriétés, un modèle logarithmique, exponentiel, rationnel, puissance, etc. La sélection du modèle se fait principalement selon l'approche choisie, autrement dit les groupes structuraux, parce que ces groupes sont le résultat de l'approche choisie, et on cherche un modèle qui relie de groupes structuraux à la propriété qu'on veut estimer.

La première étape dans la recherche d'un modèle est de sélectionner le type du modèle. Plusieurs outils et méthodes sont utilisés pour la sélection du modèle, les outils informatiques comme MATLAB, R, etc, peuvent aider au développement de l'équation du modèle en utilisant *les diagrammes de dispersion* qui sont présentés par la propriété d'un côté et les groupes structuraux de l'autre côté, le graphe aide à déterminer s'il existe une relation entre les deux variables et le type d'équation approprié (linéaire, non linéaire) (*Dodge, 2002*).

Pour éviter d'être perdus avec les méthodes de choix des modèles, on a pensé que la meilleure voie est d'essayer avec les mêmes modèles disponibles dans la littérature pour chaque propriété, et cette voie sera la plus courte qui nous ramène à trouver les meilleurs modèles.

Les modèles sélectionnés à partir de la littérature sont ; le modèle de *Marrero et Pardillo (1999)*, ce modèle a été développé avec la même approche employée dans ce travail, le modèle de *Marrero et Gani (2001)* et le modèle de *Nannoolal, et al., (2004)*.

Pour chaque modèle on réalise les étapes réalisées par ces auteurs pour l'obtention des constants d'ajustement (a, b, c, etc), les contributions, les valeurs estimées, ensuite on calcule les paramètres de validation comme la déviation standards (DS), l'erreur moyenne absolue (EMA), le pourcentage d'erreur moyenne absolu (PEMA) afin de pouvoir réaliser la comparaison entre ces modèles.

Dans ce travail, on va développer des modèles pour sept (7) propriétés (mentionnée dans le tableau 2.4), puis qu'il est trop long de décrire le développement de toutes ces propriétés, donc, on va décrire seulement la procédure pour la température d'ébullition et pour les autres propriétés elle sera la même.

✚ Le premier modèle est celui de *Marrero et Pardillo (1999)* qui ont utilisé la contribution des interactions de groupes comme approche, leur modèle est le suivant:

$$T_{eb} = M^a \sum_i N_i C_i + b \quad (2.1)$$

La première étape est de réaliser la régression non-linéaire avec nos groupes structuraux, ces groupes sont obtenus selon l'approche de l'interaction de groupes, ensuite, on a obtenu les constantes du modèle : $a=0,6891$ et $b=231,94$, alors, le modèle devient :

$$T_{eb} = M^{0,6891} \sum_i N_i C_i + 231,948 \quad (2.2)$$

Ensuite, on a réalisé la régression linéaire pour obtenir les valeurs des contributions de chaque groupe structural, et cette étape va nous permettre de calculer les valeurs estimées et les paramètres de validation.

✚ Le deuxième modèle est celui de *Marrero et Gani (2001)* qui ont utilisé la contribution de groupes mais effectuée à trois niveaux, dans notre cas, on a utilisé un seul niveau, alors, le modèle devient :

$$T_{eb} = a \ln(\sum_i N_i C_i) \quad (2.3)$$

Pour ce modèle la procédure a été la même avec une constante $a=224,564$, le modèle devient ;

$$T_{eb} = 224,564 \ln(\sum_i N_i C_i) \quad (2.4)$$

- ✚ Le troisième modèle est le modèle de *Nannoolal, et al., (2004)*, qui ont utilisé le nombre d'atomes présents dans la molécule sans compter l'hydrogène au lieu de la masse molaire dans leur modèle :

$$T_{eb} = \frac{\sum_i N_i C_i}{n^a + b} + c \quad (2.5)$$

Ainsi pour ce modèle, on a réalisé une régression non linéaire et on a obtenu les valeurs suivantes des constants du modèle, $a = 0,0053$, $b = -1,0056$ et $c = -899,5$.

Donc, le modèle devient ;

$$T_{eb} = \frac{\sum_i N_i C_i}{n^{0,0053} - 1,0056} - 899,5 \quad (2.6)$$

Ensuite, on a réalisé la régression linéaire pour obtenir les valeurs des contributions de chaque groupe structurel, en utilisant ces valeurs on peut obtenir les valeurs estimées des propriétés et on les compare avec les valeurs expérimentales.

L'étape suivante est de calculer les paramètres de validation, le pourcentage d'erreur moyenne absolu (PEMA), l'erreur moyenne absolue (EMA) et la déviation standard (DS), les formules de ces trois paramètres sont les suivantes :

Pourcentage d'erreur moyenne absolu (*Constantinou, et al., 1994*).

$$PEMA = \frac{1}{N} \sum_i \frac{|X_{i,est} - X_{i,exp}|}{X_{i,exp}} * 100\%$$

Erreur moyenne absolue (*Constantinou, et al., 1994*).

$$EMA = \frac{1}{N} \sum_i |X_{i,est} - X_{i,exp}|$$

Déviation standard (*Constantinou, et al., 1994*).

$$DS = \sqrt{\frac{\sum_i (X_{i,est} - X_{i,exp})^2}{N}}$$

Ces trois paramètres sont utilisés afin de comparer les trois modèles, on obtient ces trois paramètres après l'obtention des valeurs estimées par la régression des valeurs expérimentales, le tableau 2.5 résume les résultats de ces trois modèles :

Tableau 2.5: Résultats de la régression des trois modèles.

	<i>Marrero et Pardillo (1999)</i>	<i>Marrero et Gani (2001)</i>	<i>Nannoolal, et al., (2004)</i>
PEMA (%)	27,1735	2,26	32,98
EMA (K)	115,8595	8,74	106,58
DS (K)	121,8968	14,39	613,97

Selon les résultats de la régression mentionnés dans le tableau ci-dessus, on peut noter que l'erreur est très grande avec le modèle de *Marrero et Pardillo (1999)*, avec ce modèle la contribution des interactions de groupes qui a été utilisée. Cependant, les trois paramètres d'ajustement montre que ce modèle ne convient pas à nos données, c'est-à-dire il ne représente pas bien la relation entre la propriété (la température d'ébullition) et les groupes structurels.

En utilisant le modèle *Nannoolal, et al., (2004)* dans la régression de nos données, on a obtenu aussi une grande erreur, mais elle inférieure de celle de *Marrero et Pardillo (1999)*, mais elle est toujours inacceptable, là aussi le modèle ne représente pas bien la relation entre la propriété (la température d'ébullition) et les groupes structurels.

On note également que le modèle de *Marrero et Gani (2001)*, donne la plus faible erreur en comparant aux deux autres modèles, et les trois paramètres d'ajustement sont dans la gamme acceptable.

A partir de ces résultats, on peut dire que le type du modèle qui représente mieux la relation entre la propriété (T_{eb}) et les groupes structurels, c'est *la fonction logarithmique*, donc, on essaie de trouver un modèle de ce type qui peut minimiser l'erreur le plus possible et nous aide à obtenir les valeurs estimées les plus proches possible aux valeurs expérimentales.

Après la recherche d'un modèle de type logarithmique, on a trouvé le modèle suivant :

$$\frac{T_{eb}}{K} = -a + b \ln \left(c + \sum_i N_i C_i \right) \quad (2.7)$$

Où a et b sont les constants du modèle, et K représente l'unité de la propriété (kelvin)

Le choix de ce modèle est basé principalement sur l'expérience acquise et l'analyse des méthodes déjà mentionnées dans le chapitre 01, ainsi que les résultats obtenus de la régression des trois modèles mentionnés au-dessus.

L'obtention des constants du modèle a été réalisée en se basant sur deux procédures de régression, la première est la régression non-linéaire, qui a été utilisée pour obtenir les paramètres a et b, et la deuxième est la régression linéaire, elle a été utilisée pour l'obtention de la contribution des groupes structuraux.

2.5.2. Modèles proposés des autres propriétés

La même procédure décrite au-dessus pour le développement du modèle de la température d'ébullition a été employée pour l'obtention des autres modèles pour les autres propriétés, les modèles proposés pour les autres propriétés sont les suivants :

$$\text{Température critique :} \quad T_c = \frac{T_{eb}}{\left(c + \sum_i N_i C_i\right)^2 - a} \quad (2.8)$$

$$\text{Pression critique :} \quad \frac{P_c}{(\text{bar})} = \frac{a}{\left(c + \sum_i N_i C_i\right)} - b \quad (2.9)$$

$$\text{Volume critique :} \quad \frac{V_c}{(\text{cm}^3 / \text{mol})} = \left(c + \sum_i N_i C_i\right) \quad (2.10)$$

$$\text{Température de fusion :} \quad \frac{T_f}{(\text{K})} = a * \left(c + \sum_i N_i C_i\right)^b \quad (2.11)$$

$$\text{Enthalpie de fusion :} \quad \frac{\Delta H_{fus}}{(\text{kJ} / \text{mol})} = T_f * \left(c + \sum_i N_i C_i\right)^2 \quad (2.12)$$

$$\text{Enthalpie standard de formation:} \quad \frac{\Delta_f H^\circ_{(gaz)}}{(\text{kJ} / \text{mol})} = \left(c + \sum_i N_i C_i\right) \quad (2.13)$$

Les résultats de cette régression (les paramètres structuraux) sont présentés dans l'annexe 01.

2.6. Ajustement du modèle

Après la préparation des valeurs expérimentales des propriétés et le choix des groupes structuraux ainsi que le modèle reliant ces deux, l'étape suivante est l'ajustement du modèle et l'obtention des paramètres de ce modèle (les contributions).

2.6.1. Méthodes de régression

Historiquement, le mot *régression* revient au *Sir Francis Galton (1885)*, travaillant sur l'hérédité, il a cherché à expliquer la taille des fils en fonction de celle des pères, ses résultats lui permettent de considérer sa théorie de « *regression toward mediocrity* », (régression vers la médiocrité).

La régression est une méthode avec laquelle on étudie ou on mesure la relation entre deux variables, Y la variable à expliquer (dépendante). Dans notre travail, Y représente la propriété P, et X variable explicative (indépendante) (ou X plusieurs variables explicatives), qui représente (ent) dans notre travail les groupes structurels, cette relation nous permet de prédire les valeur de Y en fonction de X, (de prédire la valeur de la propriété P en fonction des groupes structurels).

2.6.1.1. Régression linéaire

La régression linéaire est la méthode avec laquelle on cherche à trouver une fonction f qui relie deux variables p et c, dans ce cas ce type de régression est appelée *régression linéaire simple*.

Si la relation à chercher est entre plusieurs variables, c'est-à-dire, entre p et $c_{i1}, c_{i2}, \dots, c_{ip}$, ce type de régression est appelée *régression linéaire multiple*, tels que :

$$p_i \approx f(c_{i1} + c_{i2} + \dots + c_{ip})$$

Il existe deux modèles mathématiques dans la régression :

1. Modèle exacte (modèle déterministe) , $m_i = a + bc_i$ dans le cas où par exemple on cherche à trouver la masse molaire en fonction des groupes structurels.
2. Modèle aléatoire (modèle stochastique)

$$p_{a \times 1} = c n_{l \times 1} + \varepsilon_{a \times 1}$$

Où ;

- p est un vecteur aléatoire de dimension a ;
- c est une matrice de taille a×l connue ;
- β est le vecteur de dimension l des paramètres inconnus du modèle ;
- ε est le vecteur centré, de dimension a, des erreurs (*Cornillon, et al., 2011*).

2.6.1.2. Régression non-linéaire

La différence principale entre la régression linéaire et non-linéaire (modèle linéaire et non-linéaire) réside dans la linéarité de la courbe qui représente la relation entre le variable dépendant (Y) et les variables indépendants (X_1, X_2, \dots, X_n).

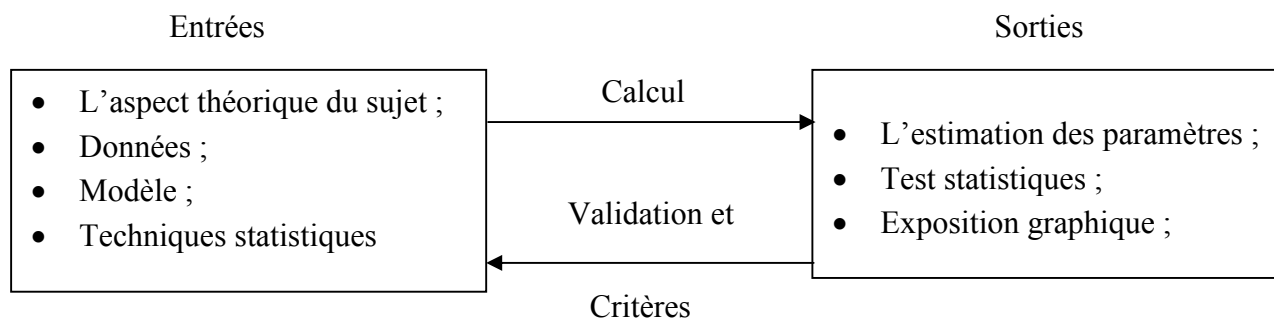


Figure 2.4 : Schéma illustrant le processus de la régression.

2.6.2. Outils informatiques

L'évolution dans le domaine informatique ne permet pas seulement d'améliorer la qualité de la recherche scientifique mais aussi de développer des nouvelles branches et spécialité comme par exemple la *chimométrie*.

L'influence de ce développement sur les méthodes de prédiction est présentée dans l'augmentation du nombre de données et de paramètres, la méthode de *Lydersen (1955)* a été développée par 35 groupes structurels et 557 valeurs expérimentales de la température critique, en 2004, *Nannoolal, et al.*, ont développé leur modèle par 212 groupes structurels et 2800 valeurs expérimentales.

Les outils informatiques ne sont pas uniquement utilisés dans le développement des méthodes de prédiction, mais aussi dans l'utilisation de ces méthodes. *Nannoolal, et al., 2004* ont développé un logiciel appelée *Artist* dans lequel ils ont intégré plusieurs méthodes d'estimation.

Une vaste gamme d'outils informatiques est disponible dans le marché, leur but principal est de faciliter l'utilisation des techniques statistiques, on cite par exemple MATLAB, XLstat, Knime, R, RapidMiner, Weka, etc. Les performances de ces logiciels sont bien distinctes, selon les applications intégrées dans chacun d'eux.

Dans ces logiciels, il existe l'option de *l'ajustement d'un modèle (fit model)*, cette option permet de choisir le meilleur modèle qui relie la propriété étudiée et les paramètres choisis (groupes structuraux).

Dans notre travail, on a besoin d'un logiciel performant afin de permettre le traitement d'une grande base de données avec un nombre de groupes structuraux considérable, (ce qui pose un problème pour quelques logiciels), minimiser le temps de calcul et d'offrir tout les paramètres de validation et de vérification du modèle, PEMA, EMA, DS, etc.

Le logiciel qui a été utilisé dans ce travail pour l'obtention des constants du modèle a et b en utilisant la régression non-linéaire est R.

R est un logiciel de manipulation de données, calcul et d'exposition graphique. Il est effectif dans la manipulation et le stockage de données, il représente une large collection, cohérente et intégrée d'outils pour l'analyse des données, etc. Ce logiciel est utilisé dans plusieurs domaines, *la statistique* est parmi eux. Dans ce logiciel plusieurs techniques classiques et modernes sont implémentées, certaines sont construites dans l'environnement de base de R, et d'autres sont fournies sous forme des packages.

2.6.2.1. Transformation de la base de données

Notre base de données est réalisée sous forme d'un fichier Excel, alors, la première étape est de transférer ce fichier Excel.XLSX à un fichier Excel.CSV(DOC), cette étape est réalisée par l'utilisation des options suivantes dans l'Excel :

[Botton office-> Enregistrer sous-> Type (CSV(DOC))].

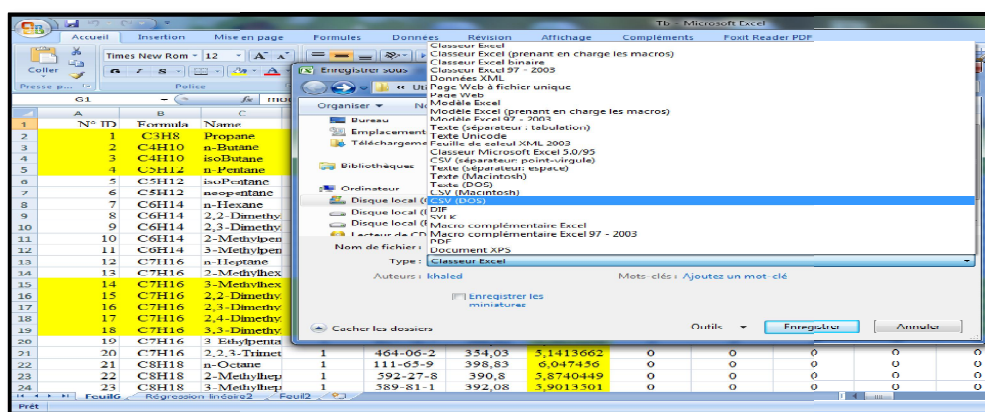


Figure 2.5: Transformation de la base de données d'une feuille XLXS à une feuille CSV.

2.6.2.2. Création d'un fichier en R

La deuxième étape est de former un fichier dans le logiciel R pour réaliser les étapes de la régression, ce fichier contient les commandes avec lesquelles on réalise la régression ainsi que tous les paramètres d'ajustement permettant de juger la fiabilité et la précision du modèle (PEMA, EMA, DS, R^2 , etc). On peut aussi ajouter des commandes pour l'obtention des graphes.

La troisième étape est d'importer le fichier qui contient la température d'ébullition ou bien les autres propriétés avec les groupes structuraux au logiciel R, en suivant ces étapes ;[*Envirenment->Import dataset->From Text File*].

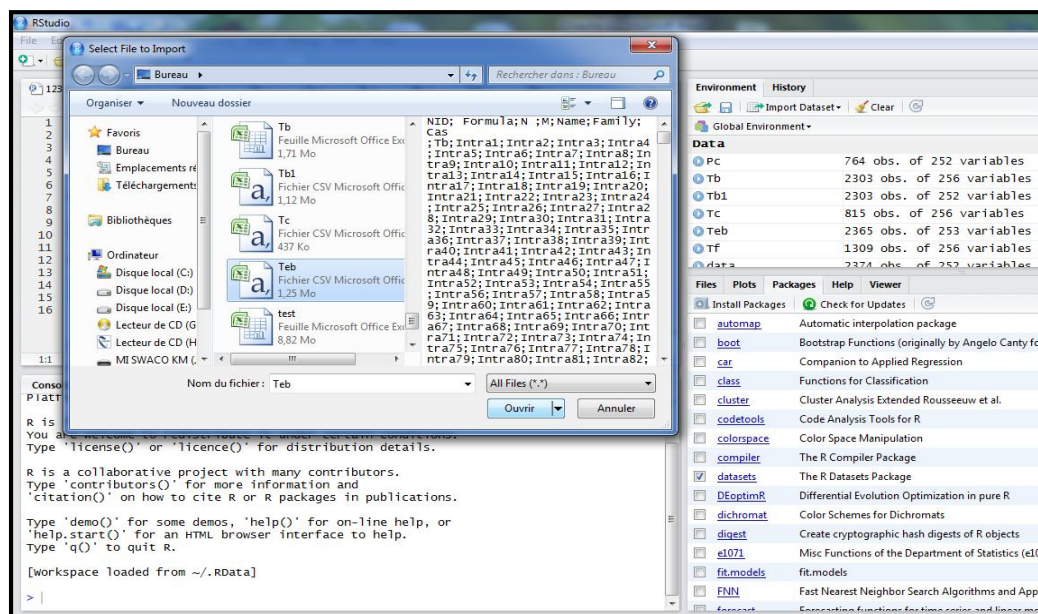


Figure 2.6 : Importation de la base de données dans le logiciel R.

Le logiciel R contient *des packages*, chaque package est désigné pour un type de calcul, pour la régression non-linéaire on a utilisé le package [*Mosaic*]. Le contenu de ce package est les techniques mathématiques et statistiques. Dans ce package il existe des commandes pour préciser l'opération avec laquelle on veut obtenir nos résultats. Dans cette étape et pour réaliser la régression non-linéaire on a utilisé la commande *fitModel*.

L'expression suivante représente le début du programme dans le logiciel R.

```
f<-fitModel(log(Intra1 +Intra2+...+Intra244+Intra245)~(Tb+a)/b,data=data)  
Model (f)
```

Le premier lettre représente le nom de notre fonction, suivi par la commande (fitModel), ensuite, on met les paramètres du modèle (groupes structurels) d'un coté et de l'autre coté la propriété sous forme du modèle proposé avec les constants, où le but de cette opération est la résolution de l'équation et l'obtention des valeurs des constants a, b, c, etc. L'objectif de la commande *model(f)* est d'afficher les valeurs des constants et la forme du modèle.

La forme mathématique du modèle proposé, pour la température d'ébullition est l'équation (2.7) :

$$\ln \left(c + \sum_i N_i C_i \right) = \left(\frac{T_{eb} + a}{b} \right) \quad (2.7)$$

Après la réalisation de la régression on a obtenu les valeurs suivantes des constants du modèle : a=84,565 et b=265,476.

2.6.2.3. Obtention des contributions de groupes

L'étape de la régression linéaire est précédée par la *linéarisation* de notre modèle, c'est-à-dire on le transforme d'un modèle non-linéaire à un modèle linéaire. Alors, on met les constants du modèle a et b avec la valeur de la température d'ébullition d'un côté, et on met les paramètres du modèle (groupes structurels) de l'autre côté, donc, la forme de l'équation (2.7) devient :

$$\exp \left(\frac{T_{eb} + a}{b} \right) = \left(c + \sum_i N_i C_i \right) \quad (2.7)$$

On peut réaliser la régression linéaire aussi avec R mais cette fois avec la commande *[lm]* cette commande utilise la méthode des moindre carrée pour l'obtention des résultats, et cette commande se trouve dans le package *[stats]*.

Avant de réaliser la régression linéaire pour obtenir les valeurs de contributions ainsi que la constante c avec notre base de données on doit diviser la base en deux ensembles, 90% pour le développement (obtention des valeurs de contributions de groupes) et 10% pour le tester la fiabilité et la précision du modèle.

Dans la régression linéaire, il est mieux de faire la régression de tous les paramètres simultanément pour obtenir la valeur optimale de la fonction objective au lieu de faire la régression pour les groupes structurels.

Ensuite on réalise la deuxième régression avec les contributions des groupes structurels obtenues de la première régression pour obtenir les contributions des groupes de correction, c'est-à-dire régression *pas à pas*.

Après la réalisation de la régression, on obtient les valeurs des contributions des groupes structurels, les programmes de la régression de chaque propriété sont présentés dans l'annexe 03.

2.6.3. Validation des modèles

La validation est basée sur des hypothèses, ces dernières concernent principalement la base de données et le modèle choisi (*Chattefuee, et al., 2006*).

Concernant la base de données, on a utilisé tous les moyens pour assurer que ces valeurs sont des valeurs expérimentales réelles, et qu'elles sont présentées dans la base de données avec leurs valeurs correctes, parce que plusieurs problèmes peuvent se poser durant le développement due à la base de données, on cite à titre d'exemple :

- Une valeur incorrecte de la propriété, par exemple au lieu de 485,15 la valeur 48515 perturbe tous les résultats ;
- Une valeur répétée ;
- Une fragmentation incorrecte ou répétée d'un composé, etc,

Donc, nos sources ont été nombreuses et variées, livres, sites internet, bases de données Excel, etc, pour éviter les problèmes dus à la base de données, et tout ça a été bien expliqué dans la préparation de la base de données.

2.6.3.1. R^2 et R^2 ajusté

On mesure avec R^2 (le coefficient de la détermination) la précision de la droite de régression. Il mesure la proportion de la variation dans la variable Y, qui s'explique par la présence de la variable X (ou par la droite de régression). R^2 est le rapport entre la variation expliquée et la variation totale. Il est, en principe, souhaitable que la valeur de R^2 soit très grande, car plus la valeur de R^2 est grande, plus celle de la variation inexpliquée est petite (*Dodge, 2002*).

$$R^2 = 1 - \frac{SCE}{SCT} \quad SCE = \sum (Y_i - \hat{Y}_i)^2 \quad SCT = \sum (Y_i - \bar{Y})^2$$

Où :

SCE somme carrée des erreurs ;

SCT somme carrée total (la différence entre la valeur réelle et la valeur moyenne) ;

Y_i Valeur expérimentale;

\hat{Y}_i Valeur estimée ;

\bar{Y} La moyenne des valeurs expérimentales (*Chattefuee, et al., 2006*).

$$R^2_{ajusté} = 1 - \frac{(1 - R^2)(N - 1)}{(N - P - 1)}$$

R^2 le coefficient de la détermination ;

N le nombre de donnée de la variable dépendants, le nombre des valeurs expérimentales ;

P le nombre des variables indépendants, le nombre des groupes structurels.

2.6.3.2. Analyse des résidus

a. Analyse graphique des résidus

Les représentations graphiques satisfaisant au besoin de l'analyse de voir le comportement des données, des ajustements, des mesures de diagnostic et des résidus et donc de saisir les caractéristiques inattendues ainsi que les régularités familières (*Dodge, 2002*).

Deux types de graphes ont été employés pour la validation des modèles :

- La comparaison entre les valeurs expérimentales et estimées ;
- La dispersion des résidus en fonction des valeurs expérimentales et estimées.

La forme de la dispersion des points dans le graphe des résidus peut montrer la qualité de la modélisation, et elle peut nous aider à juger le modèle.

b. Analyse quantitative des résidus

Les résidus sont la différence entre les valeurs observées (expérimentales) et les valeurs estimées, L'examen des résidus constitue une étape primordiale de la régression linéaire (*Cornillon, et al., 2011*).

Les paramètres de validation utilisés pour juger les modèles sont les mêmes utilisés pour choisir le modèle de chaque propriété (PEMA, EMA, DS).

$$\text{Résidu} = \text{expérimentale} - \text{estimée}$$

Chapitre 03 :

Résultats et discussion

3.1. Introduction

Deux types de représentations des résultats sont illustrés dans ce chapitre les tableaux et les graphes, dont le but est de discuter les performances de chaque modèle. Pour chaque propriété il y a des tableaux dans lesquels il existe PEMA (pourcentage d'erreur moyenne absolu), EMA (Erreur moyenne absolue) et DS (Déviation standard) pour les différentes classes ainsi que le total de la base de données, et d'autres tableaux pour la comparaison entre les modèles proposés dans ce travail et les modèles déjà publiés.

Concernant les graphes, un pour la comparaison entre les valeurs expérimentales et les valeurs estimées de chaque propriété, et les deux autres graphes sont utilisés pour la comparaison entre les résidus et les valeurs expérimentales d'un part, et les résidus et les valeurs estimées de l'autre part.

On sait qu'il existe une relation structure-propriété, mais peut-on obtenir la propriété en fonction de la structure par un modèle mathématique ? Donc, le but de l'analyse des résultats est de répondre aux questions suivantes :

- Est ce que l'approche choisie représente bien cette relation ?
- Est ce que la base de données sélectionnée représente l'ensemble des composés organiques ?
- Est ce que le modèle choisi montre cette relation mathématiquement ?

Les paramètres dans les tableaux dans ce chapitre (PEAM, EMA et DS) sont utilisés pour nous aider à répondre aux trois questions ci-dessus.

3.2. Température d'ébullition

Les résultats du tableau 3.1 montrent les performances du modèle de la température d'ébullition pour la base de données total et pour chaque classe.

Tableau 3.1 : Performances du modèle de la CIG pour la (T_{eb}) pour les différentes classes.

Modèle proposé	$\frac{T_{eb}}{(K)} = -84,565 + 265,476 \ln \left(1,8214 + \sum_i N_i C_i \right)$			
Classes	Nombre de données	PEMA (%)	EMA (K)	DS (K)
Hydrocarbures	838	0,26	1,22	3,44
Contient l'oxygène	702	0,59	2,85	6,63
Contient l'azote	207	2,13	8,91	11,92
Contient le soufre	101	1,08	4,32	7,01
Halogènes	379	1,38	4,75	8,53
Multifonctionnels	85	2,07	9,10	12,14
Ensemble de test	230	2,58	9,82	12,603
Total	2312	1,6	6,44	9,12
$R^2 = 0,98$		$R^2 \text{ ajusté} = 0,978$		

D'après le tableau, on note que la classe la plus précise est celle des hydrocarbures, ceci montre que les groupes structuraux choisis pour représenter la température d'ébullition en fonction de la structure moléculaire des hydrocarbures sont les meilleurs ainsi que ces groupes restent avec toutes les autres classes. Donc, cette classe a le minimum nombre de groupes structuraux. Concernant le modèle, on peut dire qu'il a montré bien cette relation, sans oublier que cette classe est la plus grande dans la base de données et ça représente un autre facteur qui améliore la précision des résultats.

La deuxième classe qui vient après les hydrocarbures est celle des composés contenant l'oxygène, ensuite, les composés contenant le soufre, les halogènes, les composés multifonctionnels, et enfin les composés contenant l'azote.

On note que la précision augmente avec le nombre de données dans chaque classe, donc, on sait maintenant que le nombre de données utilisés dans le développement du modèle a une influence sur la précision.

Concernant la classe des composés qui ont été sélectionnés pour tester la fiabilité du modèle, on note que la précision diminue avec cette classe, mais toujours elle est dans la gamme acceptable.

Deux autres paramètres de validation, ce sont, le coefficient de la détermination R^2 et R^2 ajusté, la valeur de R^2 est très proche de 1 ce qui prouve que le modèle est très précis, le cas est le même pour R^2 ajusté, sa valeur prouve la précision du modèle.

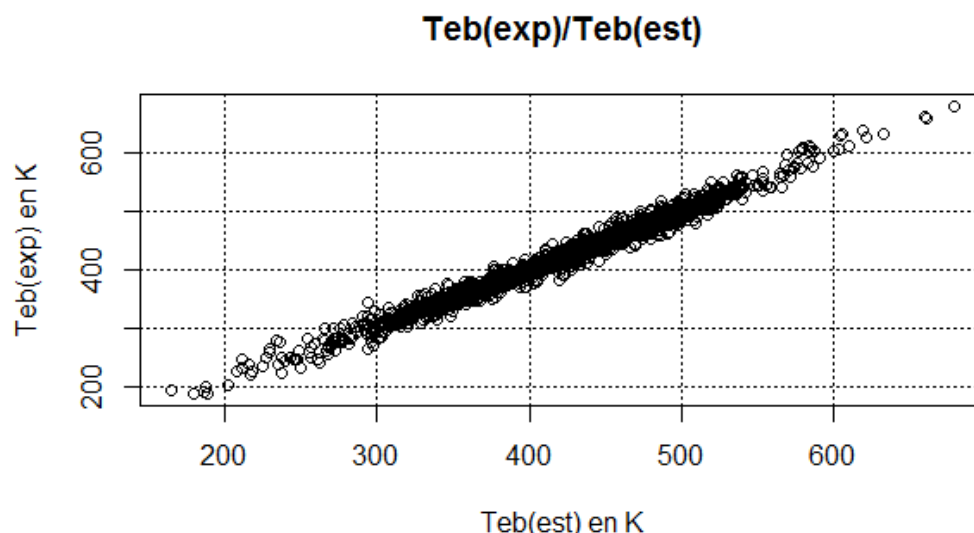


Figure 3.1 : Comparaison entre les valeurs expérimentales et estimées de (T_{eb}).

La figure 3.1 représente la comparaison entre les valeurs expérimentales et estimées de la température d'ébullition, alors, on voit que la forme est une ligne, donc, on peut dire que la précision est bonne.

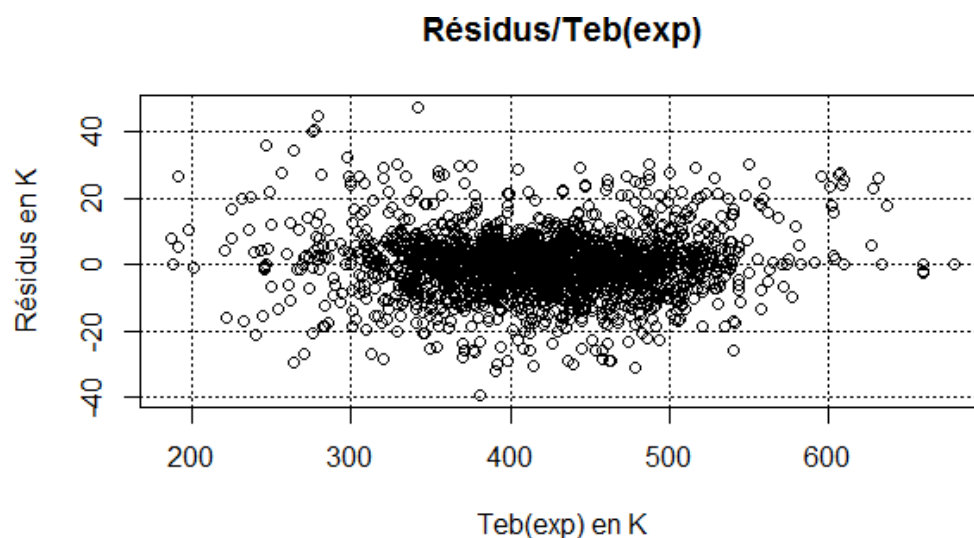


Figure 3.2 : Dispersion des résidus de la (T_{eb}) en fonction des (T_{eb}) expérimentales.

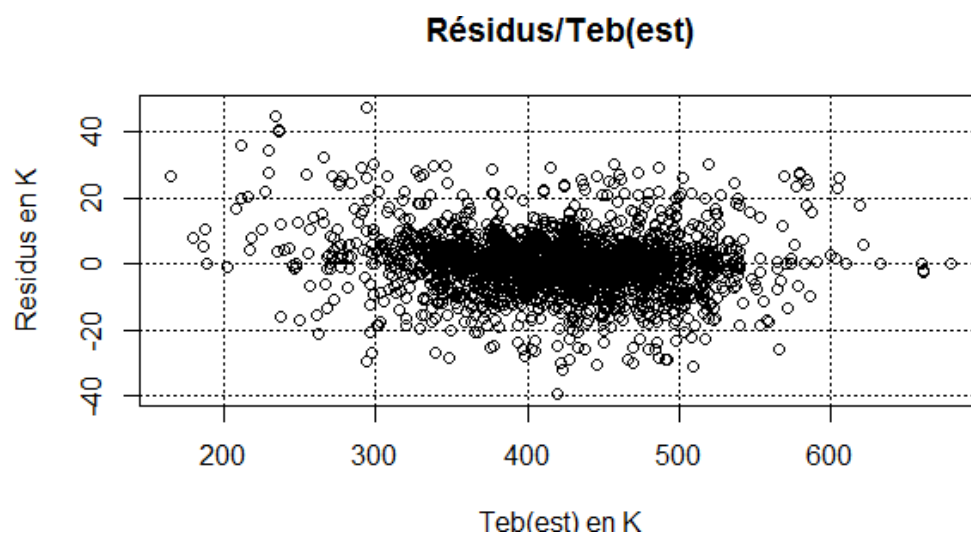


Figure 3.3 : Dispersion des résidus de la (T_{eb}) en fonction des (T_{eb}) estimées.

Les deux figures ci-dessus, servent à la comparaison entre les résidus et les valeurs expérimentales d'un part, et entre les résidus et les valeurs estimées de l'autre part.

A partir de ces deux figures, on peut noter qu'il existe une ressemblance, ce que permet de conclure que le modèle est le meilleur pour représenter la relation entre la propriété et les groupes structuels.

On peut voir aussi que la dispersion se déroule aux environs de l'axe des abscisses, c'est-à-dire autour de zéro, et ceci nous permet de dire que la précision est excellente.

Le tableau 3.2 sert à la comparaison entre le modèle présenté dans ce travail et les autres modèles déjà publiés, y compris le modèle de la même approche.

La comparaison entre le PEMA, EMA, et la DS de chaque modèle est en fonction du nombre de données et de groupes structuels.

Tableau 3.2 : Comparaison entre les performances du modèle de la CIG et des modèles précédents pour la (T_{eb}).

Méthodes	Nombre de données	Nombre de groupes	PEMA (%)	EMA (K)	DS (K)
<i>Constantinou et Gani (1994)</i>	392	78 1 ^{er} ordre	2,04	7,71	10,48
		43 2 ^{eme} ordre	1,42	5,35	7,7
<i>Marrero et Pardillo (1999)</i>	507	174	1,34	4,99	7,35
<i>Marrero et Gani (2001)</i>	1794	176 1 ^{er} ordre	1,80	7,90	11,11
		106 2 ^{eme} ordre	1,40	6,38	8,96
		50 3 ^{eme} ordre	1,40	5,89	8,09
<i>Nannoolal, et al., (2004)</i>	2812	212	NA	6,52	ND
<i>Wang, et al., (2008) CPG</i>	417	91	1,00	4,2	ND
<i>Ce travail</i>	2312	228	1,6	6,44	9,12

Le modèle qui a présenté la meilleure précision est notre modèle, soit en fonction de PEMA, EMA ou DS les valeurs de ces trois paramètres sont les meilleurs pour le modèle choisi, et le modèle qui suit, est celui de *Nannoolal, et al., (2004)*, parce que dans le développement de ce modèle ils ont utilisé le plus grand nombre de données, ainsi on note que le nombre de groupe structuraux est inférieur à notre méthode, mais, cette méthode a un problème avec la distinction entre les isomères.

Pour la méthode de *Marrero et Gani (2001)*, ils ont utilisé le plus grand nombre de groupe structuraux, ce facteur diminue la description de la molécule si elle n'est pas utilisée dans de développement du modèle.

Malgré que le modèle de *Wang, et al., (2008)* a offert la plus basse erreur, mais, dans le développement de ce modèle ils ont utilisé le plus bas nombre de données, ce qui influe la précision du modèle avec les nouveaux composés qu'on veut prédire.

3.3. Température critique

Tableau 3.3 : Performances du modèle de la CIG pour la (T_c) pour les différentes classes.

Modèle proposé	$T_c = \frac{T_{eb}}{\left(6,8356 + \sum_i N_i C_i\right)^2 - 46,1301}$			
Classes	Nombre de données	PEMA (%)	EMA (K)	DS (K)
Hydrocarbures	335	0,15	1,06	4,30
Contient l'oxygène	188	0,40	2,54	5,06
Contient l'azote	95	0,86	5,18	8,66
Contient le soufre	13	0,60	3,38	5,41
Halogènes	134	0,44	2,32	3,83
Multifonctionnels	46	1,27	8,36	15,11
Ensemble de test	81	1,47	8,91	16,28
Total	811	0,74	4,39	7,57
$R^2 = 0,994$		$R^2 \text{ ajusté} = 0,992$		

Pour le modèle de la température critique on note que la précision est très bonne avec les hydrocarbures, ensuite, les composés contenant l'oxygène, les halogènes, les composés contenant le soufre, l'azote, les composés multifonctionnels, et enfin, les composés qu'ont été choisis pour tester le modèle.

A partir du tableau, on constate clairement que la précision augmente avec l'augmentation du nombre de données qui a été utilisé pour le développement du modèle, donc, le même cas se répète ici comme avec la température d'ébullition.

Les valeurs de R^2 et R^2 ajusté dans le tableau ci-dessus, représente une autre preuve de la parfaite précision du modèle proposé, les deux valeurs sont presque identiques et les deux au même temps sont proches de 1.

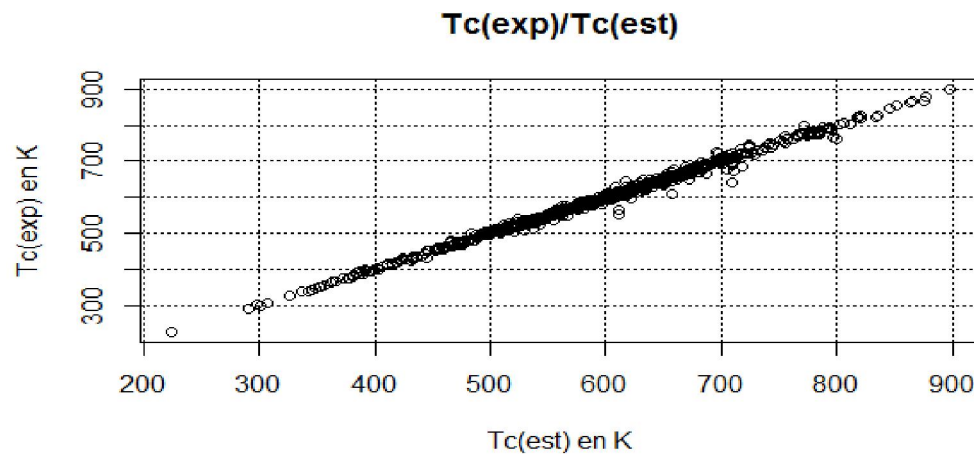


Figure 3.4 : Comparaison entre les valeurs expérimentales et estimées de (T_c).

La figure 3.4 se ressemble à une ligne, ce qui montre que les valeurs expérimentales et estimées sont très proche les unes aux autres, ceci représente un autre moyen pour confirmer la précision du modèle qui permet de prédire la (T_c).

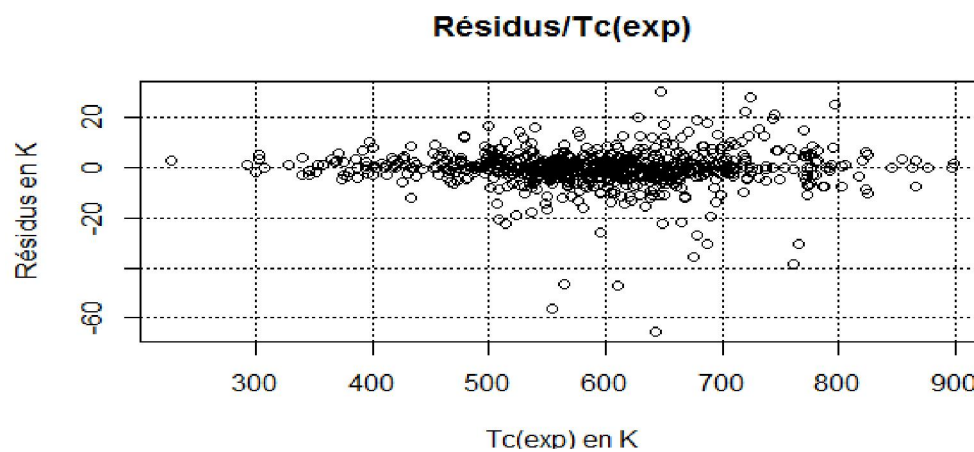


Figure 3.5 : Dispersion des résidus de la (T_c) en fonction des (T_c) expérimentales.

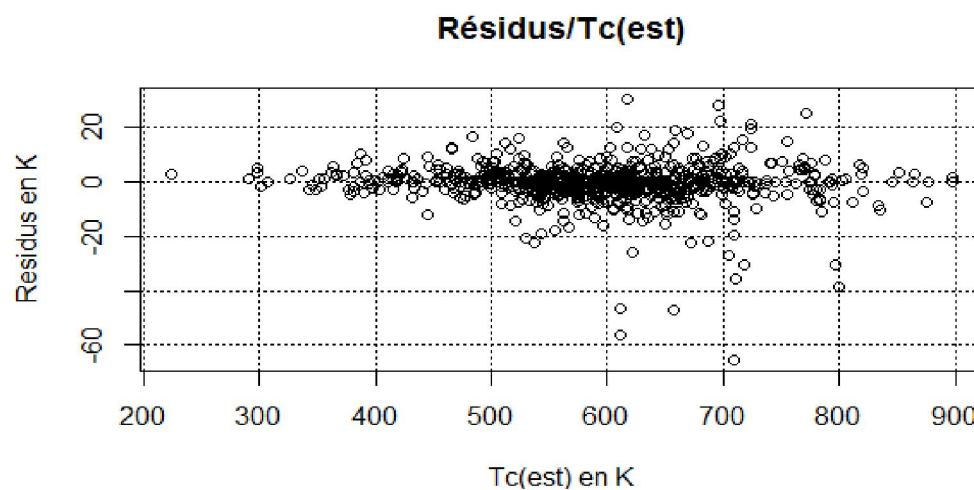


Figure 3.6 : Dispersion des résidus de la (T_c) en fonction des (T_c) estimées.

A partir des deux figures, on constate que la dispersion est presque la même, ce qui montre que le modèle choisi a bien représenté la propriété, et cette dispersion est toujours autour de l'axe des abscisses, donc, on conclut que la précision est excellente.

Tableau 3.4 : Comparaison entre les performances du modèle de la CIG et des modèles précédents pour la (T_c).

Méthodes	Nombre de données	Nombre de groupes	PEMA (%)	EMA (K)	DS (K)
<i>Constantinou et Gani (1994)</i>	285	67 1 ^{er} ordre	1,62	9,12	12,96
		41 2 ^{eme} ordre	0,85	4,85	6,98
<i>Marrero et Pardillo (1999)</i>	491	173	0,48	2,79	4,39
<i>Wen et Quiang (2001)</i>	471	140	0,73	4,35	ND
<i>Marrero et Gani (2001)</i>	783	125 1 ^{er} ordre	1,40	8,75	17,25
		78 2 ^{eme} ordre	0,90	5,67	8,50
		31 3 ^{eme} ordre	0,80	4,93	6,99
<i>Wang, et al., (2008) CPG</i>	467	73	1,10	6,6	ND
<i>Nannoolal, et al., (2006)</i>	588	218	0,74	4,31	ND
<i>Ce travail</i>	811	179	0,74	4,39	7,57

La plus basse erreur entre tous les modèles présentés dans le tableau 3.4 est notre modèle, ce que signifie que notre modèle est le plus précis. Cette précision vient principalement du nombre de données. En effet, dans notre travail on a utilisé le plus grand nombre, et un nombre de groupes structuraux qui est un peu bas par rapport aux autres modèles.

3.4. Pression critique

Tableau 3.5 : Performances du modèle de la CIG pour la (P_c) pour les différentes classes.

Modèle proposé	$\frac{P_c}{(bar)} = \frac{239,15}{\left(2,8529 + \sum_i N_i C_i\right)} - 1,855$			
Classes	Nombre de données	PEMA (%)	EMA (bar)	DS (bar)
Hydrocarbures	333	1,84	0,56	0,91
Contient l'oxygène	163	2,89	1,02	1,54
Contient l'azote	91	2,99	1,23	1,90
Contient le soufre	6	4,67	1,88	2,43
Halogènes	125	3,46	1,49	2,04
Multifonctionnels	38	3,79	1,55	2,38
Ensemble de test	75	4,37	1,65	2,29
Total	756	2,60	0,95	1,53
$R^2 = 0,973$		$R^2 \text{ ajusté} = 0,965$		

Les résultats illustrés dans le tableau 3.5 montrent que la précision est meilleure pour les hydrocarbures, suivis par les composés contenant l'oxygène, ensuite, les composés contenant l'azote, les halogènes, les multifonctionnels, les composés contenant le soufre, on note cette fois ci que la dernière classe est celle des composés contenant le soufre, qui ont la plus basse précision et ceci est due au nombre de données qui est aussi le plus bas, même les composés qui sont hors de la base de données, ils ont une erreur faible de cette classe, parce qu'ils ont un nombre de données plus grand.

Le coefficient de la détermination R^2 ainsi que R^2 ajusté sont proche l'un à l'autre et les deux sont proches de 1. Ces valeurs sont suffisantes pour dire que le modèle a offert une bonne précision.

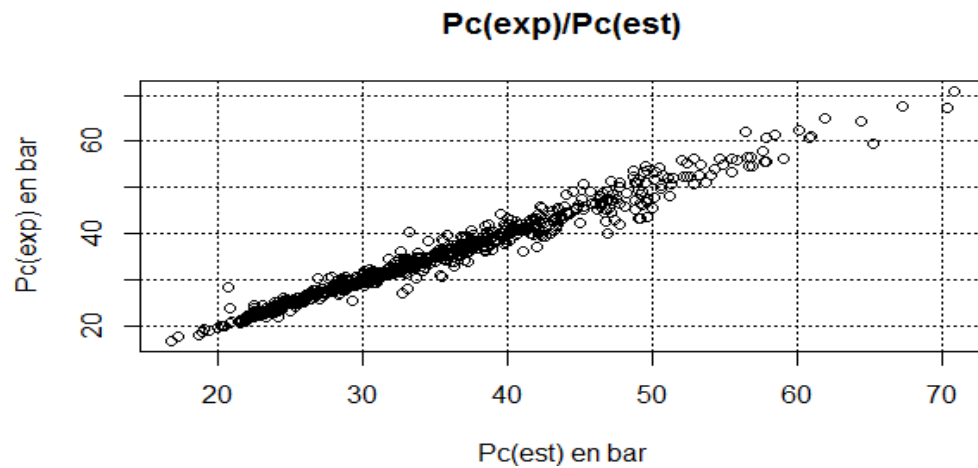


Figure 3.7 : Comparaison entre les valeurs expérimentales et estimées de (P_c).

La figure 3.7 se ressemble à la forme d'une ligne, ce qui prouve que le modèle est précis, et représente bien la propriété.

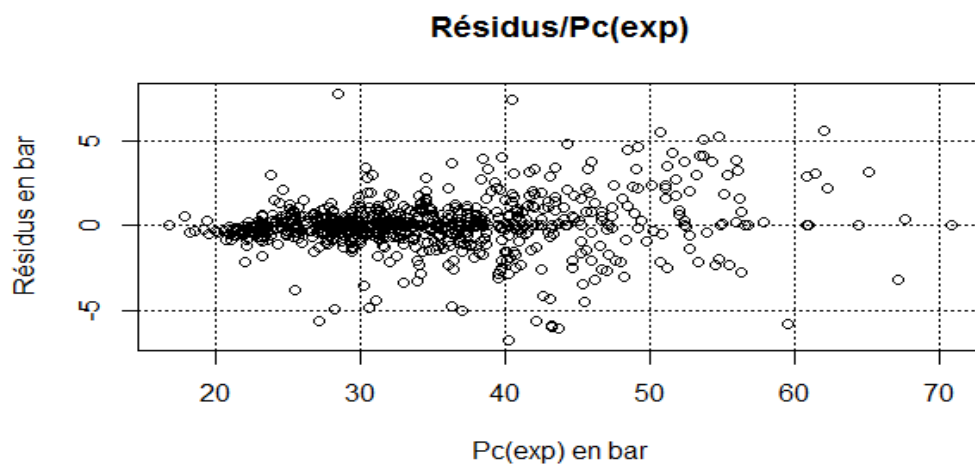


Figure 3.8 : Dispersion des résidus de la (P_c) en fonction des (P_c) expérimentales.

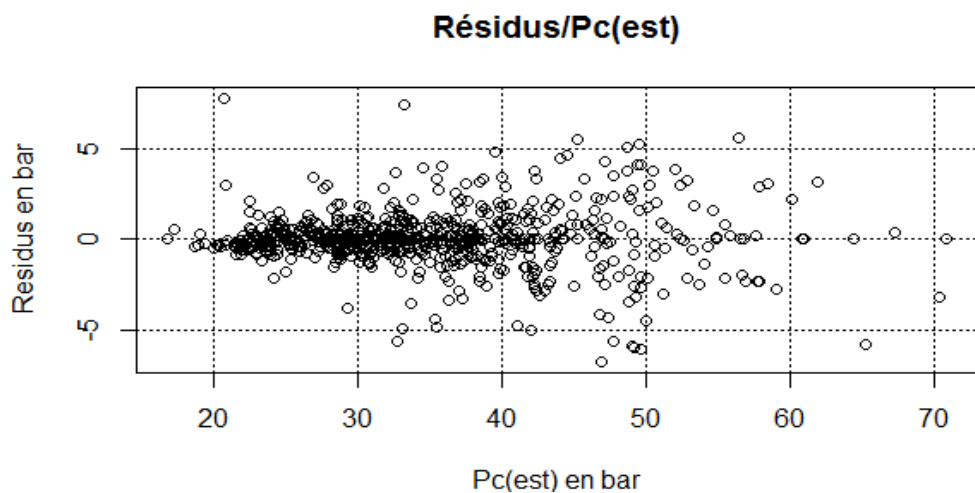


Figure 3.9 : Dispersion des résidus de la (P_c) en fonction des (P_c) estimées.

La dispersion des points dans les deux figures est presque identique, ceci signifie que le modèle est adéquat à la propriété, ainsi que cette dispersion est à l'entour de l'axe des abscisses, ce qui montre que l'erreur est petite pour la plus part de la base de données.

Tableau 3.6 : Comparaison entre les performances du modèle de la CIG et des modèles précédents pour la (P_c).

Méthodes	Nombre de données	Nombre de groupes	PEMA (%)	EMA (bar)	DS (bar)
<i>Constantinou et Gani (1994)</i>	269	67	3,72	1,43	2,34
		40	2,89	1,13	2,02
<i>Marrero et Pardillo (1999)</i>	445	165	2,92	1,06	1,70
<i>Wen et Quiang (2001)</i>	408	140	2,72	0,96	ND
<i>Marrero et Gani (2001)</i>	775	124 1 ^{er} ordre	2,90	1,02	1,73
		79 2 ^{eme} ordre	2,60	0,87	1,53
		32 3 ^{eme} ordre	2,30	0,79	1,39
<i>Nannoolal, et al., (2006)</i>	486	218	2,96	0,9982	ND
<i>Ce travail</i>	756	175	2,60	0,95	1,53

Malgré que le modèle de *Marrero et Gani (2001)* a offert la plus basse erreur, mais avec un nombre groupe structurels très grand, ceci influe la précision du modèle avec les nouveaux composés si on veut prédire leurs propriétés. Après ce modèle, notre modèle a la faible d'erreur avec un nombre de données le plus grand après le modèle de *Marrero et Gani (2001)*, avec un nombre de groupes structurels acceptable.

3.5. Volume critique

Tableau 3.7 : Performances du modèle de la CIG pour le (V_c) pour les différentes classes.

Modèle proposé	$\frac{V_c}{(cm^3 / mol)} = \left(39,1941 + \sum_i N_i C_i \right)$			
Classes	Nombre de données	PEMA (%)	EMA (cm ³ /mol)	DS (cm ³ /mol)
Hydrocarbures	331	0,24	1,11	3,26
Contient l'oxygène	159	0,99	3,69	8,13
Contient l'azote	76	2,25	7,85	13,90
Contient le soufre	5	6,47E-14	1,93E-13	2,17E-13
Halogènes	120	1,55	4,17	9,33
Multifonctionnels	39	6,26	15,65	33,86
Ensemble de test	73	7,81	23,91	35,46
Total	730	2,09	7,24	13,09
$R^2 = 0,988$		$R^2 \text{ ajusté} = 0,984$		

Les résultats du tableau 3.7 montrent que la plus basse erreur est notée pour la classe des composés contenant le soufre, contrairement à la pression critique, et ça est dû aux groupes structuraux. Ces groupes représentent bien la relation entre la structure du composé et le volume critique, et le modèle aide à montrer cette relation. Concernant les autres classes, on peut dire que l'ordre du classement est presque le même, et la raison est la même, c'est le nombre de données.

Pour le volume critique, le R^2 et R^2 ajusté montrent que le modèle proposé pour la prédiction de cette propriété a offert une bonne précision.

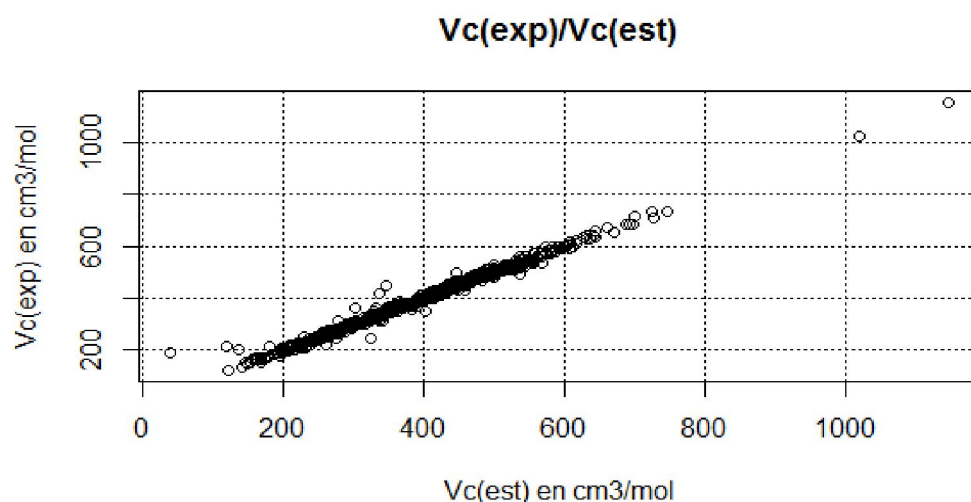


Figure 3.10 : Comparaison entre les valeurs expérimentales et estimées du (V_c).

La forme de la dispersion des points entre les valeurs expérimentales et estimées du volume critique se ressemble à une ligne, ce qui nous conduit à conclure que la précision obtenue par ce modèle de cette propriété est bonne.

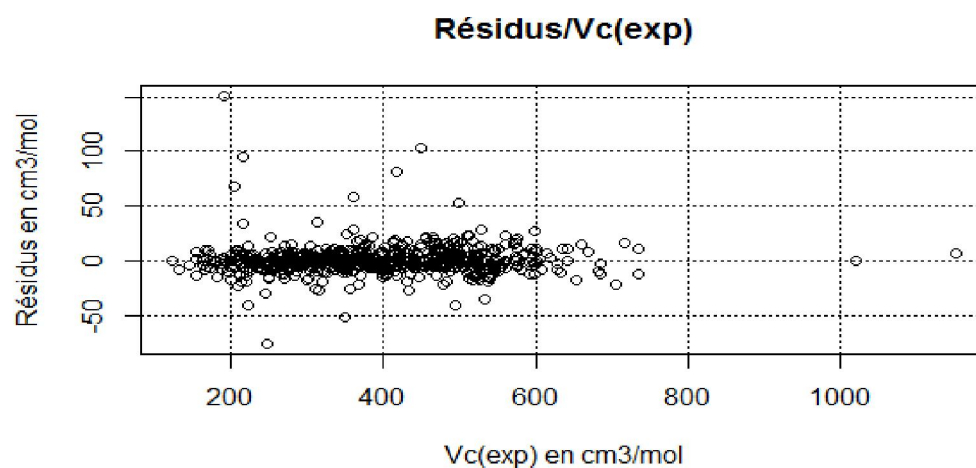


Figure 3.11 : Dispersion des résidus du (V_c) en fonction des (V_c) expérimentales.

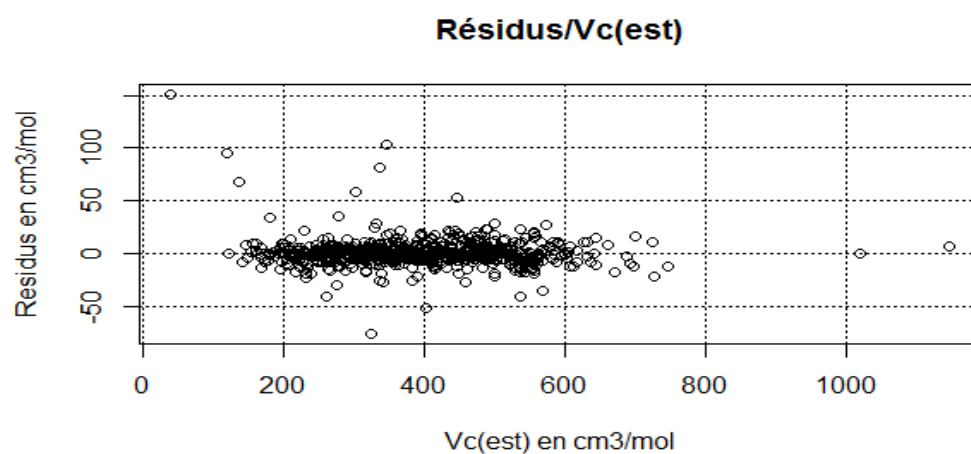


Figure 3.12 : Dispersion des résidus du (V_c) en fonction des (V_c) estimées.

Lorsqu'on compare les deux dispersions des points pour la propriété, on note qu'il existe quelques points qui sont loin de l'axe des abscisses, mais ces points n'ont pas une grande influence sur la précision du modèle, et concernant la forme de la dispersion, on constate que sont presque pareilles, alors, ce modèle est le meilleur pour représenter cette propriété.

Tableau 3.8 : Comparaison entre les performances du modèle de la CIG et des modèles précédents pour le (V_c).

Méthodes	Nombre de données	Nombre de groupes	PEMA (%)	EMA (cm^3/mol)	DS (cm^3/mol)
<i>Constantinou et Gani (1994)</i>	251	66	2,30	8	12
		40	1,79	6	10
<i>Marrero et Pardillo (1999)</i>	289	139	1,45	4,56	6,68
<i>Wen et Quiang (2001)</i>	275	140	1,41	4,83	ND
<i>Marrero et Gani (2001)</i>	762	124 1 ^{er} ordre	2,20	9,12	13,36
		79 2 ^{eme} ordre	1,90	7,85	11,57
		32 3 ^{eme} ordre	1,80	7,33	10,74
<i>Nannoolal, et al., (2006)</i>	348	218	1,79	6,40	ND
<i>Ce travail</i>	730	191	2,08	9,5	19,40

Les modèles de *Nannoolal, et al., (2004)*, *Constantinou et Gani (1994)*, ont offert une basse erreur, mais les deux modèles ont un petit nombre de données et un grand nombre de groupes structuraux, ce qui minimise la précision avec des nouveaux composés. Cependant, notre modèle a le plus grand nombre de données après le modèle de *Marrero et Gani (2001)*, mais ce dernier a le plus grand nombre de groupes structuraux. Concernant le modèle de *Wen et Quiang (2001)*, dans le développement de leur modèle, ils ont utilisé un petit nombre de données, et ça a une influence sur les performances du modèle avec les nouveaux composés.

3.6. Température de fusion

Tableau 3.9 : Performances du modèle de la CIG pour la (T_f) pour des différentes classes.

Modèle proposé	$\frac{T_f}{(K)} = 10,944 * \left(3,6155 + \sum_i N_i C_i \right)^{1,522}$			
Classes	Nombre de données	PEMA (%)	EMA (K)	DS (K)
Hydrocarbures	419	8,23	14,83	18,53
Contient l'oxygène	299	6,44	14,79	18,77
Contient l'azote	123	6,12	14,55	18,74
Contient le soufre	60	6,93	11,82	15,90
Halogènes	192	8,27	16,76	20,30
Multifonctionnels	29	4,60	11,69	15,68
Ensemble de test	112	8,35	16,78	20,36
Total	1124	7,35	14,85	18,71
$R^2 = 0,901$		$R^2 \text{ ajusté} = 0,879$		

Les résultats de tableau 3.9 montrent que la plus basse erreur est celle des composés multifonctionnels, suivis par les composés contenant l'azote, les composés contenant l'oxygène, les composés contenant le soufre, les hydrocarbures, et enfin les halogènes.

Malgré que les valeurs de R^2 et R^2 ajusté sont inférieures par rapport aux propriétés mentionnées au-dessus, mais ils restent toujours dans la gamme acceptable en comparaison avec les autres modèles.

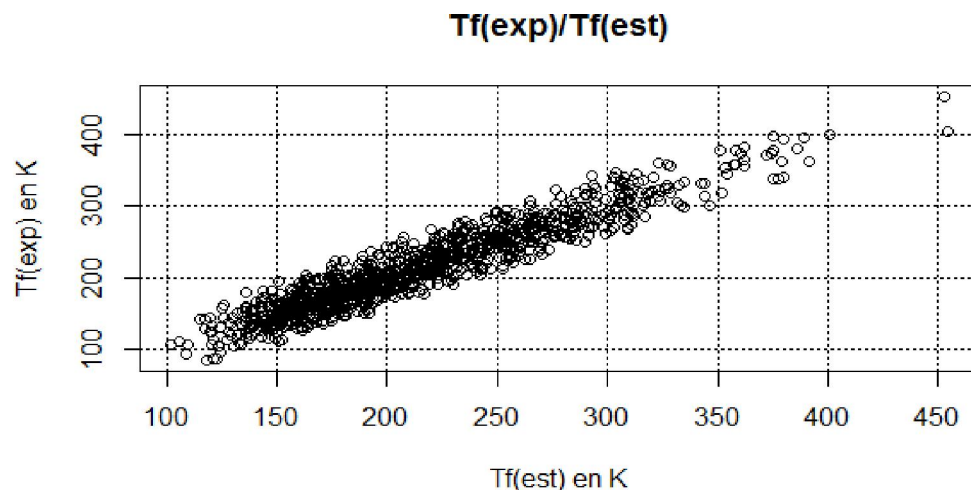


Figure 3.13 : Comparaison entre les valeurs expérimentales et estimées de (T_f).

La dispersion des valeurs expérimentales en fonction des valeurs estimées a l'aspect d'une ligne. Mais cette fois, elle est un peu plus large, et pour cette propriété (température de fusion) c'est le meilleur résultat.

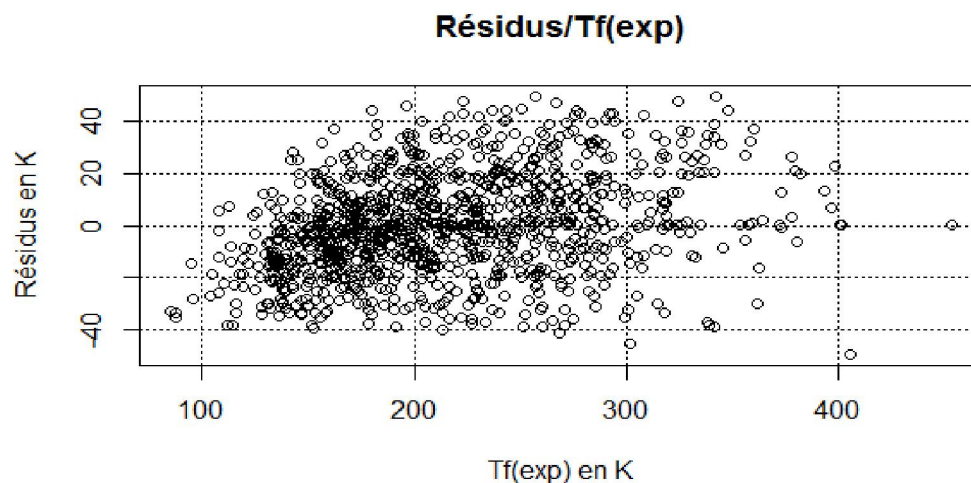


Figure 3.14 : Dispersion des résidus de la (T_f) en fonction des (T_f) expérimentales.

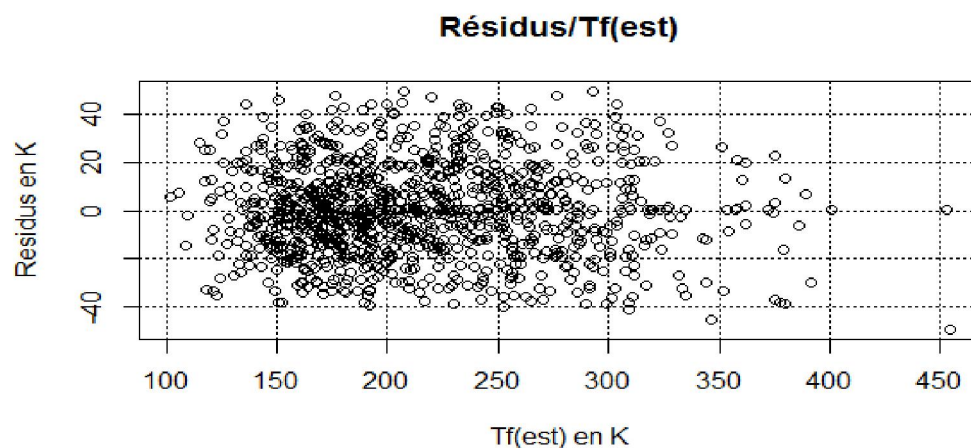


Figure 3.15 : Dispersion des résidus de la (T_f) en fonction des (T_f) estimées.

La forme de la dispersion des points des deux figures se ressemble, mais, le nombre des points qui sont aux environ de l'axe des abscisses est faible, mais pour cette propriété c'est l'optimum.

Tableau 3.10 : Comparaison entre les performances du modèle de la CIG et des modèles précédents pour la (T_f).

Méthodes	Nombre de données	Nombre de groupes	PEMA (%)	EMA (K)	DS (K)
<i>Constantinou et Gani (1994)</i>	312	66 1 ^{er} ordre	8,90	17,39	22,51
		41 2 ^{eme} ordre	7,23	14,03	18,28
<i>Marrero et Gani (2001)</i>	1547	166 1 ^{er} ordre	9,30	24,90	33,87
		117 2 ^{eme} ordre	7,90	21,41	29,52
		62 3 ^{eme} ordre	7,60	20,22	27,67
<i>Wang, et al., (2008) CPG</i>	730	93	6,67	14,46	ND
<i>Ce travail</i>	1124	203	7,35	14,85	18,71

La première remarque qu'on peut noter d'après le tableau 3.10, est que les modèles de la température de fusion sont rares par rapport aux autres propriétés, ce qui montre que le développement d'un modèle pour prédire cette propriété est très difficile, ainsi que la gamme d'erreur de cette propriété est très grand par rapport aux autres propriétés.

Le modèle qui a présenté la moindre erreur est celui de *Constantinou et Gani (1994)*, mais le nombre de données qui a été employé pour le développement de ce modèle est le plus petit, ce qui diminue la précision du modèle si on veut prédire cette propriété avec des nouveaux composés.

Le nombre le plus grand qui a été employé dans le développement d'un modèle est celui utilisé dans le modèle de *Marrero et Gani (2001)*, et cette fois aussi, ce modèle a le plus grand nombre de groupes structurels.

Malgré que ce modèle *Wang, et al., (2008)* a la plus basse erreur et nombre de groupes structurels, mais, son inconvénient est le nombre de données participant dans le développement du modèle, qui est aussi le plus bas.

Entre tous ces avantages et inconvénients notre modèle est au milieu, il a un nombre de données plus grand que certains modèles, et une erreur plus basse que d'autre modèle, donc, il est le meilleur choix pour obtenir les valeurs estimées de la température de fusion.

3.7. Enthalpie de fusion

Tableau 3.11 : Performances du modèle de la CIG pour la (ΔH_{fus}) pour les différentes classes.

Modèle proposé	$\frac{\Delta H_{fus}}{(kJ / mol)} = T_f * \left(0,14327 + \sum_i N_i C_i \right)^2$			
Classes	Nombre de données	PEMA (%)	EMA (kJ/mol)	DS (kJ/mol)
Hydrocarbures	264	17,39	1,25	1,64
Contient l'oxygène	95	10,25	1,07	1,56
Contient l'azote	37	10,11	1,15	1,6
Contient le soufre	33	10,7	0,84	1,3
Halogènes	62	17,76	1,01	1,43
Multifonctionnels	9	11,06	1,31	2,03
Ensemble de test	50	16,52	1,55	1,98
Total	501	14,95	1,15	1,58
$R^2 = 0,90$		$R^2 \text{ ajusté} = 0,87$		

Pour cette propriété, la plus basse erreur a été enregistrée avec les composés contenant l'azote, suivis par les composés contenant l'oxygène, ensuite, les composés contenant le soufre, les composés multifonctionnels, les hydrocarbures, et enfin les halogènes.

Les composés qui ont été choisis pour tester la précision du modèle ont présenté des bons résultats par rapport aux classes des hydrocarbures et les halogènes.

On note également pour cette propriété que les valeurs de R^2 et R^2 ajusté sont un peu inférieurs par rapport aux valeurs des autres propriétés, cependant, pour cette propriété les valeurs sont dans la gamme acceptable.

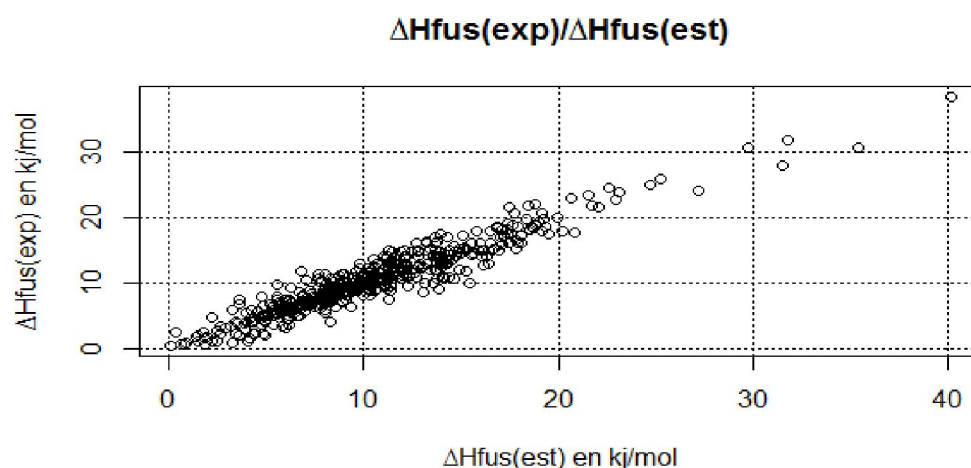


Figure 3.16 : Comparaison entre les valeurs expérimentales et estimées de (ΔH_{fus}).

L'aspect de la dispersion dans la figure ci-dessus se ressemble à une ligne, ce qui montre que la précision du modèle est bonne, et pour cette propriété cette forme est suffisante.

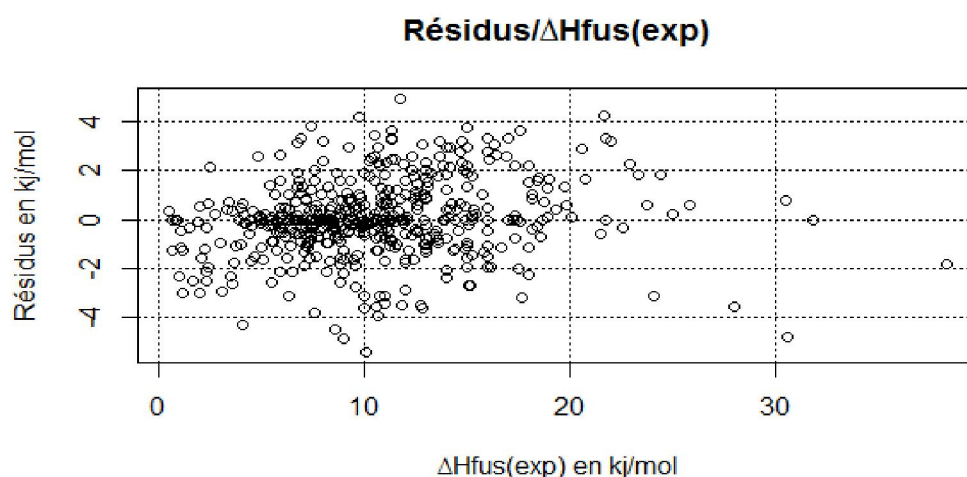


Figure 3.17 : Dispersion des résidus de la (ΔH_{fus}) en fonction des (ΔH_{fus}) expérimentales.

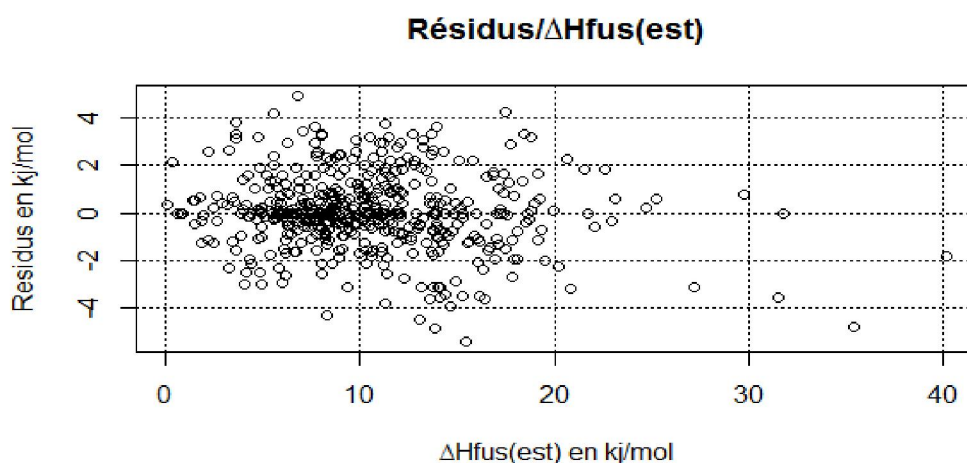


Figure 3.18 : Dispersion des résidus de la (ΔH_{fus}) en fonction des (ΔH_{fus}) estimées.

Malgré que la disposition des points par rapport à l'axe des abscisse est un peu grande, mais elle est acceptable pour cette propriété, et concernant la forme de la dispersion, les deux figures se ressemblent, ce qui montre que le modèle choisi est convenable.

Tableau 3.12 : Comparaison entre les performances du modèle de la CIG et le modèle de la CG de 3^{ème} ordre pour la (ΔH_{fus}).

Méthodes	Nombre de données	Nombre de groupes	PEMA (%)	EMA (kJ/mol)	DS (kJ/mol)
<i>Marrero et Gani (2001)</i>	711	131 1 ^{er} ordre	18,30	2,58	4,16
		68 2 ^{ème} ordre	16,40	2,32	3,88
		40 3 ^{ème} ordre	15,70	2,17	3,65
<i>Ce travail</i>	501	143	14,95	1,15	1,58

Un seul modèle a été proposé pour la prédiction de l'enthalpie de fusion. La raison est le manque des données expérimentales de cette propriété, ce qui explique la grande gamme d'erreur. Le modèle proposé dans ce travail a offert une précision satisfaisante ainsi qu'un nombre de groupes structuraux acceptable par rapport au modèle de *Marrero et Gani (2001)* malgré que le nombre de données soient un peu moins.

3.8. Enthalpie standard de formation (phase gazeuse)

Tableau 3.13 : Performances du modèle de la CIG pour la ($\Delta_f H^\circ_{\text{gaz}}$) pour les différentes classes.

Modèle proposé	$\frac{\Delta_f H^\circ_{\text{(gaz)}}}{(kJ / mol)} = \left(128,599 + \sum_i N_i C_i \right)$			
Classes	Nombre de données	PEMA (%)	EMA (kJ/mol)	DS (kJ/mol)
Hydrocarbures	143	5,33	1,33	1,89
Contient l'oxygène	101	0,67	2,05	2,64
Contient l'azote	38	1,76	1,01	1,56
Contient le soufre	14	4,17	1,99	2,84
Halogènes	42	5,28	2,15	2,84
Multifonctionnels	13	0,64	1,99	3,27
Ensemble de test	35	3,68	2,33	3,01
Total	351	3,38	1,65	2,32
$R^2 = 0,9998$		$R^2 \text{ ajusté} = 0,9997$		

Le modèle de l'enthalpie standard de formation (gaz) a donné la faible erreur avec les composés multifonctionnels suivis par les composés contenant l'oxygène, ensuite, les composés contenant l'azote, les composés contenant le soufre, les halogènes et enfin les hydrocarbures. Pour les composés qui ont été choisis pour tester le modèle, la précision a été acceptable surtout par rapport aux hydrocarbures et les halogènes.

Les valeurs de R^2 et R^2 ajusté pour ($\Delta_f H^\circ_{\text{gaz}}$) sont excellentes, ce qui prouve que le modèle représente bien la relation entre la propriété et les groupes structuraux, et ces valeurs signifient aussi que la précision est excellente.

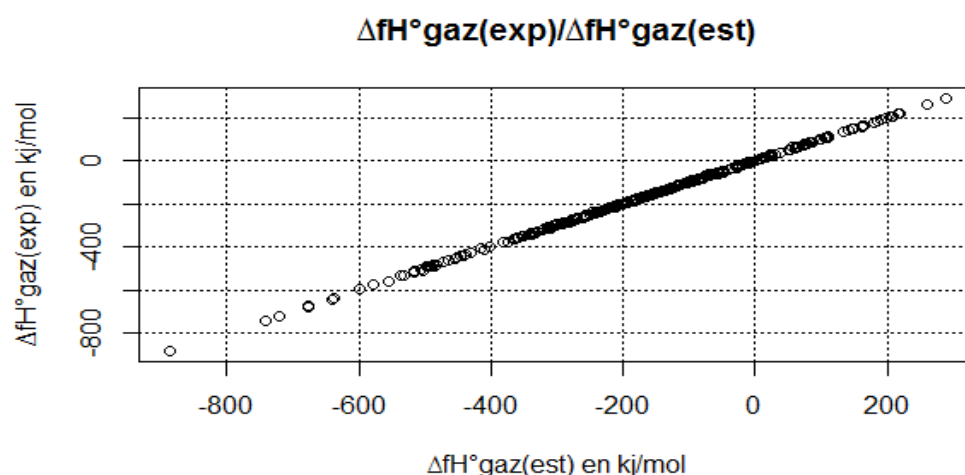


Figure 3.19 : Comparaison entre les valeurs expérimentales et estimées de ($\Delta_f H^\circ_{\text{gaz}}$).

La dispersion des points entre les valeurs expérimentales et les valeurs estimées est sous forme d'une ligne, ce que indique que la précision est excellente.

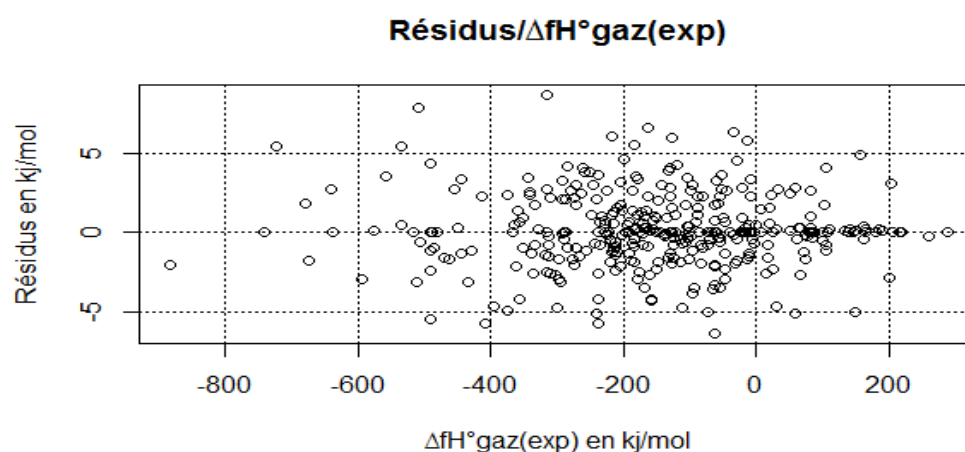


Figure 3.20 : Dispersion des résidus de la ($\Delta_f H^\circ_{\text{gaz}}$) en fonction des ($\Delta_f H^\circ_{\text{gaz}}$) expérimentales.

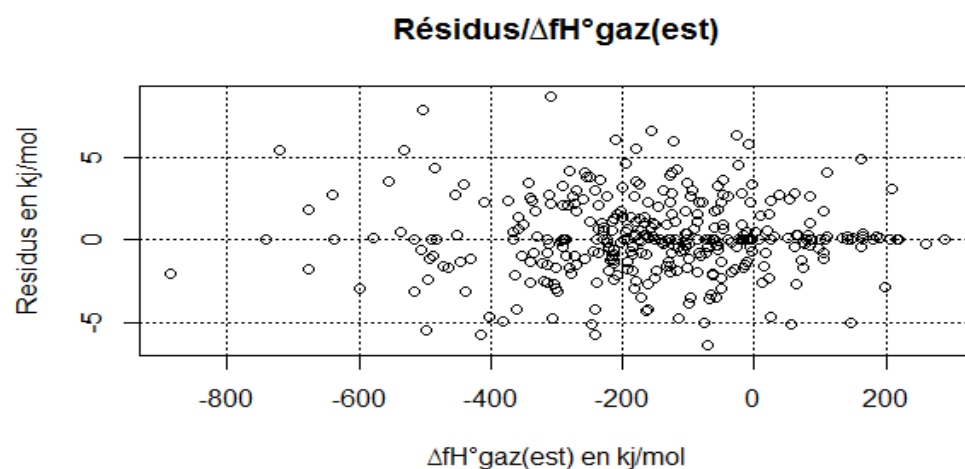


Figure 3.21 : Dispersion des résidus de la ($\Delta_f H^\circ_{\text{gaz}}$) en fonction des ($\Delta_f H^\circ_{\text{gaz}}$) estimées.

La ressemblance des deux dispersion est claire, ce qui signifie que le choix du modèle a été bien, et la majorité des points sont proches de l'axe des abscisses, une autre indication que la précision est bonne en utilisant ce modèle pour la prédiction de cette propriété.

Tableau 3.14 : Comparaison entre les performances du modèle de la CIG et le modèle de la CG de 3^{ème} ordre pour la ($\Delta_f H^\circ_{\text{gaz}}$).

Méthodes	Nombre de données	Nombre de groupes	PEMA (%)	EMA (kJ/mol)	DS (kJ/mol)
<i>Hukkerika, et al., 2013</i>	799	137 1 ^{er} ordre	ND	1,71	2,53
		90 2 ^{ème} ordre			
		36 3 ^{ème} ordre			
<i>Ce travail</i>	351	139	3,016	1,53	2,24

Le seul modèle qui permet de prédire ($\Delta_f H^\circ_{\text{gaz}}$) dans la littérature est celui de *Hukkerika, et al., 2013*, non seulement en raison du manque de données, mais aussi en à cause de la difficulté de trouver un modèle qui offre une précision acceptable. *Hukkerika, et al., 2013* ont utilisé la même méthode de *Marrero et Gani (2001)* pour l'obtention des valeurs estimées, le modèle proposé dans ce travail a présenté une précision meilleure que l'autre modèle, malgré que le nombre de données soient plus grand, sans oublier que le nombre de groupes structurels est considérable.

Conclusion générale

Conclusion générale

Le travail réalisé dans ce mémoire avait pour objectif de développer des modèles qui permettent la prédiction des propriétés physiques et thermodynamiques des composés organiques.

L'approche choisie *la contribution des interactions de groupes* a offert une bonne représentation à la relation structure-propriété, cette approche a basé principalement sur le type de la liaison dans le choix des groupes structurels, ainsi que les familles chimiques, ce qui montre que cette approche inclut un aspect théorique.

Une somme des données expérimentales de sept (7) propriétés qui ont été structurées sous forme d'une base de données, ces données ont été employées pour l'obtention des contributions de groupes.

Les sept modèles développés ont présenté des bons résultats, concernant la précision, la fiabilité et la bonne représentation de la relation entre les groupes structurels et la propriété.

La comparaison qui a été réalisée entre les performances des modèles proposés dans ce travail et les modèles déjà développés a prouvé que les résultats obtenus sont excellents, soit par rapport à la même méthode développée par la même approche (la contribution des interactions de groupes), ou par rapport aux autres méthodes, concernant, la distinction entre les isomères y compris les structures cis et trans ou ortho, meta et para, ainsi que la précision et la fiabilité.

Pour améliorer la méthode développée dans ce travail, on va élargir la base de données, en augmentant le nombre des valeurs expérimentales des propriétés déjà disponibles dans la base de données, et en ajoutant d'autres propriétés pour élargir la gamme d'application de la méthode.

Introduire des niveaux supérieurs de l'interaction entre les groupes, en utilisant l'interaction entre trois groupes organiques fonctionnels (ITGOF), ainsi que l'interaction à longue distance visant l'amélioration de la méthode.

Chercher d'autres modèles qui peuvent représenter la relation structure-propriété mieux que les modèles proposés.

Bibliographie

Atkins, P. William et Loretta, Jones. 1998. Chimie: molécule, matière, métamorphoses. s.l. : De Boeck, p. 318.

Benson, S. W, et al. 1969. Additivity rules for the estimation of thermochemical properties. Chem. Rev, Vol. 69 (3), p. 279-324.

Boethling, Robert. S et Mackay, Donald. 2000. Handbook of Property Estimation Methods for Chemicals Environmental and Health Sciences. New York : CRC Press.

Chattefuee, Amprit et Had, i Ali S. 2006. Regression Analysis by Example. 4 édition. New York, United States of America : John Wiley & Sons, Inc. p. 7.

Chemfinder 2013

<http://chemfinder.cambridgesoft.com/chembiofinder/Forms/Home/ContentArea/Home.aspx>.

Chemspider 2013. <http://www.chemspider.com/>.

Constantinou, Leonidas et Gani, Rafiqul. 1994. New Group Contribution Method for Estimating Properties of Pure Compounds, Vol. 40, 10, pp. 1697-1710.

Cordes, Wilfried et Rarey, Jürgen. 2002. A new method for the estimation of the normal boiling point of non-electrolyte organic compounds. Fluid Phase Equilibria, Vol. 201, pp. 409–433.

Cornillon, Pierre-André et Matzner-Løber, Eric. 2011. Régression avec R. Verlag, France : Springer, pp. 1-9, 29-32.

Dean, John A. 1999. Lange's handbook of chemistry. 5 ed. New York : McGraw-Hill.

Dodge, Yadolah. 2002. Premier pas en statistique. Neuchatel : Springer, 2002. pp. 337-378.

Eagleson, Mary. 1993. Concise Encyclopedia Chemistry. Berlin : Walter de Gruyter & CO, p. 277.

Fahim, Mohamed.A, Alsahhaf, Taher.A et Elkilani, Amal. 2010. Fundamentals of Petroleum Refining. 1er. s.l. : Elsevier B.V. pp. 11-31.

Jia, Qingzhu, et al. 2012. Prediction of the Flash Point Temperature of Organic Compounds with the Positional Distributive Contribution Method. J. Chem. Eng. Data. Vol. 57, p. 3357–3367.

Johnson, A. William. 2002. Invitation à la chimie organique. 1re Edition. s.l. : De Boeck, p. 27.

Bibliographie

Lide, David R. 2010. CRC Handbook of Chemistry and Physics. 90eme. Boca Raton, FL : CRC Press/Taylor and Francis.

Marrero, Jorge et Gani, Rafiqul. 2001. Group-contribution based estimation of pure component properties. Fluid Phase Equilibria, pp. 183–208.

Marrero-Morejon, Jorge et Pardillo-Fontdevila, Eladio. 1999. Estimation of Pure Compound Properties Using Group-Interaction Contributions. AIChE Journal, Vol. 45, 3, pp. 615-621.

Myers, Richard. 2003. The Basics of Chemistry. London : Greenwood press, p. 85.

Nannoolal, Yash. 2006. Development and Critical Evaluation of Group Contribution Methods for the Estimation of Critical Properties, Liquid Vapour Pressure and Liquid Viscosity of Organic Compounds. Durban : University of Kwazulu-Natal, Durban Campus, pp. 08, 12-13, Thèse de doctorat.

Nannoolal, Yash. 2004. Development of a Group Contribution Method for the Prediction of Normal Boiling Points of Non-electrolyte Organic Compounds. Durban : Université de Kwazulu-Natal, Mémoire de Master.

Nannoolal, Yash, et al. 2004. Estimation of pure component properties Part 1. Estimation of the normal boiling point of non-electrolyte organic compounds via group contributions and group interactions. Fluid Phase Equilibria, Vol. 226, pp. 45–63.

Nannoolal, Yash, Rarey, Jürgen et Ramjugernath, Deresh. 2007. Estimation of pure component properties Part 2. Estimation of critical property data by group contribution. Fluid Phase Equilibria, Vol. 252, pp. 1–27.

NIST 2013. <http://webbook.nist.gov/chemistry/>.

Poling, Bruce. E, Prausnitz, John M. et O'Connell, John P. 2001. The Properties of Gases and Liquids. 5e édition. New York : McGRAW-HILL, pp. 3.1-3.3, A1.

Reinhard, martin et Drefahl, axel. 1999. Handbook for estimating physicochemical properties of organic compounds. California, USA et Freiberg, Sachsen, Germany. : John Wiley and Sons, Inc. pp. 94-105.

Stefanis, Emmanuel, Constantinou, Leonidas et Panayiotou, Costas. 2004. A Group-Contribution Method for Predicting Pure Component Properties of Biochemical and Safety Interest. Ind. Eng. Chem. Res, Vol. 43, 19, pp. 6253-6261.

Wang, Qiang, et al. 2009. Position Group Contribution Method for Predicting the Normal Boiling Point of Organic Compounds. Chinese Journal of Chemical Engineering, Vol. 17, 2, pp. 254-258.

Bibliographie

Wang, Qiang, et al. 2008. Position Group Contribution Method for the Prediction of Critical Temperatures of Organic Compounds. *Journal of Chemical & Engineering Data*, Vol. 53, 5, pp. 1103–1109.

Wen, Xu et Qiang, Yang. 2001. A New Group Contribution Method for Estimating Critical Properties of Organic Compounds. *Ind. Eng. Chem. Res.*, 40, pp. 6245-6250.

Yaws, Carl.L. 2008. *Thermophysical Properties of Chemicals and Hydrocarbons*. Texas : William Andrew, pp. 01-809.

Annexe 01 : Contributions des interactions de groupes

		Propriétés physiques et thermodynamiques						
		C _{Teb}	C _{Tc}	C _{Pc}	C _{Vc}	C _{Tf}	C _{ΔHfus}	C _{ΔfH°(gaz)}
constants	constants	1,8214	6,8356E+00	2,8529	39,1941	3,6155	0,14327	128,599
	<i>cis</i>	0,0696	2,2037E-05	0,0332	-0,6378	-0,2031	0,00537	-5,476
	<i>trans</i>	0,0567	1,1092E-04	0,0675	2,1514	0,013	0,0135	-11,3111
	<i>ortho</i>	-0,0478	-2,8190E-04	0,0153	3,8892	0,0184	0,00661	0,3203
	<i>meta</i>	0,0136	-2,3371E-04	0,0297	-1,335	-0,3758	0,00358	-0,0601
corrections	<i>para</i>	0,1682	-2,0236E-04	0,0126	1,5608	0,4988	0,01306	1,4066
	>CH ₂	0,6128	1,4948E-03	1,1231	83,0554	0,5794	0,0245	-116,0138
	>CH-	0,5622	1,2022E-03	1,0277	74,4742	0,4161	0,01622	-87,2977
	>C<	0,519	1,0371E-03	0,9483	68,5264	0,5985	0,00241	-73,8847
	=CH-	0,5629	1,3652E-03	1,0108	84,8804	0,8216	-0,0041	-0,5443
CH ₃ -	=C<	0,6466	1,1742E-03	1,2884	103,8781	0,2978	0,00131	-12,35
	≡C-	0,6195	5,7176E-04	0,1508	33,3619	0,0084	0,01426	-54,7882
	>CH- ®	0,4843	1,1671E-03	0,9998	78,2538	0,2902	-0,03169	-78,52
	>C< ®	0,2165	1,6672E-03	1,5632	-27,4913	-2,363	-0,1125	-70,3454
	=C< ®	0,8978	1,3944E-04	-1,3231	19,3728	0,4869	0,05235	262,401
	=C< @	0,4324	1,2770E-03	0,9256	64,1346	0,7882	0,01371	-62,3174
	-O-	0,8097	1,4067E-03	0,8975	60,8809	1,1252	0,02486	-161,7534
	>CO	1,3855	1,5472E-03	0,8939	80,4082	1,1979	0,02722	-169,824
	-COO- (o)	2,4806	3,1483E-03	1,254	128,4141	2,4829	0,05914	-400,8798
	-COO- (c)	0,371	1,1503E-03	0,9831	58,5751	0,0378	0,00875	-138,7385
	>NH	0,9549	1,4775E-03	0,7432	70,3593	1,4857	0,02011	-73,6995
	>N-	0,6941	5,6873E-04	0,8164	69,0626	0,639	0,0205	-50,733
	>N-®	2,8119	1,7090E-03	3,7851	321,0519	2,9358	0,09733	ND
	-S-	1,293	1,0164E-03	0,7807	80,903	1,1707	0,03756	ND
-CH ₂ - via liaison simple (sans cycle)	-CH ₂ -	0,5957	8,8430E-04	0,7331	54,8872	0,2798	0,021	-20,706
	>CH-	0,5133	6,7790E-04	0,5662	38,927	-0,0192	0,00805	9,4019
	>C<	0,5157	2,3420E-04	0,3112	26,0082	-0,0294	0,00315	25,7868
	=CH-	0,441	9,2601E-04	0,7129	60,8256	0,2934	-0,02914	94,5225
	=C<	0,5294	4,5024E-04	0,7563	73,3681	-0,3418	-0,01514	84,9016
	≡C-	0,5647	-6,5216E-06	-0,201	4,7608	-1,2317	0,00124	41,5567
	>CH- ®	0,4877	4,7453E-04	0,7047	44,6214	-0,4014	0,00718	15,4302
	>C< ®	0,1325	1,6158E-03	1,0994	16,5713	-3,1363	ND	ND
	=C< ®	0,7521	-3,6550E-03	0,2046	-97,6257	-0,2361	ND	362,4148
	=C< @	0,2922	7,9225E-04	0,6264	36,8134	-0,2235	0,00126	33,1177
	-F	0,7711	1,1964E-03	0,5851	41,5623	1,1683	0,01797	-277,7792
	-Cl	1,5617	1,0230E-03	0,5748	72,1772	1,558	0,01781	-125,3572
	-Br	1,9308	5,3987E-04	-0,1179	85,2539	1,4865	0,01882	-76,5715
	-I	2,8055	7,4397E-04	0,7013	114,7474	1,7649	0,03245	-17,2942

	-OH	2,3847	2,8813E-03	-0,0071	50,21	1,8456	0,00577	-248,159
	-O-	0,5936	6,0257E-04	0,3808	35,5863	0,0921	0,0161	-72,4227
	>CO	1,1741	1,2218E-03	0,5927	59,8164	1,0423	0,0199	-77,5822
	-CHO	2,0578	1,8470E-03	-0,0153	81,412	1,9762	0,04758	-198,8483
	-COOH	4,3534	4,7888E-03	0,6977	109,8257	2,9685	0,00699	-465,5064
	-COO- (o)	2,2011	2,8277E-03	0,9321	100,9101	1,6397	0,05586	-318,7311
	-COO- (c)	0,2497	3,4824E-04	0,6927	29,6159	-0,473	-0,01322	-42,9678
	-NH2	1,7601	2,0421E-03	0,2569	75,666	2,1162	0,05978	-62,3816
	>NH	1,0355	1,4063E-03	0,4936	48,5072	0,714	ND	15,2203
	>N-	0,5587	9,1122E-04	0,6334	35,5667	0,119	ND	41,073
	-N=	2,013	1,3896E-04	-0,0941	77,7799	ND	0,02625	ND
	-SH	2,0559	8,5054E-04	0,2377	84,7506	1,3825	0,03492	-59,4408
	-S-	1,199	6,2236E-04	0,163	56,3476	0,4038	0,02314	-70,8612
>CH- via liaison simple (sans cycle)	>CH-	0,5343	8,5204E-05	0,1689	15,1677	0,1245	-0,00392	45,8397
	>C<	0,6213	-1,6857E-04	-0,1829	-0,6053	-0,2247	0,00488	66,4565
	=CH-	0,3647	7,5227E-04	0,5098	45,8259	0,0461	-0,0527	123,2588
	=C<	0,5025	3,5161E-05	0,5211	59,1644	-0,4722	-0,03298	116,4554
	≡C-	0,1588	4,3930E-05	-0,5018	-7,0408	-1,3576	ND	ND
	>CH- ®	0,4781	9,9048E-04	0,424	21,7978	-0,2429	ND	ND
	=C< ®	0,8086	ND	ND	ND	ND	ND	ND
	=CH-®	1,0589	1,8972E-03	1,1479	81,1662	2,4778	ND	ND
	>C=@	0,1048	3,7120E-04	0,349	27,3754	-0,2912	-0,00916	62,5298
	-F	0,3159	9,9479E-04	0,5517	33,5649	0,5427	0,0091	ND
	-Cl	1,0636	8,8411E-04	0,6007	67,1799	0,8645	0,02298	-81,4858
	-Br	1,7935	5,8302E-04	-0,1588	85,2728	1,4257	0,0195	-50,7346
	-I	2,4488	5,9042E-04	0,5178	103,4353	2,2782	ND	5,6964
	-OH	1,8019	3,2549E-03	-0,2693	44,8613	1,9673	-0,00184	-227,5706
	-O-	0,4077	8,1651E-04	0,2631	27,6767	0,3005	0,02255	-49,997
	>CO	0,9932	1,1651E-03	0,3008	41,4563	0,7497	0,01503	-44,2735
	-CHO	1,8358	2,8279E-03	0,9165	93,6026	2,1867	ND	ND
	-COOH	3,9194	5,0318E-03	1,0174	106,9438	3,342	-0,00437	ND
	-COO- (o)	1,9549	2,4508E-03	0,838	97,5618	1,5507	ND	-296,8651
	-COO- (c)	-0,0211	3,2317E-05	0,366	19,7577	-0,3723	ND	ND
	-NH2	1,5009	1,1290E-03	-0,1885	73,0302	2,0674	0,03947	-39,6382
	>NH	0,4795	ND	0,0943	ND	0,4071	ND	38,3959
	>N-	0,6013	ND	ND	ND	ND	ND	ND
	-SH	1,7907	8,6403E-04	0,367	71,4622	0,2348	0,0268	-32,7175
	-S-	1,0093	ND	ND	ND	0,3744	0,0192	ND
>C< via liaison simple (sans cycle)	>C<	0,7207	-4,7599E-04	-0,5646	9,0025	-0,6375	0,01622	91,3306
	=CH-	0,3184	4,5546E-04	0,3133	24,0233	-0,1549	-0,06248	ND
	=C<	0,5109	-4,8159E-04	0,2726	43,2687	-0,6435	-0,1534	140,4142
	≡C-	0,081	-1,5442E-04	-0,7591	-12,9297	-1,9001	ND	ND

	>C=@	0,1151	2,8304E-05	0,3429	28,9892	-0,8428	0,00344	85,6744
	>CH- ®	0,4368	ND	ND	ND	-0,362	ND	ND
	=C< ®	0,6484	ND	ND	ND	ND	ND	ND
	-F	-0,0203	9,2262E-04	0,7748	32,4218	0,3112	-0,00821	-240,8509
	-Cl	0,8022	6,3764E-04	0,5689	59,0924	0,8639	-0,00425	-57,032
	I	2,2571	ND	ND	ND	2,088	ND	20,9552
	-Br	1,6107	5,7048E-04	0,4659	80,5347	1,5877	-0,01134	-11,1747
	-OH	1,3392	3,3507E-03	-0,2117	37,2395	2,1292	-0,02336	-219,6736
	-O-	0,4589	4,3970E-04	0,1984	18,376	-0,2523	0,03955	-24,6945
	>CO	0,8936	7,3916E-04	0,2275	47,6377	0,2208	0,04863	-27,7208
	-CHO	1,5142	ND	ND	ND	0,8884	ND	-154,1031
	-COO- (o)	1,8322	1,8280E-03	0,6343	87,4391	2,5226	ND	-284,7063
	-COO- (c)	-0,1383	ND	ND	ND	-1,2519	ND	-2,7324
	-COOH	3,6946	5,0919E-03	1,7677	ND	3,2118	-0,03105	-398,2448
	-NH2	1,2204	1,3898E-03	ND	48,2267	1,0206	-0,0851	-27,9448
	>NH	0,3932	1,0080E-03	0,3555	66,7415	-0,3223	ND	ND
	-SH	1,5617	ND	ND	ND	ND	-0,11583	-17,7025
	-S-	1,1752	ND	ND	ND	-0,0405	0,02187	ND
CH2=	=CH-	0,7387	7,7569E-04	0,9253	64,2034	0,4168	0,07214	-106,5276
	=C<	0,657	1,0525E-03	0,25	-5,3827	1,2087	0,07114	-120,5089
	=C< ®	0,4343	8,4149E-03	1,7393	274,663	5,6803	0,17879	ND
	=C=	0,7561	5,0444E-04	0,7701	57,8809	0,7789	0,02366	30,9939
=CH- via liaison double (sans cycle)	=CH- d	0,8234	5,0215E-04	0,6746	32,6761	0,1317	0,08691	-129,0726
	=C< d	0,6777	6,5524E-04	-0,0713	-40,3737	0,5978	0,08595	-139,6868
	=C=	0,8374	5,8255E-04	0,6196	40,048	0,0897	0,04004	2,8382
	=CH- s	0,5045	5,4978E-04	0,5317	53,398	1,1431	-0,11664	194,689
	=C< s	0,5731	ND	0,6844	74,1072	-0,0206	-0,09118	ND
	≡C-	0,3138	2,2430E-05	-0,338	0,5151	-1,3212	ND	ND
	=CH- ®	0,1344	7,9915E-04	0,5462	44,7392	ND	ND	ND
	=C< ®s	0,4209	5,8616E-04	-1,6059	-13,9202	0,0039	ND	ND
	=C< @s	0,4854	1,1594E-03	0,5162	35,8741	0,202	-0,06499	141,4478
	-Br	1,5834	4,1012E-04	-0,4893	96,6026	1,3122	-0,01375	ND
	-F	0,3826	6,8389E-04	0,6525	44,8994	0,3963	ND	ND
	-Cl	1,046	5,0116E-04	0,3711	75,3004	1,3693	-0,01387	2,0878
	I	2,4124	ND	ND	ND	ND	ND	ND
	-O-	0,5368	1,5431E-03	0,4574	36,5737	0,541	ND	25,3711
	>CO	1,2983	4,1743E-03	0,3165	156,6026	5,8505	ND	44,8037
	-CHO	2,4148	2,7765E-03	0,8545	91,098	2,1026	ND	-99,5821
	-COOH	4,439	ND	ND	ND	4,1842	-0,02249	-338,5632
	-COO- (o)	2,0129	3,2765E-03	ND	ND	2,3311	-0,00757	-197,7329
	-COO- ©	0,268	ND	ND	ND	0,0684	-0,05079	73,856
	>NH	1,5286	-7,8044E-05	-0,7686	76,5169	3,1314	ND	ND

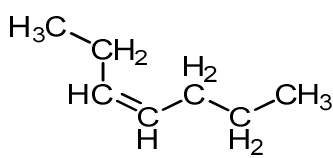
	=N-	3,5538	-4,8253E-04	-0,7117	29,0581	4,7427	ND	ND
=C< via liaison double (sans cycle)	=C<	0,5069	6,6433E-04	-0,5035	-85,8594	1,2879	0,05907	-139,499
	=C=	0,5128	5,2137E-04	-0,169	-13,8312	0,8247	0,05369	ND
=C< via liaison simple (sans cycle)	=C<	-0,2614	-4,3478E-04	0,8052	126,712	-0,6276	ND	ND
	-C≡	0,4031	-1,5146E-03	-0,2522	23,4123	-1,9401	ND	ND
	>C=@	0,2493	4,2505E-04	0,6888	56,8916	-0,0651	-0,04135	ND
	>CO	0,8615	ND	ND	ND	0,0828	ND	ND
	-O-	0,4228	ND	ND	ND	ND	ND	ND
	-CHO	2,0689	1,3568E-03	1,0689	117,3106	1,7972	ND	ND
	-COOH	4,7431	ND	ND	ND	2,508	ND	ND
	-COO- (o)	1,9864	ND	ND	ND	1,1407	ND	ND
	-COO- (c)	0,0808	ND	ND	ND	-0,3174	-0,02185	ND
	-Br	1,5538	-5,4956E-04	0,1683	117,0715	0,4575	ND	ND
	I	2,4827	ND	ND	ND	ND	ND	ND
	-F	0,1132	7,3769E-04	0,8911	56,198	-0,0394	0,007	ND
	-Cl	1,0015	6,3696E-04	0,8587	92,6355	0,4955	ND	ND
≡CH via liaison triple (sans cycle)	≡C-	0,9277	1,3614E-03	1,1657	95,3532	2,6476	0,03324	111,2784
≡C- via liaison triple (sans cycle)	≡C-	1,1497	1,2791E-03	1,7972	116,2699	3,2622	0,02416	126,6774
	>C=@	-0,3903	-1,1015E-03	-0,3525	-27,9805	-1,3466	-0,01651	102,7202
	≡N	3,1676	3,6077E-03	1,6815	95,4433	3,0663	0,01302	ND
-CH2- ® via liaison simple (avec cycle)	-CH2- ®	0,6222	5,4848E-04	0,4549	45,6267	0,454	-0,00819	-42,1309
	>CH- ®	0,5185	3,9433E-04	0,4085	37,0442	0,2853	0,0394	-15,2314
	>C< ®	0,7339	-1,6085E-04	-0,2248	89,8952	3,1951	0,0963	ND
	=CH- ®	0,8487	2,2949E-03	1,6989	131,2574	0,8418	0,04002	31,8046
	=C< ®	0,532	-3,2284E-03	0,7919	-61,2761	0,4485	ND	15,9582
	=C< @	0,3976	6,2006E-04	0,1931	26,3376	0,0807	0,02783	-5,8706
	-O- ®	0,678	5,5730E-04	0,1809	27,2783	0,9742	0,01455	ND
	>CO ®	1,3397	2,0886E-03	0,0149	55,6267	0,9324	0,08152	-359,2167
	>NH ®	0,8967	7,2222E-04	0,0771	43,9374	1,3429	0,06365	-3,2454
	>N- ®	-1,3156	ND	-2,4803	-197,1261	-1,5465	ND	ND
	-S- ®	1,2299	2,4575E-04	ND	ND	0,9565	0,04254	-14,9704
>CH- ® via liaison simple (avec cycle)	>CH- ®	0,5203	4,8797E-04	0,2304	23,1091	0,377	-0,00336	17,8555
	>C< ®	0,5803	-9,2891E-04	-0,4113	93,3203	2,8064	0,06872	ND
	=C<	0,9504	ND	ND	ND	0,9325	ND	ND
	=CH- ®	0,5934	2,1046E-03	1,6307	126,5756	1,1971	-0,04488	54,0787
	=C< ®	0,4828	-6,7296E-03	-1,1485	-46,9983	0,2178	ND	ND
	-O- ®	0,8565	ND	ND	ND	0,9194	ND	ND
	=C< @	0,3374	1,1804E-03	3,3737	17,2224	-0,0912	0,00032	69,3066

	>CO®	1,0637	ND	ND	ND	1,2021	ND	ND
	–OH	1,595	4,7092E-03	-0,2992	18,2271	2,273	ND	-251,2918
	>NH2	1,0517	2,0020E-03	0,0328	64,2106	1,4523	0,01143	-30,1281
	>NH ®	0,6549	1,3554E-03	0,3304	32,0164	0,0794	ND	ND
	–N=	1,4405	2,0976E-03	0,3595	54,3054	ND	ND	ND
	–SH	1,6282	1,2024E-03	ND	ND	0,512	0,03365	-22,8281
	–S– ®	1,0453	ND	ND	ND	0,6085	ND	ND
>C< ® via liaison simple (avec cycle)	>C< ®	0,3469	-1,5647E-03	-0,0626	41,4017	3,8879	ND	ND
	=CH– ®	0,8653	ND	ND	ND	ND	ND	ND
	=C< ®	0,6049	ND	ND	ND	ND	ND	ND
	>CO®	1,2738	ND	ND	ND	3,1743	ND	ND
	=C< @	0,807	ND	ND	ND	-12,7743	ND	ND
	–OH	0,9662	ND	ND	ND	ND	ND	ND
	–F	0,0387	ND	ND	ND	ND	ND	ND
=CH– ® via liaison double (avec cycle)	–S– ®	1,1015	ND	ND	ND	ND	ND	ND
	=CH– ®	0,0564	-3,0154E-03	-2,3188	-145,5737	-0,5657	-0,06828	-71,2
	=C< ®	-0,0429	3,3935E-03	0,6699	82,6613	-0,4823	0,02056	-358,301
=CH– ® via liaison simple (avec cycle)	=N ®	5,0097	ND	ND	ND	-1,1374	ND	-128,701
	=CH– ®	1,1373	3,5860E-03	2,8872	213,9015	1,6954	0,08585	122,001
	=C< ®	1,3697	-1,6408E-03	0,0171	ND	1,4163	ND	362,35
	=C< @	0,5709	2,6402E-03	1,9088	150,7504	1,9825	0,00449	57,9388
	–O– ®	-2,5269	ND	ND	ND	3,2714	0,06503	27,901
	>NH ®	1,6018	2,5117E-03	1,2523	134,0259	1,813	0,04271	ND
	>N– ®	ND	ND	ND	ND	ND	ND	ND
	=N– ®	3,1654	ND	ND	ND	-1,907	ND	ND
=C< via liaison double (avec cycle)	–S– ®	0,956	ND	ND	ND	1,0846	-0,00326	ND
	=C< ®	-0,2181	ND	ND	ND	-0,1818	ND	ND
=C< via liaison simple (avec cycle)	=C< ®	1,3303	ND	ND	ND	1,3085	ND	ND
	–O– ®	6,6296	ND	ND	ND	ND	ND	ND
	–F	ND	ND	ND	ND	ND	ND	ND
	–Cl	1,2243	ND	ND	ND	ND	ND	ND
	=C< @	0,1588	ND	ND	ND	ND	ND	ND
	–S– ®	0,8278	ND	ND	ND	0,9214	ND	ND
–O– via liaison simple	>CO	0,9536	1,1873E-03	0,2375	44,9046	1,2196	ND	-194,813
>CO®	>NH®	4,3501	9,0159E-04	-0,0434	34,3617	ND	-0,05599	ND
	>N–®	2,4037	ND	ND	ND	ND	ND	ND

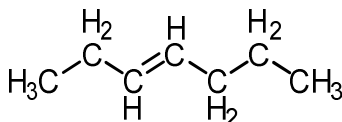
	-O-®	3,0608	ND	ND	ND	0,9598	ND	ND
>CO via liaison simple (sans cycle)	>CO	ND	4,2049E-03	ND	85,9133	0,0506	ND	ND
	F	0,4883	ND	ND	ND	ND	0,03102	ND
	-Cl	1,0804	8,1439E-04	0,6143	ND	1,0028	0,0277	-167,4653
	-NH2	5,6588	ND	ND	ND	ND	ND	-197,075
	>NH	4,2653	2,6038E-03	0,4121	71,4852	3,1357	ND	ND
	>N-	3,4907	2,9239E-03	0,1611	84,6758	2,1312	ND	-219,533
	=N-	0,4846	2,5100E-03	0,7754	21,9052	ND	ND	ND
	-S-	0,9116	8,9409E-04	0,5803	59,9948	ND	ND	ND
>NH via liaison simple	>NH	5,813	ND	0,1823	ND	2,0412	ND	ND
-S- via liaison simple	-S-	1,4131	9,3058E-04	ND	ND	0,5189	-0,00928	ND
-COO- ©	-COO-	2,2393	4,6088E-03	ND	116,3692	ND	ND	-276,8724
-N=	-N=	0,5811	3,2794E-04	0,8694	0,2924	-0,6993	ND	ND
	>NH	1,0591	1,4633E-03	0,0523	64,7195	1,4311	ND	ND
=CH- @	=CH- @ s	0,598	-4,5498E-04	0,0036	-7,014	-0,3479	0,02288	33,8375
	=CH- @ d	0,5927	1,3975E-03	0,6373	81,378	1,3224	-0,00709	-48,1961
	=C< @ s	0,7114	4,9190E-05	0,0863	7,7608	-0,1164	0,01416	32,0926
	=C< @ d	0,6737	8,6304E-04	0,3893	59,4993	0,7322	-0,00204	-18,9947
	=N- s	1,5919	3,6414E-04	-0,1894	-2,6681	0,6128	0,04309	62,9023
	=N- d	0,188	3,5065E-04	-0,0155	76,7255	1,2739	-0,01063	-22,3362
=C< @ s	=C< @ s	0,9123	2,9930E-04	0,1155	16,0303	-0,0788	0,00679	33,9044
	=C< @ d	0,8239	7,5361E-05	0,2522	33,5701	0,2277	-0,01499	20,9232
	-O-	0,7275	8,4309E-04	0,1289	29,464	0,3013	0,0072	-19,1264
	-OH	1,9678	1,8034E-03	-0,5067	-34,5173	2,0711	0,01921	-206,4913
	>CO	0,845	ND	ND	ND	ND	ND	ND
	-COO- c	-0,6959	-1,2129E-05	0,4548	74,5303	-1,3481	-0,03546	ND
	-F	-0,2869	9,3138E-04	0,512	24,8789	0,7494	0,00859	ND
	-Br	1,4584	5,7342E-04	0,3041	64,0929	1,4607	0,01946	ND
	-Cl	0,886	1,0659E-03	0,6523	56,1183	1,1877	0,01723	-58,4924
	-I	2,237	5,2374E-04	0,4732	95,8178	1,4619	ND	51,9202
	-SH	1,7148	ND	ND	ND	1,9012	ND	ND
	-S-	1,2477	2,3867E-04	ND	26,009	1,2886	0,00443	5,9043
	-NH2	1,9921	2,0118E-03	0,1254	48,2552	1,4971	0,00822	-25,4798
	>NH	1,2898	1,1188E-03	0,4212	41,9051	-0,0705	-0,00181	60,9697
	>N-	0,8219	1,2236E-04	-0,054	66,3864	0,7405	ND	88,9862
	=N- s	1,4007	7,9462E-04	0,3187	63,9126	0,7349	0,01082	58,2528

Annexe 02 : Exemples

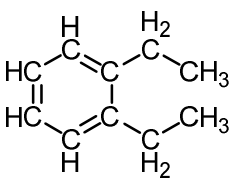
Exemple 01

Température d'ébullition T _{eb}			
<u>Nom</u> : cis-3-Haptène <u>Cas Number</u> : 7642-10-6 <u>Formule moléculaire</u> : C ₇ H ₁₄ T _{eb} (exp) = 368,9 K		<u>Structure moléculaire</u> : 	
Méthode des contributions des interactions de groupes			
Groupe i	Nombre d'apparition N _i	Contribution C _i	N _i * C _i
-CH3 et >CH2	2	0,6128	1,2256
>CH2 et >CH2	1	0,5957	0,5957
>CH2 et =CH-	2	0,441	0,882
-CH= et =CH-	1	0,8234	0,8234
cis	0,0696		0,0696
Σ _i N _i C _i			3,5963
c			1,8214
$\frac{T_{\acute{e}b}}{K} = -a + b \ln\left(c + \sum N_i C_i\right) \quad T_{\acute{e}b} = 265,476*(1,6897) - 84,565 \quad T_{\acute{e}b} = 364K$			
T _{eb} (exp)= 368,9 K, erreur = 364-368,9 = 4,89 K, 1,32 %			

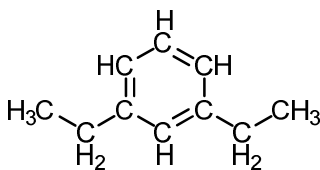
Exemple 02

Température normale d'ébullition T _{eb}			
<u>Nom</u> : trans-3-Heptène <u>Cas Number</u> : 14686-14-7 <u>Formule moléculaire</u> : C ₇ H ₁₄ T _{eb} (exp) = 368,82 K		<u>Structure moléculaire</u> : 	
Méthode des contributions des interactions de groupes			
Groupe i	Nombre d'apparition N _i	Contribution C _i	N _i * C _i
-CH3 et >CH2	2	0,6128	1,2256
>CH2 et >CH2	1	0,5957	0,5957
>CH2 et -CH=	2	0,441	0,882
-CH= et =CH-	1	0,8234	0,8234
trans	0,0567		0,0567
Σ _i N _i C _i			3,5834
c			1,8214
$\frac{T_{\acute{e}b}}{K} = -a + b \ln\left(c + \sum N_i C_i\right) \quad T_{\acute{e}b} = 265,476*(1,6872) - 84,565 \quad T_{\acute{e}b} = 363,37K$			
T _{eb} (exp)= 368,82 K, erreur = 363,37-368,82 = 5,45K, 1,48 %			

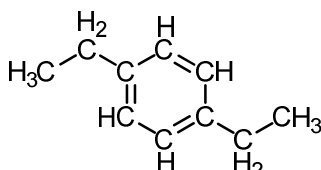
Exemple 03

Température normale d'ébullition T _{eb}			
<u>Nom</u> : o-diethylbenzene <u>Cas Number</u> : 135-01-3 <u>Formule moléculaire</u> : C ₁₀ H ₁₄ T _{eb} (exp) = 456,61 K		<u>Structure moléculaire</u> : 	
Méthode des contributions des interactions de groupes			
Groupe i	Nombre d'apparition N _i	Contribution C _i	N _i * C _i
-CH3 et >CH2	2	0,6128	1,2256
>CH2 et >C=@	2	0,2922	0,5844
=CH- @ =CH- @ s	2	0,5980	1,196
=CH- @ =CH-@ d	1	0,5927	0,5927
=CH- @ =C< @d	2	0,6737	1,3474
=C< @ =C< @s	1	0,9123	0,9123
ortho	-0,0478		-0,0478
Σ _i N _i C _i			5,8106
c			1,8214
$\frac{T_{éb}}{K} = -a + b \ln\left(c + \sum N_i C_i\right)$ T _{éb} = 265,476*(2,0323)– 84,565 T _{éb} = 454,97K			
T _{eb} (exp)= 456,61 K, erreur = 454,97-456,61 = 1,63 K, 0,358 %			

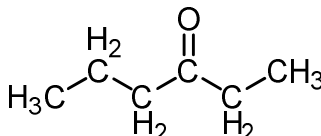
Exemple 04

Température normale d'ébullition T _{eb}			
<u>Nom</u> : m-diethylbenzene <u>Cas Number</u> : 141-93-5 <u>Formule moléculaire</u> : C ₁₀ H ₁₄ T _{eb} (exp) = 454,29 K		<u>Structure moléculaire</u> : 	
Méthode des contributions des interactions de groupes			
Groupe i	Nombre d'apparition N _i	Contribution C _i	N _i * C _i
-CH3 et >CH2	2	0,6128	1,2256
>CH2 et >C=@	2	0,2922	0,5844
=CH- @ =CH- @ s	1	0,5980	0,5980
=CH- @ =CH-@ d	1	0,5927	0,5927
=CH- @ =C< @ s	2	0,7114	1,4228
=CH- @ =C< @d	2	0,6737	1,3474
meta	0,0136		0,0136
Σ _i N _i C _i			5,7845
c			1,8214
$\frac{T_{éb}}{K} = -a + b \ln\left(c + \sum N_i C_i\right) \quad T_{éb} = 265,476*(2,0289) - 84,565 \quad T_{éb} = 454,06K$			
T _{eb} (exp)= 454,29 K, erreur = 454,06-454,29 = 0,23 K, 0,05 %			

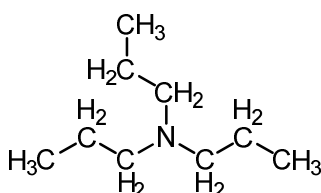
Exemple 05

Température normale d'ébullition T _{eb}			
<u>Nom</u> : p-diethylbenzene <u>Cas Number</u> : 105-05-5 <u>Formule moléculaire</u> : C ₁₀ H ₁₄ T _{eb} (exp) = 456,94 K		<u>Structure moléculaire</u> : 	
Méthode des contributions des interactions de groupes			
Groupe i	Nombre d'apparition N _i	Contribution C _i	N _i * C _i
-CH3 et >CH2	2	0,6128	1,2256
>CH2 et >C=@	2	0,2922	0,5844
=CH- @ =CH- @ s	1	0,5980	0,5980
=CH- @ =CH-@ d	1	0,5927	0,5927
=CH- @ =C< @ s	2	0,7114	1,4228
=CH- @ =C< @d	2	0,6737	1,3474
para	0,1682		0,1682
Σ _i N _i C _i			5,9391
c			1,8214
$\frac{T_{éb}}{K} = -a + b \ln\left(c + \sum N_i C_i\right)$ T _{éb} = 265,476*(2,049) – 84,565 T _{éb} = 459,407K			
T _{eb} (exp)= 456,94 K, erreur = 459,407-456,94 = 2,467 K, 0,539%			

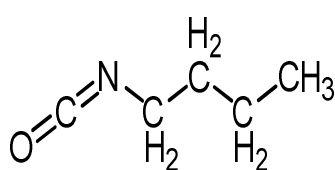
Exemple 06

Température critique T _c			
<u>Nom</u> : 3-Hexanone <u>Cas Number</u> : 589-38-8 <u>Formule moléculaire</u> : C ₆ H ₁₂ O T _c (exp) = 582,82 K T _{eb} (exp)= 396,65 K		<u>Structure moléculaire</u> : 	
Méthode des contributions des interactions de groupes			
Groupe i	Nombre d'apparition N _i	Contribution C _i	N _i * C _i
-CH3 et >CH2	2	1,4948E-03	2,9896E-03
>CH2 et >CH2	1	0,8843E-03	0,8843E-03
>CH2 et >CO	2	1,2218E-03	2,4436E-03
Σ _i N _i C _i			6,3175E-03
c			6,8356
$T_c = \frac{T_{eb}}{\left(c + \sum N_i C_i\right)^2 - 46,1301} \quad T_c(est) = \frac{396,65}{46,811 - 46,1301} \quad T_c(est) = 581,82K$			
T _c (exp)= 582,82 K, erreur = 581,82-582,82 = 0,995 K, 0,17 %			

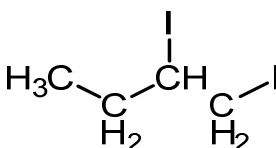
Exemple 07

Pression critique P _c			
<u>Nom</u> : Tripropylamine <u>Cas Number</u> : 102-69-2 <u>Formule moléculaire</u> : C ₉ H ₂₁ N P _c (exp) = 22,0084 bar		<u>Structure moléculaire</u> : 	
Méthode des contributions des interactions de groupes			
Groupe i	Nombre d'apparition N _i	Contribution C _i	N _i * C _i
-CH3 et >CH2	3	1,1231	3,3693
>CH2 et >CH2	3	0,7331	2,1993
>CH2 et >N-	3	0,6334	1,9002
Σ _i N _i C _i			7,4688
c			2,8529
$\frac{P_c}{bar} = \frac{239,15}{\left(c + \sum N_i C_i\right)} - 1,855 \qquad \frac{P_c}{bar} = \frac{239,15}{10,3217} - 1,855 \qquad P_c = 21,31bar$			
P _c (exp)= 22,0084 bar, erreur = 21,31-22,0084 =0,6984, 3,17 %			

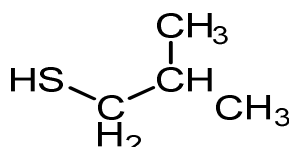
Exemple 08

Volume critique V _c			
<u>Nom</u> : butyl isocyanate <u>Cas Number</u> : 111-36-4 <u>Formule moléculaire</u> : C ₅ H ₉ NO V _c (exp) = 360 cm ³ /mol		<u>Structure moléculaire</u> : 	
Méthode des contributions des interactions de groupes			
Groupe i	Nombre d'apparition N _i	Contribution C _i	N _i * C _i
-CH3 et >CH2	1	83,0554	83,0554
>CH2 et >CH2	2	54,8872	109,7744
>CH2 et -N=	1	77,7799	77,7799
>CO et -N=	1	21,9052	21,9052
Σ _i N _i C _i			292,5149
c			39,1941
$\frac{V_c}{cm^3/mol} = \left(c + \sum N_i C_i \right) \quad V_c = 331,709 cm^3/mol$			
V _c (exp)= 360 cm ³ /mol, erreur = 331,709-360 = 28,29 cm ³ /mol, 7,85 %			

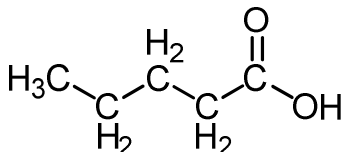
Exemple 09

Température de fusion T _f			
<u>Nom</u> : 1,2-Diiodobutane <u>Cas Number</u> : 53161-72-1 <u>Formule moléculaire</u> : C ₄ H ₈ I ₂ T _f (exp) = 279,06 K		<u>Structure moléculaire</u> : 	
Méthode des contributions des interactions de groupes			
Groupe i	Nombre d'apparition N _i	Contribution C _i	N _i * C _i
-CH3 et >CH2	1	0,5794	0,5794
>CH2 et >CH-	2	-0,0192	-0,0384
>CH2 et -I	1	1,7649	1,7649
>CH- et -I	1	2,2782	2,2782
Σ _i N _i C _i			4,5841
c			3,6155
$\frac{T_f}{K} = 10,944 * \left(c + \sum N_i C_i\right)^{1,522}$ $\frac{T_f}{K} = 10,944 * (8,1996)^{1,522}$ T _f = 269,13K			
T _f (exp)= 269,13 K, erreur = 269,13-279,06 = 9,926 K, 3,55 %			

Exemple 10

Enthalpie de fusion ΔH_{fus}			
<u>Nom</u> : isobutyl mercaptan <u>Cas Number</u> : 513-44-0 <u>Formule moléculaire</u> : C ₄ H ₁₀ S ΔH_{fus} (exp) = 4,98 kJ/mol à T _f = 128,31K		<u>Structure moléculaire</u> : 	
Méthode des contributions des interactions de groupes			
Groupe i	Nombre d'apparition N _i	Contribution C _i	N _i * C _i
-CH3 et >CH-	2	0,01622	0,03244
>CH2 et >CH-	1	0,00805	0,00805
>CH2 et -SH	1	0,03492	0,03492
$\sum_i N_i C_i$			0,07541
c			0,14327
$\frac{\Delta H_{fus}}{kJ / mol} = T_f * (c + \sum N_i C_i)^2 \quad \frac{\Delta H_{fus}}{kJ / mol} = 128,31*(0,21868)^2 \quad \Delta H_{fus} = 6,135kJ / mol$			
ΔH_{fus} (exp) = 4,98 kJ/mol, erreur = 6,135-4,98 = 1,15 kJ/mol, 23,21 %			

Exemple 11

Enthalpie de formation phase gazeuse $\Delta H^\circ_{\text{for}}(\text{gaz})$			
<u>Nom</u> : Pentanoic acid <u>Cas Number</u> : 109-52-4 <u>Formule moléculaire</u> : C ₅ H ₁₀ O ₂ $\Delta H^\circ_{\text{for}}(\text{gaz}) (\text{exp}) = -491,9 \text{ kJ/mol}$		<u>Structure moléculaire</u> : 	
Méthode des contributions des interactions de groupes			
Groupe i	Nombre d'apparition N _i	Contribution C _i	N _i * C _i
-CH3 et >CH2	1	-116,0138	-116,0138
>CH2 et >CH2	2	-20,7060	-41,412
>CH2 et -COOH	1	-465,5064	-465,5064
$\sum_i N_i C_i$			-622,9322
c			128,5990
$\frac{\Delta_f H^\circ_{(\text{gaz})}}{\text{kJ} / \text{mol}} = \left(c + \sum N_i C_i \right) \quad \Delta_f H^\circ_{(\text{gaz})} = -494,33 \text{ kJ} / \text{mol}$			
$\Delta_f H^\circ_{(\text{gaz})} (\text{exp}) = -491,9 \text{ kJ/mol}$, erreur = $ -494,33+491,9 = 2,43 \text{ K}$, 0,49 %			

Annexe 03 : Programme sur R**# Température d'ébullition T_{eb}** **# 1. Régression non linéaire**

```
Tb1<-fitModel(log(cis +trans+ ortho+ meta+para+Inter1 +...+ Inter240+Inter241+Inter242+
Inter243+Inter244+Inter245)~(Tb+a)/b,data=Tball)
```

```
coef(Tb1)
```

```
model(Tb1)
```

2. Régression linéaire

```
trainTeb <- sp[1:2082,]
```

```
testTeb <- sp[2083:2312,]
```

```
Tb<-lm(mod~(cis +trans+ ortho+ meta+ para+ Inter1 +...+Inter242+Inter243+Inter244+
Inter245),data=Tb)
```

```
summary(Tb)$r.squared
```

```
summary(Tb)$adj.r.squared
```

```
summary(Tb)$residuals
```

```
coefficients(Tb)
```

```
fitted.values(Tb)
```

```
predict(f, newdata=test)
```

```
write.csv(Tb$coef, "C:\\Users\\khaled\\Desktop\\R data\\Les contributions\\coefTb.csv")
```

```
write.csv(Tb$fit, "C:\\Users\\khaled\\Desktop\\R data\\Les contributions\\fitTb.csv")
```

```
plot(PlotTb$Tb~PlotTb$Pred,xlab="Teb(est) en K",ylab="Teb(exp) en K",type="p",main=
"Teb(exp)/Teb(est)")
```

```
panel.first = grid(col="black")
```

```
plot(PlotTb$Res~PlotTb$Tb,xlab="Teb(exp) en K",ylab="Résidus en K",type="p",main=
"Résidus/Teb(exp)")
```

```
panel.first = grid(col="black")
```

```
plot(PlotTb$Res~PlotTb$Pred,xlab="Teb(est) en K",ylab="Residus en K",type="p",main=
"Résidus/Teb(est)")
```

```
panel.first = grid(col="black")
```

Température de fusion T_f **# 1. Régression non linéaire**

```
Tf1<-fitModel((cis +trans+ ortho+ meta+para+Inter1+...+Inter242+Inter243+Inter244+
Inter245)~(Tf/a)^b,data=Tfall)
coef(Tf1)
model(Tf1)
```

#2. Régression linéaire

```
trainTf <- Tf[1:1012,]
testTf <- Tf[1013:1124,]
Tf<-lm(mod~(cis +trans+ ortho+ meta+ para+ Inter1+...+ Inter242+Inter243+
Inter244+Inter245),data=trainTf)
summary(Tf)$r.squared
summary(Tf)$adj.r.squared
summary(Tf)$residuals
coefficients(Tf)
fitted.values(Tf)
write.csv(Tf$coef, "C:\\Users\\khaled\\Desktop\\R data\\Les contributions\\coefTf.csv")
write.csv(Tf$fit, "C:\\Users\\khaled\\Desktop\\R data\\Les contributions\\fitTb.csv")
plot(PlotTf$Tf~PlotTf$Pred,xlab="Tf(est) en K",ylab="Tf(exp) en K",type="p",main=
"Tf(exp)/Tf(est)")
panel.first = grid(col="black")
plot(PlotTf$Res~PlotTf$Tf,xlab="Tf(exp) en K",ylab="Résidus en K",type="p",main=
"Résidus/Tf(exp)")
panel.first = grid(col="black")
plot(PlotTf$Res~PlotTf$Pred,xlab="Tf(est) en K",ylab="Residus en K",type="p",main=
"Résidus/Tf(est)")
panel.first = grid(col="black")
```


Pression critique P_c **# 1. Régression non linéaire**

```
P1<-fitModel((cis +trans+ ortho+ meta+para+Inter1+...+ Inter242+Inter243+Inter244+
Inter245)~(a/(Pc+b)),data=Pc)
coef(P1)
model(P1)
```

#2. Régression linéaire

```
trainPc <- Pc[1:681,]
testPc <- Pc[682:756,]
P<-lm(mod~(cis +trans+ ortho+ meta+ para+ Inter1+...+ Inter242+Inter243+Inter244+
Inter245),data=trainPc)
summary(P)$r.squared
summary(P)$adj.r.squared
summary(P)$residuals
coefficients(P)
fitted.values(P)
write.csv(P$coef, "C:\\Users\\khaled\\Desktop\\R data\\Les contributions\\coefPc.csv")
write.csv(P$fit, "C:\\Users\\khaled\\Desktop\\R data\\Les contributions\\fitPc.csv")
plot(PlotP$Pc~PlotP$Pred,xlab="Pc(est) en bar",ylab="Pc(exp) en bar",type="p",main=
"Pc(exp)/Pc(est)")
panel.first = grid(col="black")
plot(PlotP$Res~PlotP$Pc,xlab="Pc(exp) en bar",ylab="Résidus en bar",type="p",main=
"Résidus/Pc(exp)")
panel.first = grid(col="black")
plot(PlotP$Res~PlotP$Pred,xlab="Pc(est) en bar",ylab="Residus en bar",type="p",main=
"Résidus/Pc(est)")
panel.first = grid(col="black")
```

Température Critique T_c**#1. Régression non linéaire**

```
Tc1<-fitModel((cis +trans+ ortho+ meta+para+Inter1+...+Inter242+Inter243+Inter244+  
Inter245)~((Tb/Tc)+a)^0.5,data=Tc)
```

```
coef(Tc1)
```

```
model(Tc1)
```

#2. Régression linéaire

```
trainTc <- Tc[1:730,]
```

```
testTc <- Tc[731:811,]
```

```
Tc<-lm(mod~(cis +trans+ ortho+ meta+ para+ Inter1+...+ Inter242+Inter243+Inter244+  
Inter245),data=trainTc)
```

```
summary(Tc)$r.squared
```

```
summary(Tc)$adj.r.squared
```

```
summary(Tc)$residuals
```

```
coefficients(Tc)
```

```
fitted.values(Tc)
```

```
write.csv(Tc$coef, "C:\\Users\\khaled\\Desktop\\R data\\Les contributions\\regTc.csv")
```

```
plot(PlotTc$Tc~PlotTc$Pred,xlab="Tc(est) en K",ylab="Tc(exp) en K",type="p",main=  
"Tc(exp)/Tc(est)")
```

```
panel.first = grid(col="black")
```

```
plot(PlotTc$Res~PlotTc$Tc,xlab="Tc(exp) en K",ylab="Résidus en K",type="p",main=  
"Résidus/Tc(exp)",col="black")
```

```
panel.first = grid(col="black")
```

```
plot(PlotTc$Res~PlotTc$Pred,xlab="Tc(est) en K",ylab="Residus en K",type="p",main=  
"Résidus/Tc(est)")
```

```
panel.first = grid(col="black")
```

Volume Critique V_c**#1. Régression linéaire**

```

trainVc <- Vc[1:657,]
testVc <- Vc[658:730,]
Vc<-lm(Vc~(cis +trans+ ortho+ meta+ para+ Inter1+...+ Inter242+Inter243+Inter244+
Inter245),data=trainVc)
summary(Vc)$r.squared
summary(Vc)$adj.r.squared
summary(Vc)$residuals
coefficients(Vc)
fitted.values(Vc)
write.csv(Vc$coef, "C:\\Users\\khaled\\Desktop\\R data\\Les contributions\\regVc.csv")
plot(PlotV$Vc~PlotV$Pred,xlab="Vc(est)      en      cm3/mol",ylab="Vc(exp)      en
cm3/mol",type="p",main= "Vc(exp)/Vc(est)")
panel.first = grid(col="black")
plot(PlotV$res~PlotV$Vc,xlab="Vc(exp)      en      cm3/mol",ylab="Résidus      en
cm3/mol",type="p",main= "Résidus/Vc(exp)",col="black")
panel.first = grid(col="black")
plot(PlotV$res~PlotV$Pred,xlab="Vc(est)      en      cm3/mol",ylab="Residus      en
cm3/mol",type="p",main= "Résidus/Vc(est)")
panel.first = grid(col="black")

```

Enthalpie de fusion ΔH_{fus} **# 1. Régression non linéaire**

```
Hf1<-fitModel((cis+ trans+ ortho+ meta+para+Inter1+...+Inter242+Inter243+Inter244+
Inter245)~((Tf/Hf)+a)^0.5,data=Hf)
```

```
coef(Hf1)
```

```
model(Hf1)
```

2. Régression linéaire

```
trainHf <- Hf[1:451,]
```

```
testHf <- Hf[452:501,]
```

```
Hf<-lm(mod~(cis +trans+ ortho+ meta+ para+ Inter1+...+ Inter242+Inter243+Inter244+
Inter245),data=trainHf)
```

```
summary(Hf)$r.squared
```

```
summary(Hf)$adj.r.squared
```

```
summary(Hf)$residuals
```

```
coefficients(Hf)
```

```
fitted.values(Hf)
```

```
predict(Hf, newdata=testHf)
```

```
write.csv(Hf$coef, "C:\\Users\\khaled\\Desktop\\R data\\Les contributions\\coefHf.csv")
```

```
write.csv(Hf$fit, "C:\\Users\\khaled\\Desktop\\R data\\Les contributions\\fitHf.csv")
```

```
plot(PlotHf$Hf~PlotHf$Pred,xlab="ΔHfus(est) en kj/mol",ylab="ΔHfus(exp) en
kj/mol",type="p",main= "ΔHfus(exp)/ΔHfus(est)")
```

```
panel.first = grid(col="black")
```

```
plot(PlotHf$Res~PlotHf$Hf,xlab="ΔHfus(exp) en kj/mol",ylab="Résidus en
kj/mol",type="p",main= "Résidus/ΔHfus(exp)")
```

```
panel.first = grid(col="black")
```

```
plot(PlotHf$Res~PlotHf$Pred,xlab="ΔHfus(est) en kj/mol",ylab="Residus en
kj/mol",type="p",main= "Résidus/ΔHfus(est)")
```

```
panel.first = grid(col="black")
```

Enthalpie de formation (phase gazeuse) $\Delta H^\circ_{\text{for (gaz)}}$ **# 1. Régression linéaire**

```
trainHfr <- Hfr[1:316,]
testHfr <- Hfr[317:351,]
Hfr<-lm(Hfr~(cis +trans+ ortho+ meta+ para+ Inter1+...+ Inter242+Inter243+Inter244+
Inter245),data=trainHfr)
summary(Hfr)$r.squared
summary(Hfr)$adj.r.squared
summary(Hfr)$residuals
coefficients(Hfr)
fitted.values(Hfr)
predict(Hfr, newdata=testHfr)
write.csv(Hfr$coef, "C:\\Users\\khaled\\Desktop\\R data\\Les contributions\\coefHfr.csv")
write.csv(Hfr$fit, "C:\\Users\\khaled\\Desktop\\R data\\Les contributions\\fitHf.csv")
plot(PlotHfr$Hfr~PlotHfr$Pred,xlab="ΔfH°gaz(est) en kj/mol",ylab="ΔfH°gaz(exp) en
kj/mol",type="p",main= "ΔfH°gaz(exp)/ΔfH°gaz(est)")
panel.first = grid(col="black")
plot(PlotHfr$Res~PlotHfr$Hfr,xlab="ΔfH°gaz(exp) en kj/mol",ylab="Résidus en
kj/mol",type="p",main= "Résidus/ΔfH°gaz(exp)")
panel.first = grid(col="black")
plot(PlotHfr$Res~PlotHfr$Pred,xlab="ΔfH°gaz(est) en kj/mol",ylab="Residus en
kj/mol",type="p",main= "Résidus/ΔfH°gaz(est)")
panel.first = grid(col="black")
```

ملخص :

في هذا العمل تم عرض نماذج جديدة تسمح بتنبؤ الخواص الفيزيائية والحرارية للجزيئات العضوية. وتعتمد هذه النماذج على طريقة نشرت في وقت سابق وهو مساهمة روابط المجموعات التي تم تمديدها من خلال توسيع قاعدة البيانات، إضافة مجموعات هيكلية جديدة ومجموعات إضافية لتصحيح لتقديم المزيد من الدقة والموثوقية وتوسيع نطاق تطبيق النماذج. هذه النماذج تسمح بتقدير الخصائص التالية درجة الغليان، درجة الحرارة الحرجة، الضغط الحرج، الحجم الحرج، نقطة انصهار، المحتوى الحراري الانصهاري، والمحتوى الحراري للتشكيل باستخدام 2312، 811، 756، 730، 1124، 501 و 351 نقطة بيانات تجريبية لهذه الخصائص على التوالي. النهج يقدم تمييزا جيدا بين الصيغ المتشابهة بما في ذلك الروابط *cis* أو *trans* و *ortho*، *meta* أو *para*. أظهرت الإضافات التي أدخلت على المنهج تحسنا كبيرا بالمقارنة مع الطرق التي نشرت في وقت سابق بما في ذلك الطريقة التي استخدمت نفس المنهج.

الكلمات الدلالية : الخواص الفيزيائية والحرارية، النمذجة، التحسين، طرق التنبؤ، مساهمة روابط المجموعات، البترول و الغاز الطبيعي.

Résumé :

Des nouveaux modèles permettent la prédiction des propriétés physiques et thermodynamiques des composés organiques sont présentés. Ces modèles sont basés sur une méthode publiée précédemment, c'est *la contribution des interactions de groupes*. Cette méthode a été étendue par l'élargissement de la base de données, l'addition des nouveaux groupes structuraux ainsi que d'autres groupes pour la correction afin d'améliorer la précision, la fiabilité et d'élargir la gamme d'application. Ces modèles permettent l'estimation des propriétés suivantes : la température d'ébullition (T_{eb}), la température critique (T_c), la pression critique (P_c), le volume critique (V_c), la température de fusion (T_f), l'enthalpie de fusion (ΔH_{fus}) et l'enthalpie de formation (phase gazeuse) ($\Delta_f H^\circ_{gaz}$) en utilisant 2312, 811, 756, 730, 1124, 501 et 351 données expérimentales des ces propriétés respectivement. L'approche offre une bonne distinction entre les isomères y compris les structures *cis* ou *trans* et *ortho*, *meta* ou *para*. Les additions ajoutées à l'approche ont montré une amélioration significative à la méthode en comparaison avec les méthodes déjà publiées y compris la méthode de la même approche.

Mots-clés : Propriétés physiques et thermodynamiques ; Modélisation ; Optimisation ; Méthodes de prédiction ; la contribution des interactions des groupes ; Domaine pétrolier et gaz naturel.

Abstract:

New models allow the prediction of the physical and thermodynamic properties of organic molecules are presented. These models are based on a method published earlier which is the group's interaction contributions which was extended by broadening the data set, adding new structural groups and additional groups correction to offer more accuracy, reliability and extend the range of applicability of the models. These models allow the estimation of the following properties: boiling point (T_b), critical temperature (T_c), critical pressure (P_c), critical volume (V_c), melting point (T_m), enthalpy of fusion (ΔH_{fus}) and enthalpy of formation ($\Delta_f H^\circ_{gas}$), using 2312, 811, 756, 730, 1124, 501 and 351 experimental data point of these properties respectively. The approach offer a good distinction between isomers including *cis* or *trans* and *ortho*, *meta* or *para* structure. The additions made to the approach showed significant improvement in comparison with the earlier published methods including the method of the same approach.

Keywords: Physical and thermodynamic properties; Modeling; Optimization; Prediction methods; Group interaction contributions; Petroleum and natural gas.