



**MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET  
DE LA RECHERCHE SCIENTIFIQUE**

**UNIVERSITE « ABBES LAGHROUR » KHENCHELA**

**Faculté des sciences de la Technologie**

**Département Des Sciences de la matière**

**N° de série : .....**

**Mémoire de Fin d'études**

**Pour l'obtention du diplôme de Master (L.M.D)**

**Filière: Chimie**

**Spécialité: Chimie Analytique et environnement**

**Intitulé :**

**Modélisation du point d'éclair d'un ensemble des  
solvants de type n-alcanes**

**Présenté par:**

**Nouna Hounaida**

**Bourdoucen Meroua**

**Dirigé par :**

**M<sup>me</sup> Lachekheb Messouda : MAA**

**Membres du jury :**

**President: Golea Lynda : M.C.A .**

**Examineur : Baaziz Sonia : M.A.A**

**2021-2022**

# Remerciement

*En premier lieu, nous tenons à remercier ALLAH, notre créateur pour nous*

*Avoir donné la force pour accomplir ce travail.*

*Au terme de ce travail nous tenons à remercier en premier lieu*

*M<sup>me</sup>. Lechkheb Messouda, pour son aide précieuse, pour ses remarques  
Pertinentes et surtout pour le temps qu'il nous a accordé et ceci malgré ses  
Nombreuses charges.*

*Nous tenons à remercier les membres de jury M<sup>me</sup>. Gelea Lynda &*

*M<sup>me</sup>. Baaziz Sonia qui ont bien voulu nous faire l'honneur de  
Participer à ce jury.*

*Et nos familles : Bourdoucen & Nouna*

*En particulier, nos parents pour leurs efforts.*

*Nous remercions également tous les enseignants du département des Sciences de  
La matière.*

*Merci à tous ceux qui nous ont contribués de près ou de loin à l'élaboration de  
Ce mémoire.*

# SOMMAIRE

SYMBOLES ET ABREVIATIONS

LISTE DES TABLEAUX

LISTE DES FIGURES

INTRODUCTION GENERALE..... 2

## LA PARTIE BIBLIOGRAPHIQUE

### CHAPITRE I : MISE AU POINT BIBLIORAPHIQUE SUR LES SOLVANTS

<b>I.1.Introduction</b> .....	<b>7</b>
<b>I.2.Définition d'un solvant</b> .....	<b>7</b>
<b>I.3. Propriétés physico-chimiques</b> .....	<b>8</b>
I.3.1 Densité .....	<b>8</b>
I.3.2: Point d'ébullition.....	<b>8</b>
I.3.3. Tension de vapeur.....	<b>9</b>
I.3.4. Chaleur de vaporisation.....	<b>9</b>
I.3.5. Taux d'évaporation.....	<b>9</b>
I.3.6. Viscosité.....	<b>10</b>
I.3.7. Tension superficielle.....	<b>10</b>
I.3.8. Paramètre de solubilité.....	<b>10</b>
I.3.9. Point d'éclair.....	<b>11</b>
I.3.10. Point d'inflammation.....	<b>11</b>
I.3.11. .... Limite d'inflammabilité.....	<b>12</b>
I.3.12. Explosivité.....	<b>13</b>
I.3.13. Auto-inflammation.....	<b>14</b>
<b>I.4. Comment mesurer le point d'éclair ?</b> .....	<b>14</b>
I.4.1. Domaine d'application.....	<b>15</b>
<b>I.5. Classification des solvants selon le règlement CLP</b> .....	<b>15</b>
I.5.1. Comprendre le règlement CLP.....	<b>15</b>

I.5.2. Classification des liquides inflammables selon le règlement CLP.....	16
I.5.2.1. Définition.....	16
I.5.2.2. Critères de classification.....	16
I.5.2.3. Évolution de l'étiquetage des produits chimiques.....	17
I.5.2.4. Résumé et comparaison des méthodes d'évaluation.....	20
<b>I.6. Principales catégories de solvants.....</b>	<b>23</b>
I.6.1. Hydrocarbures.....	23
I.6.2. Solvants halogénés.....	23
I.6.3. Solvants oxygénés.....	23
I.6.4. D'autres solvants.....	23
<b>I.7. La toxicité et les maladies professionnelles des solvants organiques.....</b>	<b>23</b>
<b>I.8. Risque pour l'environnement.....</b>	<b>25</b>
<b>I.9. Quelques accidents.....</b>	<b>25</b>

## CHAPITRE II : ÉTUDE THEORIQUE

<b>II.1. La modélisation moléculaire.....</b>	<b>28</b>
<b>II.2. Optimisation des molécules.....</b>	<b>28</b>
II.2.1. La Méthode de HARTREE-FOCK-ROOTHAAN (Méthode de HFR).....	28
II.2.1.1. Energie d'un micro système représenté par un déterminant de Slater.....	28
II.2.1.2. Détermination des Orbitales ou équations de Hartree-Fock.....	31
II.2.1.3. Equations de Roothaan et Hall.....	32
II.2.1.4. Quelques remarques sur les processus de résolution des équations de Hartree-Fock-Roothaan.....	33
II.2.1.5. Détermination des intégrales de la méthode de Hartree-Fock-Roothaan (HFR). .....	34
II.2.2. Les méthodes semi-empiriques.....	35
II.2.2.1. Définition du semi-empirisme.....	35
II.2.2.2. Quelques théories semi-empiriques.....	35

II.2.2.3. Limites et avantages des méthodes semi-empiriques [II.18].....	38
II.2.3. Analyse des distributions de charges.....	41
II.2.3.1. Analyse de population de Mulliken.....	42
II.2.3.2. Calcul du moment dipolaire.....	43
II.2.3.3. Application.....	44
<b>II.3. La mécanique moléculaire.....</b>	<b>46</b>
II.3.1. Pas de calculs de champ de force sans définition préalable des types d'atomes	46
II.3.2. Forme fonctionnelle des champs de force courants	47
II.3.3. Quelques exemples	47
<b>II.4. Génération des descripteurs moléculaires</b>	<b>48</b>
<b>II.5. Méthodes appliquées pour la sélection d'échantillons</b>	<b>49</b>
II.5.1. Sélection aléatoire des échantillons	50
II.5.2. Algorithme CADEX pour la sélection d'échantillons.....	50
<b>II.6. Sélection d'un sous-ensemble de descripteurs significatifs.....</b>	<b>50</b>
II.6.1. Principe de sélection par algorithme génétique.....	51
II.6.2. Initialisation aléatoire du modèle.....	51
II.6.3. Étape de croisement.....	51
II.6.4. Étape de mutation.....	52
II.6.5. Conditions d'arrêt.....	52
<b>II.7. Développement des modèles QSPR/QSAR.....</b>	<b>52</b>
II.7.1. La régression linéaire multiple (RLM).....	53
II.7.2. Machines à vecteurs supports SVM.....	53
II.7.3. Paramètres statistiques d'évaluation d'un modèle QSAR/QSPR.....	54
II.7.4. Analyse des résidus pour la détection des observations aberrantes.....	58
II.7.5. Diagnostics d'influence.....	59
II.7.6. Statistique $DFBETAS_{ji}$ .....	60

II.7.7. COVRATIO <sub>i</sub> .....	60
II.7.8. Test de randomisation.....	60
II.7.9. Le domaine d'application du modèle QSPR.....	61

## **PARTIE PRATIQUE**

### **CHAPITRE III : RESULTATS ET DISCUSSIONS**

<b>III.1. Modélisation du point d'éclair d'un ensemble de n-alcane</b> .....	<b>64</b>
III.1.1. Introduction.....	64
III.1.2. Données et méthodes de recherches.....	64
III.1.3. Résultats et discussions.....	68
III.1.3.1. Calcul du modèle.....	68
III.1.3.2. Analyse des résidus.....	72
III.1.3.3. Diagnostics d'influence.....	78
III.1.3.4. Domaine d'applicabilité.....	82
III.1.3.5. Test de randomisation.....	83
III.1.3.6. Validation statistique externe.....	83
III.1.4. Conclusion.....	85
<b>CONCLUSION GENERALE</b> .....	<b>87</b>
<b>REFERENCES BIBLIOGRAPHIQUES</b> .....	<b>89</b>

## LISTE DES ABREVIATIONS

- **AG** : Algorithme génétique.
- **AM1** : Austin Model 1.
- **CAS** : Chemical Abstracts Service.
- **CGS** : Centimètre-Gramme-Seconde.
- **CLOA**: Combinaison Linéaire des Orbitales Atomiques.
- **CLP** : Classification, emballage et étiquetage (En anglais : Classification, Labelling, Packaging).
- **CMR** : Molécules Cancérogènes, Mutagènes et Reprotoxiques.
- **CNDO** : Complete Neglect of Differential Overlap.
- **COC** : Cleveland Open Cup (pour, Coupelle ouverte).
- **COV** : Composés Organiques Volatils.
- **DFITS** : Statistique permettant de mesurer l'influence d'une observation  $i$  sur la valeur ajustée.
- **$D_i$**  : Distance de COOK.
- **$d$**  : Statistique de Durbin-Watson.
- **$D_i$** : Résidu standardisé.
- **DNA** : DeoxyriboNucleic Acid.
- **DP** : Directive Préparations Dangereuses.
- **ECOSO** : Conseil économique et social des Nations unies.
- **EQM** : Ecart quadratique moyen.
- **EQMC** : Ecart quadratique moyen calculé sur l'ensemble de calibrage.
- **EQMP** : Ecart quadratique moyen de prédiction.
- **EQMP<sub>ext</sub>** : Ecart quadratique moyen calculé sur l'ensemble de validation externe.
- **ERENAV** : Entreprise Nationale d'Entretien et de Réparation Navale.
- **$e_i$**  : Résidu : différence entre les valeurs observée ( $y_i$ ) et estimée ( $\hat{y}_i$ ).
- **F** : Statistique de Fisher.
- **INDO** : Intermediale Neglect of Differential Overlap.
- **GA** : Genetic Algorithm ( pour, Algorithme génétique).
- **HF** : Hartree –Fock.
- **HFR** : Hartree -Fock-Roothan.
- **hii** : Eléments diagonaux de la matrice chapeau.
- **HOMO** : Highest Occupied Molecular Orbital.
- **LII** : Limite inférieure d'inflammabilité.
- **LMO** : Cross-validation by leave-many-out: Validation croisée par omission d'un ensemble d'observations.
- **LOO** : Cross-validation by leave-one-out: Validation croisée par omission d'une observation.

- **LSI** : Limite supérieure d'inflammabilité.
- **LUMO** : Lowest Unoccupied Molecular Orbital.
- **MM** : Mécanique Moléculaire.
- **N** : Dimension de la population (échantillon).
- **n-p** : Nombre de degrés de liberté.
- **NDDO** : Neglect of Diatomic Differential Overlap
- **OM** : Orbitales Moléculaires.
- **PLS(ou MCP)** : Moindres carrés partiels.
- **PM3** : Parametrization Method 3.
- **PRESS** : Somme des carrés des erreurs de prédiction.
- **p** : Nombre de descripteurs en comptant la constante (Nombre de paramètres).
- **PPP** : Pople-Pariser-Parr.
- **QSAR** : Quantitative Structure/ Activity Relationships. (Relations Quantitatives Structure/ Activité).
- **QSPR** : Quantitative Structure/ Property Relationships. (Relations Quantitatives Structure/ Propriété).
- $Q_{boot}^2$  : Coefficient de prédiction par la technique du bootstrap.
- $Q_{Loo}^2$  : Coefficient de prédiction par leave one out.
- **REACH** : enRégistrement, Evaluation et Autorisation des produits Chimiques.
- **RLM (MLR)** : Régression linéaire multiple.
- **RMSE** : Racine de l'écart quadratique moyen ( Root Mean Squared Error).
- $R^2$  : Coefficient de détermination.
- $r_i$  : Résidu studentisé interne.
- **S** : Erreur standard.
- **SCE** : Somme des carrés des écarts.
- **SCF** : Self Consistent Field.
- **SGH** : Système général harmonisé.
- **SVM** : Support Vector Machine (machine à vecteur support)
- **t** : t de Student.
- $t_i$  : Résidu studentisé externe.
- **Y<sub>i</sub>** : Valeur observée.
- $\hat{y}_i$  : Valeur estimée.
- $\hat{y}_{(i)}$  : Valeur prédite.

## LISTE DES FIGURES

<b>Chapitre</b>	<b>Figure</b>	<b>Page</b>
<b>I</b>	<b>Figure I.1:</b> Les six conditions pour une explosion	<b>14</b>
	<b>Figure I.2:</b> Signification d'une étiquette.	<b>20</b>
<b>II</b>	<b>Figure II.1:</b> Déterminants de Slater excités générés à partir d'une référence HF	<b>40</b>
	<b>Figure II.2:</b> Les indices électroniques de la méthode des orbitales moléculaires et leurs applications, d'après	<b>45</b>
<b>III</b>	<b>Figure III.1:</b> Le montage à reflux utilisé	<b>65</b>
	<b>Figure III.2:</b> Droite d'ajustement des Tec prédites en fonction des Tec expérimentales pour les ensembles de calibrage et de test	<b>71</b>
	<b>Figure III.3:</b> Diagramme de Williams	<b>82</b>
	<b>Figure III.4:</b> Test de randomisation	<b>83</b>

## LISTES DES TABLEAUX

Chapitre	Tableaux	Page
<b>I</b>	<b>Tableau I.1:</b> Critères applicables aux liquides inflammables.	<b>17</b>
	<b>Tableau I.2:</b> Les nouveaux pictogrammes selon le règlement CLP et leur signification .	<b>18</b>
	<b>Tableau I.3:</b> Résumé et comparaison des méthodes d'évaluation .	<b>22</b>
	<b>Tableau I.4:</b> Toxicité de quelques solvants organiques appartenant à différentes familles .	<b>24</b>
<b>III</b>	<b>Tableau III.1:</b> Composés étudiés et valeurs des températures d'éclair expérimentales.	<b>66</b>
	<b>Tableau III.2:</b> Classe et définition des deux descripteurs moléculaires.	<b>70</b>
	<b>Tableau III.3:</b> Résidus caractéristiques et valeurs estimées des points d'éclair.	<b>73</b>
	<b>Tableau III.4:</b> Diagnostics d'influence.	<b>78</b>
	<b>Tableau III.5:</b> Valeurs des paramètres statistiques pour tout l'ensemble étudié.	<b>84</b>
	<b>Tableau III.6:</b> Valeurs des paramètres statistiques pour le modèle inverse.	<b>84</b>

# **Introduction Générale**

# Introduction Générale

---

## Introduction générale :

Un danger chimique peut se mesurer par la toxicité du produit ou par sa capacité à produire des dégâts et des effets néfastes pour la santé humaine et l'environnement. Ces effets sur la santé peuvent concerner aussi bien le travailleur qui les produit, ou les utilise, que le consommateur final. Plus généralement, c'est l'ensemble de la population qui peut être exposée via le relargage de substances dans l'environnement.

Chaque année on note l'apparition d'un milliard de nouvelles molécules. Vingt deux millions seulement sont répertoriées auprès de la banque de données du Chemical Abstracts Service (CAS), cent mille sont commercialisées dont seulement 5% ont des propriétés (physicochimiques, activités biologiques, etc...) connues [1].

Afin de mieux connaître et maîtriser les risques liés à l'utilisation de ces produits chimiques, plusieurs réglementations, essentiellement d'origine communautaire, les encadrent en fonction de leurs usages. On distingue les règlements européens, adoptés au niveau de l'Union Européenne (UE) et applicables directement, les directives, transposées en droit français pour être applicables, d'autres enfin sont purement nationales.

Le règlement, l'enregistrement, l'évaluation et l'autorisation des substances chimiques (REACH), entré en vigueur en Europe le 1er juin 2007, vise à collecter des informations sur les propriétés des substances chimiques produites ou importées en quantités supérieures à 1 tonne/an. Cette réglementation complexe impose des obligations strictes aux industriels et constitue un outil essentiel des pouvoirs publics et de la société civile pour améliorer durablement le bien-être de la population en termes de santé et d'environnement.

Une autre mesure réglementaire importante concerne la classification, l'étiquetage et l'emballage des produits chimiques : il s'agit du règlement européen 1272/2008 CE, dit CLP (Classification, Labelling, Packaging en anglais). Sa réglementation est entrée en vigueur le 20 janvier 2009 et a vocation à être pleinement applicable aux substances au 1er décembre 2010 et pleinement applicable aux mélanges au 1er juin 2015. Ce règlement européen s'appuie sur les dispositions du Système général harmonisé (SGH) promu par le Conseil économique et social des Nations unies (ECOSOC) en juillet 2003[2].

En Algérie [3], l'arrêté interministériel 13 Safar 1437, correspondant au 25 novembre 2015, établit la liste et la classification des matières et produits chimiques dangereux (art. 1, 2, 3). Le présent arrêté a pour objet d'établir des inventaires et des classements Substances et produits chimiques dangereux. La liste référencée donne à chaque matière ou produit chimique dangereux son numéro d'identification selon la classification des Nations Unies (ONU) et ses principales classes de risque telles que définies à l'article 3.

# Introduction Générale

---

Les propriétés physico-chimiques sont utilisées pour caractériser les produits chimiques qu'ils soient naturels ou de synthèse. En prévention, elles permettent d'évaluer la dangerosité d'un produit chimique (inflammabilité, volatilité, pouvoir corrosif, pouvoir oxydant...) et les risques qui en découlent (que ce soit lors de sa manipulation, son stockage ou son élimination) tels que les risques d'incendie, les risques d'explosion, corrosivité, réactivité chimique et instabilité...

La connaissance des caractéristiques des produits (température d'ébullition, point d'éclair, pression de vapeur saturante, température d'auto-inflammation, domaine d'inflammabilité ou d'explosivité) est essentielle pour la maîtrise du risque. De nombreux exercices de mises en situation permettent d'appliquer immédiatement ces acquis théoriques.

Dans la plupart des cas, il est excessivement cher d'obtenir de telles informations expérimentalement et le recours à l'expérience pour pouvoir les identifier devient impossible. Par conséquent, les compagnies et les agences régulatrices se tournent vers la prédiction de ces propriétés à travers l'usage des relations quantitatives structure / propriété (QSPR).

Les Relations Quantitatives Structure- Activité/ Propriété (QSAR/QSPR) sont devenues un puissant outil théorique, alternatif à la mécanique quantique, pour la description et la prédiction des propriétés des systèmes moléculaires complexes dans différents environnements.

Les relations entre les structures des molécules et leurs propriétés ou activités sont généralement établies à l'aide de méthodes de modélisation par apprentissage statistique. Les techniques usuelles reposent sur la caractérisation des molécules par un ensemble de descripteurs, nombres réels mesurés ou calculés à partir des structures moléculaires. Il est alors possible d'établir une relation entre ces descripteurs et la grandeur modélisée [4].

Dans le présent travail, nous nous sommes intéressés à la modélisation d'une propriétés physico-chimiques: la température d'éclair d'une ensemble des solvants n-alcanes . Nous avons appliqué les techniques QSPR.

Notre mémoire comprend Trois chapitres :

- Dans le premier chapitre, nous avons présenté des généralités sur les solvants
- Le deuxième chapitre expose la base théorique exploitée. Nous avons décrit la modélisation moléculaire et ses principes. Nous avons également développé les connaissances théoriques de base utilisées tout au long de ce travail: algorithmes génétiques, régression multilinéaire, machine à vecteurs de support, et les paramètres statistiques utilisés pour l'évaluation de la qualité des modèles obtenus.

## Introduction Générale

---

- Dans le troisième chapitre, nous présentons et nous discutons des modèles développés :
  - Un modèle AG/MLR développé pour modéliser la température d'éclair d'une série d'hydrocarbures non-saturés.
  - Un modèle AG/MLR développé pour modéliser la température d'éclair d'une série de n-alcanes.
  - Deux modèles : un modèle linéaire AG/MLR et l'autre non linéaire AG/SVR pour la modélisation d'une série de solvants organiques formés de différentes classes, avec une comparaison des résultats obtenus selon les deux modèles.

Une analyse des résidus a été effectuée pour les trois modèles linéaires pour les trois ensembles étudiés dans le but de vérifier la validité du modèle d'une part et pour repérer les observations aberrantes et les observations qui jouent un rôle important dans la détermination de la régression.

# LA PARTIE THEORIQUE

**Chapitre I:**  
**MISE AU POINT**  
**BIBILIGRAPHIQUE SUR LES**  
**SOLVANTS**

## I.1.Introduction

Le champ d'application des solvants est des plus étendu car l'économie moderne y a recours pour la préparation d'une foule de produits nouveaux dont l'utilisation déborde dans la vie courante.

Chaque année le nombre de ces substances mises sur le marché et à la disposition des industriels augmente à travers le monde. Utilisés dans de nombreux domaines, les solvants organiques entrent dans la composition de divers produits tels que les peintures, les encres, les pesticides, les dégraissants, la production des textiles artificiels, les diluants et les colles. Parmi les matières premières utilisées au cours de la fabrication, les adhésifs, notamment les adhésifs solides et liquides naturels et les solutions adhésives préparées à partir de solvants organiques, représentent les plus importants risques professionnels [I.1]. Depuis longtemps, les procédés d'extraction ont employé les solvants comme agent d'extraction.

## I.2.Définition d'un solvant

Ayant à l'esprit le cas de l'utilisation des solvants en vue de la fabrication des laques cellulosiques, Durrans [I.2] propose la définition suivante : « Le solvant est un moyen de transfert d'un solide d'une place à une autre de manière convenable. Lorsque le transfert a été réalisé, le solvant ne présente plus d'intérêt ; on l'élimine donc aussi rapidement et aussi complètement que possible ».

Mais cette définition a le défaut de viser surtout une application particulière. Une définition technique plus large pourrait être celle-ci: « Un solvant est un composé volatil liquide, capable de se charger d'autres substances en ne formant qu'une seule phase liquide, ledit composé étant capable de restituer, par simple évaporation, ces substances inaltérées et amenées éventuellement sous un état physique plus convenable [I.3-p3].

En terme général, un solvant est une substance qui sert à dissoudre une autre substance. Dans le contexte industriel, on se limite aux solvants organiques, c'est-à-dire ceux qui contiennent au moins un atome de carbone dans leurs structures moléculaires. D'après Cohr [I.4], un solvant organique est un composé chimique ou mélange qui est liquide entre 0°C et 250°C approximativement, qui est volatil et relativement inerte chimiquement.

Les solvants sont utilisés industriellement pour extraire, dissoudre ou suspendre des substances généralement insolubles dans l'eau (l'eau n'est donc pas un solvant organique) ou pour modifier les propriétés physiques d'un matériau.

Le concept de solvant organique ne doit pas être confondu avec celui des corps organiques volatils (COV) que l'on retrouve dans la réglementation environnementale visant à protéger la qualité de l'atmosphère .Le terme solvant a une dimension utilitaire alors que les COV sont définis en terme de réactivité photochimique

dans l'atmosphère et de tension de vapeur minimale , généralement autour de 13,3 Pa (soit 0,1 mm Hg) à 25°C [I.5].

### I.3 .Propriétés physico-chimiques

En plus du coût, les propriétés physico-chimiques liées aux performances techniques, est un paramètre clé qui détermine le type d'utilisation industrielle du solvant, leur mise en œuvre. Ils peuvent également prédire certains de leurs comportements environnement, ces caractéristiques sont particulièrement pertinentes pour les risques de manipulation, d'explosion et d'incendie.

#### I.3.1. Densité

La densité d'une substance est égale à la densité de la substance divisée par la densité de la substance Densité du corps de référence à la même température. pour liquides et pour les solides, l'eau est utilisée comme référence, pour les gaz, la mesure est relative à air. Elle est représenté par **d** et n'a pas d'unité (grandeur physique sans dimension).

$$d = \frac{\rho_{\text{Substance}}}{\rho_{\text{H}_2\text{O}}} \quad (\text{I.1})$$

La masse volumique de l'eau est mesurée à une température de 4°C, correspondant à une La température à laquelle la densité atteint sa valeur maximale. Cette température est donnée par 4. Références en indice. Le symbole devient alors  $d_4$ . Pour des raisons pratiques, La densité de la matière est mesurée à température ambiante, et Généralement à 20°C. Par conséquent, la masse volumique d'un solide ou d'un liquide est généralement enregistrée en indiquant 2 températures :  $d_4^{20}$  cela signifie donc « la densité de matière par 20°C Par rapport à l'eau à 4°C » .

A l'exception des solvants halogénés, la plupart des solvants sont plus légers que l'eau. Cela explique pourquoi la plupart des feux de solvants ne peuvent pas être éteints par l'eau .

#### I.3.2 Point d'ébullition

Le point d'ébullition est la température à laquelle la pression de vapeur du liquide est égale à celle de la pression atmosphérique normale (101,325 kPa), soit la température à laquelle le solvant passe de l'état liquide à l'état gazeux. L'unité de mesure de cette variable dans le système international est le Kelvin mais, en pratique, on utilise le degré Celsius.

Le contrôle de la température d'ébullition à la pression ordinaire est facile à faire au laboratoire et ne demande que quelques instants.

Les différents solvants sont classés suivant leur ordre de volatilité en 3 catégories : solvants légers ( $T_{eb} < 100^{\circ}\text{C}$ ); solvants moyens ( $100^{\circ}\text{C} < T_{eb} < 150^{\circ}\text{C}$ ); solvants lourds ( $T_{eb} > 150^{\circ}\text{C}$ ). Il faut souligner que la présence de certaines impuretés est capable de modifier le comportement d'un solvant lors de la distillation; elles peuvent en particulier abaisser son point d'ébullition et le faire ainsi passer d'une classe à la classe inférieure [I.3-p31].

### I.3.3. Tension de vapeur

La tension ou pression de vapeur saturante est la pression exercée par la vapeur lorsqu'elle est à l'équilibre avec le liquide. L'unité de mesure de cette variable est le kilopascal (kPa) et on utilise souvent le millimètre de mercure (mmHg ou Torr,  $1\text{kPa} = 7,500\text{ mmHg}$ ). La tension de vapeur est rapportée le plus souvent à  $25^{\circ}\text{C}$  ; plus la tension de vapeur est élevée et plus le solvant a une tendance naturelle à s'évaporer.

La détermination des tensions de vapeur d'un liquide nécessite un appareillage spécial, et demande beaucoup plus de soin et de temps que celle d'une température d'ébullition. On opère le plus souvent avec une chambre barométrique à mercure. Il faut une parfaite uniformisation de la température au moment de la mesure de la pression de vapeur et il est souhaitable d'opérer sur des échantillons de produits purs [I.3-p22].

D'une manière générale il est évident que les liquides sont d'autant plus dangereux au point de vue de l'inflammabilité que leur tension de vapeur est plus forte.

### I.3.4. Chaleur de vaporisation

La chaleur de vaporisation ou chaleur latente de vaporisation d'un solvant est la quantité de chaleur nécessaire pour vaporiser une quantité définie de solvant.

Ses unités sont les kilojoules par mole (kJ/mol) dans le système international et Calories par mole (kcal/mol) dans un système CGS. Ce paramètre est particulièrement utile pour comprendre les besoins énergétiques des solvants dans le domaine du dégraissage à la vapeur des surfaces métalliques.

### I.3.5. Taux d'évaporation

Le taux d'évaporation absolu d'un solvant est la quantité de matière qui s'évapore d'une surface par unité de temps. Dans la littérature, les taux d'évaporation relatifs sont rapportés comme solvant de référence, l'éther éthylique ou l'acétate de butyle normal.

Cette variable, qui dépend des conditions environnementales, n'a pas d'unités. Les taux d'évaporation des solvants sont largement utilisés dans la formulation de revêtements, d'adhésifs et d'encres.

### I.3.6. Viscosité

Le coefficient de viscosité ou viscosité dynamique est défini comme la force nécessaire pour déplacer une surface d'un même liquide à une vitesse de 1 cm<sup>2</sup> par rapport à un autre plan qui lui est parallèle à une distance de 1 cm [4]. L'unité de viscosité dynamique dans le système international est le pascal seconde (**Pa.s**). En pratique, on utilise aussi la poise (1P = 0.1 Pas = 0.1 Ns/m<sup>2</sup> = 1 dyne.s/cm<sup>2</sup>) et la centipoise (**cP**). La viscosité est importante lors de la formation d'un mélange ainsi que lors du choix du mode d'application, par exemple par pulvérisation.

### I.3.7. Tension superficielle

La tension superficielle est définie comme la résultante des forces intermoléculaires s'exerçant sur les molécules à la surface du liquide. On peut imaginer que les molécules situées à l'intérieur du liquide sont soumises à des forces différentes de celles subies par sa surface: en effet, dans la phase liquide contrairement à la phase gazeuse, les molécules se touchent. Bien souvent, ces molécules ne sont pas électriquement neutres. Il s'ensuit que la résultante de ces forces ne peut pas être nulle.

Lorsque deux liquides dissemblables sont en contact, ces forces intermoléculaires modifieront la forme de l'interface jusqu'à ce que l'énergie potentielle de tout le système moléculaire atteigne un minimum.

Ses unités de mesure sont : dans le système international le newton par mètre (N/m), et dans le système CGS la dyne par centimètre (1 dyne/cm = 0,001N/m).

La tension superficielle est un paramètre utile à connaître pour évaluer la tendance que possède un liquide à s'étaler sur une surface.

### I.3.8. Paramètre de solubilité

La capacité de solubilisation d'un solvant pour un soluté donné, également appelé son pouvoir solvant, est une donnée essentielle d'un solvant.

Suite à une étude faite en 2005 [I.6], le concept de paramètre de solubilité a été introduit dans des travaux théoriques développés pendant la première moitié du vingtième siècle à partir d'une base thermodynamique par Joël Hildebrand. Il définit le paramètre de solubilité global d'une substance comme étant la racine carrée de l'énergie molaire de cohésion du système par unité de volume (volume molaire).

Le paramètre de solubilité est exprimé en  $(\text{J}/\text{cm}^3)^{1/2}$  et est fonction de l'énergie molaire d'évaporation (E) et du volume molaire du liquide (V).

Un solvant dissout bien le soluté lorsque leurs paramètres de solubilité sont identiques ou ayant un pouvoir de solubilisation très proche. Il faut observer qu'au point de vue incendie, mieux vaut avoir à faire à des liquide solubles dans l'eau qu'à des corps non miscibles. En effet les premiers donnent naissance à des foyers qu'on peut attaquer efficacement avec la lance à eau classique et la dilution qui résulte de cet arrosage abondant contribue directement à diminuer la puissance de la flamme. Au contraire, lorsqu'il s'agit de liquides insolubles dans l'eau, comme les hydrocarbures par exemple, même l'envoi d'une grande quantité d'eau ne suffit pas à éteindre l'incendie. Au contraire, le plus souvent, l'arrosage à la lance ne fait qu'accroître la surface du liquide enflammé en contact avec l'air, tout en risquant d'augmenter le danger de propagation du sinistre par écoulement des liquides enflammés dans les caniveaux de l'usine.

### **I.3.9. Point d'éclair**

Le point d'éclair est la température la plus basse, corrigée pour une pression de 101,325 kPa, à laquelle le liquide d'essai dégage des vapeurs, dans les conditions définies dans la méthode d'essai, en quantité telle qu'il en résulte dans le récipient d'essai un mélange vapeur/air inflammable [I.7].

C'est le phénomène qui se produit lorsqu'un liquide organique déterminé a été porté à une température suffisante pour que les vapeurs émises puissent être enflammées mais sans que la flamme persiste ou que le liquide s'enflamme au contact de l'air ambiant par l'approche d'une petite flamme. Le point d'éclair caractérise en quelque sorte la première manifestation d'un liquide dans l'aptitude à l'inflammation. Sa détermination se ramène à la mesure d'une température.

Au point d'éclair le mélange solvant - air contient deux fois plus d'oxygène que celui que nécessiterait la combustion complète [I.3-p55].

### **I.3.10. Point d'inflammation**

Le point d'inflammation est la température la plus basse à laquelle un liquide émet suffisamment de vapeurs pour former avec l'air ambiant un mélange inflammable dont la combustion une fois débutée puisse s'entretenir d'elle-même après retrait de la source d'allumage. Il est supérieur au point d'éclair de quelques degrés [I.3].

### I.3.11. Limite d'inflammabilité

Avant de définir la limite d'inflammabilité, nous allons rappeler quelques définitions [I.8]:

#### a. Combustible

Toute substance susceptible de brûler, c'est-à-dire pouvant être partiellement ou totalement détruite par le feu, est considérée comme combustible. Les solides et les liquides ne brûlent pas en tant que tels. Ce sont les gaz et les vapeurs qu'ils émettent qui brûlent.

#### b. Comburant

Un comburant est le corps qui provoque et entretient la combustion du combustible; le plus souvent, le comburant est constitué par l'oxygène présent dans l'air ambiant ; la réaction de combustion est alors une oxydation, mais il existe d'autres comburants (halogènes, soufre, phosphore); si l'oxygène est le comburant, sa concentration diminue très rapidement dans l'atmosphère (par phénomène de consommation oxydative) et expose les victimes au risque d'asphyxie. L'oxygène peut se trouver soit à l'état pur, soit en mélange avec d'autres gaz, soit lors de la décomposition de certains produits chimiques. Dans la plupart des cas, le comburant est l'oxygène de l'air ambiant (environ 21 % d'oxygène 79 % d'azote). Pour que l'air soit un comburant efficace, il faut qu'il contienne plus de 15 % d'oxygène.

#### c. Combustion

La combustion est un processus d'oxydation qui se produit par réaction chimique entre deux corps un combustible et un comburant pour donner naissance à un ou plusieurs corps différents des premiers « les produits de combustion ». Il s'agit d'une réaction chimique s'accompagnant d'un dégagement de chaleur. On parle de combustion lente lorsque l'élévation de température devient perceptible, mais sans atteindre une température donnant une lumière visible. Les combustions vives correspondent à une réaction provoquant des températures élevées. À l'extrême, quand la vitesse de propagation devient extrêmement grande, on parle d'explosion.

Les concentrations limites d'un gaz ou d'une vapeur combustible, dans l'air ou dans tout autre comburant en aval et en amont desquelles la propagation de la flamme n'est pas possible, sont appelées «limites d'inflammabilité». Si le mélange est trop pauvre en combustible, l'inflammation ne se produit pas. Le pourcentage est alors au-dessous de la limite inférieure d'inflammabilité (LII). Au-dessus de ce seuil, le mélange combustible comburant pourra brûler tant que l'on n'aura pas dépassé un taux maximum de combustible au-delà duquel le mélange serait trop pauvre en comburant; ce second seuil est la limite supérieure d'inflammabilité (LSI).

L'intervalle entre la limite inférieure et la limite supérieure d'inflammabilité s'appelle «domaine d'inflammabilité ou intervalle d'inflammabilité ». Celui-ci varie fortement selon les gaz ou vapeurs

combustibles, la température, le taux d'oxygène et la pression. Une matière combustible sous forme gazeuse ou de vapeur ne peut exploser que si elle est mélangée à de l'air avec une concentration comprise entre LII et LSI [I.8]. Ces valeurs sont généralement exprimées en pourcentage du volume de gaz inflammable dans le volume total du mélange.

### I.3.12. Explosivité

Les explosions peuvent être soit d'origine physique (par exemple, éclatement d'un récipient dont la pression intérieure est trop grande), soit d'origine chimique, cette dernière est la résultante d'une réaction chimique.

Une réaction d'origine chimique est une réaction rapide de combustion ou de décomposition entraînant une élévation de température et /ou de pression. Les six conditions à remplir pour une explosion sont réunies et symbolisées par un hexagone sont (Figure I.1) :

- La présence d'un combustible ;
- La présence d'une source d'inflammation ;
- La présence d'un comburant(en général l'oxygène de l'air) ;
- Un combustible sous forme gazeuse, d'aérosol ou poussières en suspension ;
- L'obtention d'un domaine d'explosivité (domaine de concentrations de combustible dans l'air comprises entre la LII et LSI à l'intérieur duquel les explosions sont possibles.
- Un confinement suffisant (en l'absence d'un confinement, on obtient un phénomène de combustion rapide sans effet notable de pression, type boule de feu [I.9].

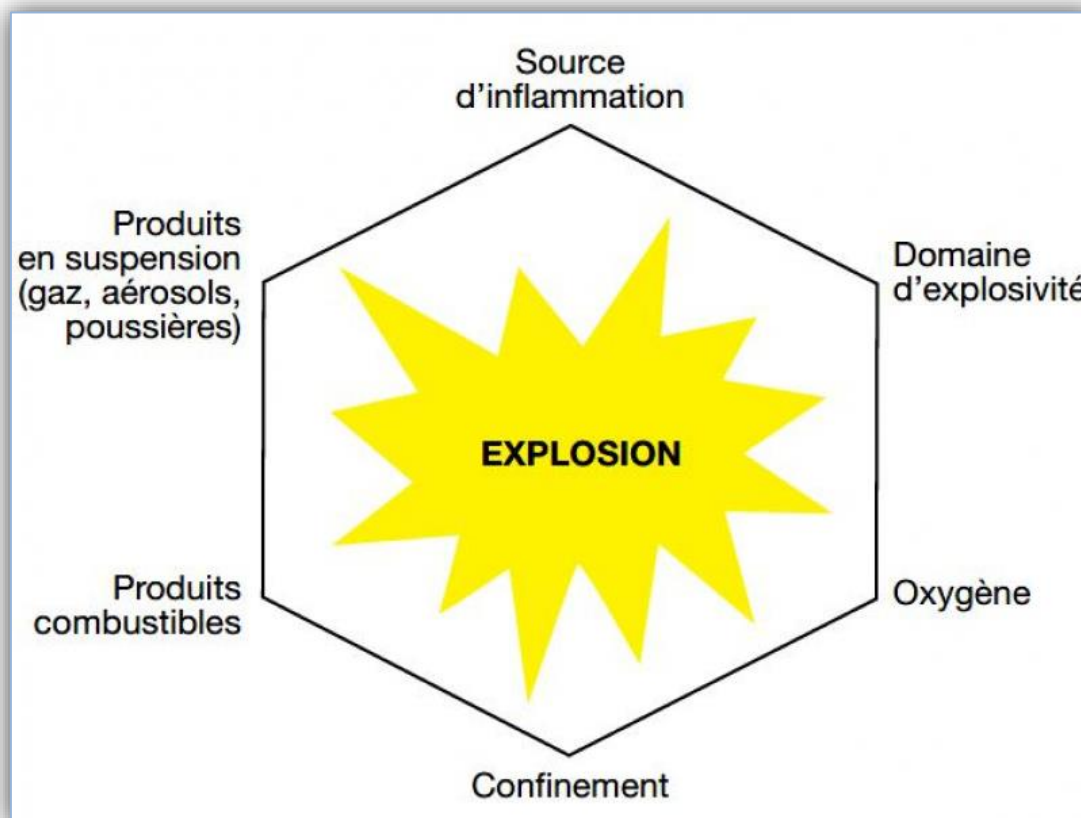


Figure I.1 : Les six conditions pour une explosion.

### I.3.13. Auto-inflammation

Un mélange d'air et de vapeur d'une substance inflammable porté à une température assez élevée peut déclencher une combustion spontanée en l'absence d'une flamme ou d'une étincelle. Ce phénomène est appelé « auto-inflammation ».

Ici, il est important de noter la grande influence de la nature des parois sur le déclenchement de cette réaction d'oxydation. Le verre a généralement une influence plus hâtive que les métaux [I.10].

### I.4. Comment mesurer le point d'éclair ?

La sécurité dans l'utilisation et le stockage de produits inflammables et des mélanges liquides combustibles est bien nécessaire à cause des accidents dramatiques qui se produisent régulièrement sous forme d'une série d'explosions d'huiles essentielles. L'importance de la sécurité du transport des liquides inflammables et des combustibles ainsi que le risque d'explosion des liquides se caractérise principalement par leur point d'éclair ou température d'éclair (Tec). Le point d'éclair est un concept pétrolier et les premiers appareils qui permettent de définir un point d'éclair ont été décrits dans des normes pétrolières. Il faut donc toujours préciser l'appareil qui a été utilisé quand on donne une valeur de point d'éclair. La valeur dépend

non seulement de l'appareil mais également de la bonne vue de l'opérateur qui doit déceler le début d'inflammation. On dénombre deux catégories de test : test à coupelle ouverte (COC: Cleveland Open Cup), et test à coupelle fermée.

Le point d'éclair en coupelle fermée se mesure à l'aide d'un appareil dit de Pensky- Martens (PMCC : Pensky –Martens *Closed Cup*). Un appareil semi-automatique de ce type est constitué d'une coupelle que l'on peut remplir du liquide dont on désire connaître le point d'éclair. On referme ensuite la coupelle. Le couvercle est muni d'un thermomètre dont l'embout se situe au-dessus du liquide dans les vapeurs. L'appareil dispose d'un chauffage qui permet d'élever la température degré par degré. Chaque fois que la température atteint un degré supérieur, une flamme est plongée dans les vapeurs. S'il y a inflammation, c'est que le point d'éclair est atteint, dans le cas contraire l'appareillage continue d'augmenter la température du liquide [I.11].

Le règlement sur les matières dangereuses, prévoit un test d'inflammabilité pour des échantillons liquides contenant ou non des solides en suspension ou en solution. L'inflammabilité d'une matière s'évalue en mesurant la température la plus basse à laquelle les vapeurs du liquide s'enflamment en présence d'une flamme.

Selon le règlement sur les matières dangereuses, toute matière liquide ou semi-liquide, autre qu'une boisson alcoolisée, est considérée comme inflammable si le point d'éclair est égal ou inférieur à 60 °C. De plus, le règlement sur les matières dangereuses permet l'utilisation d'huile usée à des fins énergétiques pour autant que le point d'éclair soit de 38 °C ou plus élevé.

### **I.4.1. Domaine d'application**

Cette méthode s'applique à la détermination du point d'éclair d'une huile usée, d'un liquide ou d'un liquide contenant des solides en suspension ou en solution.

Le domaine d'application se situe entre 25 °C et 80 °C. Un point d'éclair peut être mesuré à plus basse température, en abaissant la température de l'échantillon à l'aide de glace sèche par exemple et en chauffant par la suite.

## **I.5. Classification des solvants selon le règlement CLP**

### **I.5.1. Comprendre le règlement CLP**

CLP est l'abréviation de l'expression anglaise «Classification, Labelling and Packaging» ou «classification, étiquetage et emballage». Le règlement CLP est entré en vigueur en janvier 2009 et la

méthode de classification et d'étiquetage des produits chimiques qu'il introduit repose sur le règlement du Système Général Harmonisé (SGH) des Nations unies.

Le règlement remplace progressivement deux actes législatifs antérieurs, à savoir la directive «Substances dangereuses» (dSD) et la directive «Préparations dangereuses» (dPD). La période de transition a pris fin en 2015.

Le règlement CLP a pour objet d'assurer que les dangers présentés par les substances chimiques soient clairement communiqués aux travailleurs et aux consommateurs de l'Union européenne grâce à la classification et à l'étiquetage des produits chimiques.

Avant de procéder à la mise sur le marché de produits chimiques, l'industrie doit déterminer les risques potentiels de ces substances et mélanges pour la santé humaine et l'environnement et les classer conformément aux dangers identifiés. Les produits chimiques dangereux doivent aussi être étiquetés selon un système normalisé de sorte que les travailleurs et les consommateurs soient informés de leurs effets avant de les manipuler.

Grâce à ce processus, les dangers des produits chimiques sont communiqués en recourant à des mentions types et à des pictogrammes imprimés sur les étiquettes et les fiches de données de sécurité. Ainsi, lorsqu'un fournisseur identifie une substance comme présentant une «toxicité aiguë de catégorie 1 (oral)», l'étiquetage inclura la mention de danger «mortel en cas d'ingestion», le mot «Danger» et le pictogramme comportant une tête de mort et deux tibias [I.12].

## **I.5.2. Classification des liquides inflammables selon le règlement CLP**

### **I.5.2.1. Définition**

Selon les prescriptions relatives à la classification et à l'étiquetage des substances et mélanges dangereux, on entend par « liquide inflammable », un liquide ayant un PE  $\leq 60^{\circ}\text{C}$  [I.12].

### **I.5.2.2. Critères de classification**

Les critères de classification de la classe des liquides inflammables reposent comme dans le cas du système préexistant sur la mesure du PE et du point initial d'ébullition (Teb). On distingue trois catégories dans cette classe en fonction des différentes valeurs indiquées dans le tableau I.1.

**Tableau I.1:** Critères applicables aux liquides inflammables

Catégorie	Critères
1	Le point d'éclair est $< 23^{\circ} \text{C}$ et le point initial d'ébullition est $\leq 35^{\circ} \text{C}$
2	Le point d'éclair est $< 23^{\circ} \text{C}$ et le point initial d'ébullition est $> 35^{\circ}$
3	Le point d'éclair: $23^{\circ} \text{C} < \text{PE} \leq \text{C } 60^{\circ} (*)$

(\*) Aux fins de ce règlement, les gazoles, les carburants diesel et huiles de chauffage légères

dont le point d'éclair compris entre  $55^{\circ} \text{C}$  et  $75^{\circ} \text{C}$  peuvent être considérés comme relevant de la catégorie 3.






### 1.5.2.3. Évolution de l'étiquetage des produits chimiques




Depuis l'arrêté du 20 avril 1994 modifié, les symboles et indications de danger utilisés pour l'étiquetage des substances et préparations dangereuses sont définis par la réglementation française. Ces symboles, noirs sur un fond carré jaune-orangé, vont être progressivement remplacés par un nouveau système.


Le règlement (CE), dit "règlement CLP [I.12] définit de nouvelles règles européennes de classification, d'étiquetage et d'emballage des produits chimiques. Entré en vigueur le 20 janvier 2009, ce règlement prescrit de nouveaux pictogrammes de danger.

Les anciens et nouveaux systèmes coexisteront durant une période transitoire. La mise en application du nouveau règlement est devenue obligatoire à partir du 1er décembre 2010 pour les substances et du 1er juin 2015 pour les mélanges. Ces pictogrammes ont la forme carrée debout sur la pointe et comportent un symbole en noir sur fond blanc dans un cadre rouge suffisamment épais pour être clairement visible. Ils possèdent chacun un code différent composé de SGH suivi d'un numéro (01, 02, ...,09). On lira, sous le pictogramme, l'un ou l'autre de ces mots : « DANGER » ou « ATTENTION » (voir le tableau I.2).

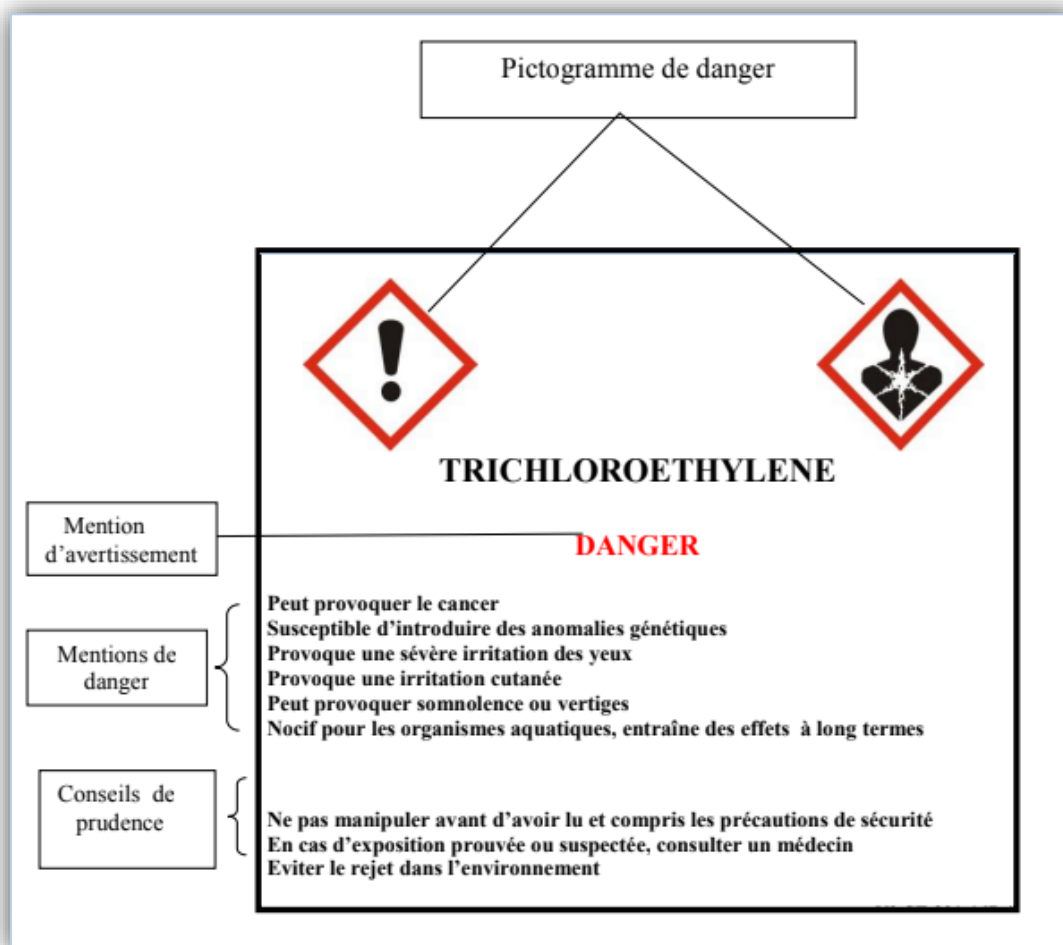
Tableau I.2: Les nouveaux pictogrammes selon le règlement CLP et leur signification [I.12].

Symbole	Signification
 <p>SGH01</p>	<p>Ces produits peuvent exploser au contact d'une flamme, d'une étincelle, d'électricité statique, sous l'effet de la chaleur, d'un choc, de frottements...</p>
 <p>SGH02</p>	<p>Ces produits peuvent s'enflammer, suivant le cas:</p> <ul style="list-style-type: none"> <li>❖ au contact d'une flamme, d'une étincelle, d'électricité statique... ;</li> <li>❖ sous l'effet de la chaleur, de frottements... ;</li> <li>❖ au contact de l'air ;</li> <li>❖ au contact de l'eau, s'ils dégagent des gaz inflammables (certains gaz s'enflamment spontanément, d'autres au contact d'une source d'énergie, flamme, étincelle...).</li> </ul>
 <p>SGH03</p>	<p>Ces produits peuvent provoquer ou aggraver un incendie, ou même provoquer une explosion s'ils sont en présence de produits inflammables. On les appelle des produits comburants.</p>
 <p>SGH04</p>	<p>Ces produits sont des gaz sous pression contenus dans un récipient. Certains peuvent exploser sous l'effet de la chaleur : il s'agit des gaz comprimés, des gaz liquéfiés et des gaz dissous. Les gaz liquéfiés réfrigérés peuvent, quant à eux, être responsables de brûlures ou de blessures liées au froid appelées brûlures et blessures cryogéniques.</p>
 <p>SGH05</p>	<p>Ces produits sont corrosifs, suivant les cas :</p> <ul style="list-style-type: none"> <li>❖ ils attaquent ou détruisent les métaux</li> <li>❖ ils peuvent ronger la peau et/ou les yeux en cas de contact ou de projection.</li> </ul>

 <p>SGH06</p>	<p>Ces produits rentrent dans une ou plusieurs de ces catégories :</p> <ul style="list-style-type: none"> <li>❖ produits cancérogènes : ils peuvent provoquer le cancer ;</li> <li>❖ produits mutagènes : ils peuvent modifier l'ADN des cellules et peuvent alors entraîner des dommages sur la personne exposée ou sur sa descendance (enfants, petits-enfants...) ;</li> <li>❖ produits toxiques pour la reproduction: ils peuvent avoir des effets néfastes sur la fonction sexuelle, diminuer la fertilité ou provoquer la mort du fœtus ou des malformations chez l'enfant à naître ;</li> <li>❖ produits qui peuvent modifier le fonctionnement de certains organes comme le foie, le système nerveux... Selon les produits, ces effets toxiques apparaissent si l'on a été exposé une seule fois ou bien à plusieurs reprises ;</li> <li>❖ produits qui peuvent entraîner de graves effets sur les poumons et qui peuvent être mortels s'ils pénètrent dans les voies respiratoires (après être passés par la bouche ou le nez ou bien lorsqu'on les vomit) ;</li> <li>❖ produits qui peuvent provoquer des allergies respiratoires (asthme, par exemple).</li> </ul>
 <p>SGH07</p>	<p>Ces produits empoisonnent rapidement, même à faible dose. Ils peuvent provoquer des effets très variés sur l'organisme : nausées, vomissements, maux de tête, perte de connaissance ou d'autres troubles plus importants entraînant la mort.</p>
 <p>SGH08</p>	<p>Ces produits chimiques ont un ou plusieurs des effets suivants :</p> <ul style="list-style-type: none"> <li>❖ ils empoisonnent à forte dose ;</li> <li>❖ ils sont irritants pour les yeux, la gorge, le nez ou la peau ;</li> <li>❖ ils peuvent provoquer des allergies cutanées (eczémas) ;</li> </ul> <p>ils peuvent provoquer une somnolence ou des vertiges.</p>

 <p>SGH09</p>	<p>Ces produits provoquent des effets néfastes sur les organismes du milieu aquatique (poissons, crustacés, algues, autres plantes aquatiques...).</p>
--	--

Un exemple d'étiquette représenté par la figure I.2, donne plus de détails et d'informations sur le pictogramme présenté



**Figure I.2:** Signification d'une étiquette.

#### I.5.2.4. Résumé et comparaison des méthodes d'évaluation

Comme illustré par le tableau I.3 (page 23), la classification d'un liquide inflammable selon l'arrêté du 20 avril 1994 modifié, est basée sur les résultats d'essai alors que celle du règlement CLP est basée sur les résultats d'essai ou sur l'application d'une méthode de calcul de type QSAR/QSPR ( Quantitative

Structure Activity / Property Relationship ) permettant de prédire de manière qualitative les propriétés physico-chimiques, biologiques et environnementales à partir de la connaissance de leur structure chimique.

Les symboles et indications de danger, pictogrammes, phrases de risques, catégories de danger, mentions d'avertissement et mentions de danger sont aussi illustrés. Par exemple, pour une substance dont la phrase de risque R10 avec  $23^{\circ}\text{C} \leq \text{PE} \leq 60^{\circ}\text{C}$  se retrouve en catégorie 3 du règlement CLP (voir tableau I.3).

Les éléments d'étiquetage des trois catégories de liquides inflammables sont menés par des phrases de risque et des conseils de sécurité dont les significations sont comme suit [I.12] :

R10 : Inflammable

R11 : Facilement inflammable

R12 : Extrêmement inflammable

H224 : Liquide et vapeurs extrêmement inflammables

H225 : Liquide et vapeurs très inflammables

H226 : Liquide et vapeurs inflammables

P210 : Tenir à l'écart de la chaleur/es étincelles/des flammes nues/des surfaces chaudes/Ne pas fumer

P233 : Maintenir le récipient fermé de manière étanche

P240 : Mise à la terre/liaison équipotentielle du récipient et du matériel de réception

P241 : Utiliser du matériel électrique/de ventilation/ d'éclairage /.../ antidéflagrant

P242 : Ne pas utiliser d'outils produisant des étincelles

P243 : Prendre des mesures de précaution contre les décharges électrostatiques

P280 : Porter des gants de protection/des vêtements de protection/un équipement de protection des yeux/du visage

Conseil de prudence et d'intervention :

P303 + P361 + P353 : En cas de contact avec la peau (ou les cheveux) :enlever immédiatement les vêtements contaminés. Rincer à l'eau/se doucher.

P370 + P378 : En cas d'incendie: utiliser ... pour l'extinction.

Conseil de prudence et de stockage






P403 + P235 : Stocker dans un endroit bien ventilé. Tenir au frais.

Conseil de prudence et d'élimination

P501 : Éliminer le contenu/réceptier dans ...

Des indications sur les dimensions minimales des étiquettes et des pictogrammes sont données par le règlement CLP par rapport à la contenance de l'emballage.

**Tableau I.3:** Résumé et comparaison des méthodes d'évaluation [I.12].

	Arrêté du 20 avril 1994 modifié			Règlement CLP		
Symbole et indication de danger, pictogramme, phrases de risque, catégorie de danger, mentions d'avertissement et mentions de danger						
	F+ - Extrêmement Inflammable  R12	F - Extrêmement Inflammable  R11	R10	Catégorie 1 : Danger H224	Catégorie 1 : Danger H 225	Catégorie 1 : Attention H 226
Critères et Méthodes	Règlement (CE) n°440/2008 A.2 et A.9  PE < 0°C Teb ≤ 35°C	Règlement (CE) n°440/2008 A.2 et A.9  PE ≤ 21°C	Règlement (CE) n°440/2008 A.9  21°C ≤ PE ≤ 55°C	Règlement CLP (idem TMD)  PE < 23°C Teb ≤ 35°C	Règlement CLP (idem TMD)  PE < 23°C Teb > 35°C	Règlement CLP (idem TMD)  23°C ≤ PE ≤ 60°C
	Classification sur la base des résultats d'essais			Classification sur la base des résultats d'essais ou de l'application d'une méthode de calcul.		
	Méthodes d'essai identiques					

## **I.6. Principales catégories de solvants**

Pour les besoins de l'hygiène du travail et pour la substitution des solvants en particulier, il est utile de classer les solvants organiques en quatre grandes familles : les hydrocarbures, les solvants halogénés, les hydrocarbures oxygénés et les autres solvants.

### **I.6.1. Hydrocarbures**

Les hydrocarbures sont constitués uniquement de carbone et d'hydrogène. La plupart des hydrocarbures sur le marché sont issus de la pétrochimie. Cette famille comprend les Hydrocarbures aliphatiques, Hydrocarbures cycliques, Hydrocarbures aromatiques, et Mélanges complexes.

Quelques hydrocarbures appelés terpènes sont extraits de végétaux, par exemple le dlimonène provenant des pelures d'agrumes.

### **I.6.2. Solvants halogénés**

Les solvants halogénés sont des hydrocarbures où l'on a remplacé un ou plusieurs atomes d'hydrogène par des atomes d'halogènes (Brome, Chlore, Fluor, Iode). Cette structure leur confère des propriétés sécuritaires intéressantes telles que l'inflammabilité et même l'incombustibilité, en plus d'un pouvoir de dissociation incomparable, d'où leur usage répandu en milieu de travail.

### **I.6.3. Solvants oxygénés**

Les solvants oxygénés contiennent des atomes d'oxygènes dans leur structure moléculaire en plus d'atomes de carbone et d'hydrogène. Cette caractéristique les rend amphiphiles. Cette famille comporte les alcools, les cétones, les esters, les éthers et les éthers de glycol.

### **I.6.4. D'autres solvants**

Il existe de nombreux autres solvants qui n'entrent pas dans les catégories listées précédemment. Plusieurs sont utilisés comme solvants réactionnels en synthèse chimique. Ils sont aussi appelés les solvants particuliers tels que les hydrocarbures nitrés, d'autres composés azotés, les dérivés soufrés et les hydrocarbures complexes. Certains solvants introduits récemment ont fait l'apparition sur le marché, mais leur utilisation est limitée notamment par leur coût élevé [I.13-Ed 4229].

## **I.7. La toxicité et les maladies professionnelles des solvants organiques**

Aucun solvant n'est inoffensif et certains solvants très toxiques doivent être évités. Beaucoup de solvants organiques ont été reconnus comme susceptibles de provoquer des maladies professionnelles lors de l'exposition des travailleurs aux solvants, dans le cadre de leur activité professionnelle.

Le pouvoir toxique des solvants se manifeste en général par une atteinte neurologique centrale et parfois périphérique, des atteintes muco-cutanées, des atteintes de la reproduction et des glandes endocrines, des troubles digestifs, hépatiques, rénaux, etc...

Certains peuvent provoquer des transformations du patrimoine génétique (effets mutagène et cancérigène) ainsi que des atteintes du système reproducteur (perturbation de la reproduction et atteinte de la descendance). Les femmes enceintes sont particulièrement vulnérables car la plupart de ces solvants franchissent la barrière placentaire. Aujourd'hui, ces dernières molécules, dites CMR (Cancérigènes, Mutagènes et Reprotoxiques) sont unanimement reconnues comme étant responsables de problèmes graves de santé. Elles doivent être éliminées de façon prioritaire pour être remplacées par des solvants de toxicité plus faible [I.14].

**Tableau I.4:** Toxicité de quelques solvants organiques appartenant à différentes familles [I.14].

Famille	Solvant pris pour exemple	Toxicité
Hydrocarbures saturés	Hexane (Inflammable)	Très toxique, puissantes propriétés neurotoxiques périphériques (polynévrite) et certainement centrales (maladie de Parkinson)
Hydrocarbures aromatiques	Benzène (Inflammable)	Extrêmement toxique : cancérigène pour l'Homme. Neurotoxique central Effets cumulatifs
Hydrocarbures halogénés	1,2 -Dichloroéthane (Inflammable)	Très toxique, reconnu cancérigène chez les animaux (rongeurs), hépatotoxique et dangereux pour les reins, le système nerveux et le cœur. Irritant de la peau, des yeux et du système respiratoire. Groupe 2B : cancérigène possible chez l'Homme, doit être éliminé.
	Méthanol	Très neurotoxique pour le nerf optique et la rétine (cécité). Provoque à forte dose de

Alcools	(Inflammable)	l'acidose (confusion mentale jusqu'au coma).
Cétones	Propanone (Très Inflammable)	Neurotoxique modéré (narcotique). Irritant des yeux, de la peau et du tractus respiratoire.
Esters	Acétate de méthyle (Facilement inflammable)	Libère par hydrolyse du méthanol et de l'acide acétique. Irritant des yeux, des muqueuses, de la peau. Atteinte à long terme du nerf optique.

### I.8. Risque pour l'environnement

Les solvants pétroliers sont tous des composés organiques volatils. Leur émission dans l'atmosphère contribue à la production d'ozone dans la troposphère par réaction photochimique, augmentant ainsi les risques pour les personnes asthmatiques ou souffrant d'insuffisance respiratoire.

En cas de rejet dans un milieu aquatique les solvants pétroliers surnageront à la surface. Il peut être envisagé de stopper leur progression par des barrages flottants et, éventuellement, de récupérer cette pollution au moyen d'absorbants par exemple.

La biodégradation est faible, variable selon leur nature: les solvants pétroliers à forte teneur en hydrocarbures aromatiques sont plus toxiques pour les organismes aquatiques.

Le risque d'incendie et d'explosion est l'un des risques majeurs lors de l'utilisation des solvants pétroliers dû à leur inflammabilité. Cependant l'utilisation à chaud de ceux qui sont non inflammables retrouvent les mêmes caractéristiques d'inflammabilité que les autres solvants inflammables car les vapeurs émises par ces substances peuvent former avec l'air des mélanges explosifs [I.13-Ed4224].

### I.9. Quelques accidents

- Le « Soummam 937 » est un navire-école des Forces navales algériennes, acquis en 2006, et dont il est la plus grosse unité en dotation. Il est conçu spécifiquement pour l'entraînement. Le 10 mai 2017 [I.15], un incendie suivi d'une explosion à l'intérieur du navire au niveau du Dock flottant de Béjaïa. L'explosion est survenue au moment de peindre, causant deux morts et quatre blessés des employés de l'Entreprise Nationale d'Entretien et de Réparation Navale (ERENAV).

- La catastrophe de GL1K de SKIKDA, le 21/01/2004 [I.16], qui s'est traduite par la destruction presque totale du complexe et la mort de 27 agents ainsi qu'un nombre important de blessés, de traumatisés avec des répercussions jusqu'à la ville de SKIKDA à environ 3 km.
- Explosion d'un bac d'hydrocarbure, s'est produite le 20 Février 2001 en France [I.17]. Le bac concerné par l'accident avait une capacité de 5090 m<sup>3</sup> . Ce jour là, le bac était vide mais contenait habituellement de l'essence de type. supercarburant. Les travaux, réalisés par des employés d'une entreprise extérieure, consistaient en un raclage du sol pour en retirer des dépôts résiduels Les deux intervenants ont été brûlés dont un grièvement. L'hypothèse d'une étincelle provoquée par un équipement porté par l'un des intervenants conduisant à l'explosion de l'atmosphère explosive.

# **Chapitre II :**

## **Étude théorique**

## II.1. La modélisation moléculaire

La modélisation moléculaire peut être considérée comme un ensemble de techniques informatiques basées sur des méthodes de chimie théorique et les données expérimentales qui peuvent être utilisées pour la description et la prédiction des propriétés des systèmes moléculaires complexes dans différents environnements.

Cette approche procède de l'hypothèse d'une correspondance univoque entre n'importe quelle propriété physique, affinité chimique, ou activité biologique d'un composé chimique et sa structure moléculaire.

La stabilité de la structure tri-dimensionnelle d'une molécule est déterminée par les interactions intramoléculaires et les interactions avec le milieu extérieur (solvant). La recherche des conformations stables d'une molécule consiste à déterminer les minima de l'énergie globale d'interaction. Cette énergie peut être calculée par des méthodes quantiques *ab initio* ou semi-empiriques généralement longues et onéreuses. Pour faciliter les calculs, on considère habituellement que le terme variable de cette énergie dépend de la construction de la molécule et de l'arrangement de ses atomes : c'est le principe des méthodes empiriques (mécanique moléculaire, dynamique moléculaire). Dans la plupart de ces méthodes, il n'est pas tenu compte des interactions avec le solvant, mais uniquement des interactions entre les atomes constitutifs de la molécule. La recherche d'une conformation consiste alors à faire une minimisation de l'énergie intramoléculaire. Cette énergie potentielle est fractionnée en un certain nombre de termes additifs indépendants. Chacun de ces termes est représenté par une fonction analytique simple justifiée par des calculs quantiques et incluant des paramètres empiriques.

## II.2. Optimisation des molécules

### II.2.1. La Méthode de HARTREE-FOCK-ROOTHAAN (Méthode de HFR)

#### II.2.1.1. Energie d'un micro système représenté par un déterminant de Slater

Les calculs quanto-mécaniques courants sont basés sur le modèle de l'électron indépendant où l'on suppose les orbitales soit vides soit garnies de deux électrons au plus.

Dans le cadre de ce modèle, la fonction d'onde polyélectronique peut s'écrire sous la forme d'un produit anti-symétrisé de spin-orbitales :

$$\Psi(1, 2, \dots, n) = \frac{1}{\sqrt{n!}} \begin{vmatrix} \Psi_1(1) & \bar{\Psi}_1(1) & \dots & \dots & \dots & \dots & \bar{\Psi}_n(1) \\ \Psi_1(2) & \bar{\Psi}_1(2) & \dots & \dots & \dots & \dots & \bar{\Psi}_n(2) \\ \vdots & \vdots & & & & & \vdots \\ \vdots & \vdots & & & & & \vdots \\ \Psi_n(n) & \bar{\Psi}_1(n) & \dots & \dots & \dots & \dots & \bar{\Psi}_n(n) \end{vmatrix} \quad (\text{II.1})$$

Les spin-orbitales sont obtenues en multipliant chaque orbitale par l'une des deux fonctions de spin possibles :

$$\Psi_m(\mathbf{n}) = \varphi_m(\mathbf{n})\alpha(\mathbf{n}) \quad (\text{II.2})$$

$$\bar{\Psi}_m(\mathbf{n}) = \varphi_m(\mathbf{n})\beta(\mathbf{n})$$

Nous considérerons le cas des systèmes à couches complètes (gaz inertes, molécules courantes dans l'état fondamental...) pour lesquels  $n=2m$ .

La fonction déterminantale  $\Psi(1, 2, 3, \dots, n)$  est appelée déterminant de Slater.

L'hamiltonien du système est l'hamiltonien résultant, à l'approximation de Born-Oppenheimer.

$$H(1, 2, \dots, n) = \sum_{i=1}^n h_{(i)}^c + \sum_{i < j} \frac{e^2}{r_{ij}} \quad (\text{II.3})$$

$h_{(i)}^c$  : est l'hamiltonien monoélectronique de cœur ; le symbole  $\sum_{i < j}$  désigne une sommation sur couples ordonnés.

Comme  $\Psi$  est normé à l'unité (constante de normalisation  $1/\sqrt{n!}$ ), l'énergie du système est donnée par :

$$E = \langle \Psi | H | \Psi \rangle \quad (\text{II.4})$$

Lorsqu'on développe cette intégrale on arrive [II.1] au résultat :

$$E = \sum_{i=1}^m 2h_{ii}^c + \sum_{i=1}^m \sum_{j=1}^m (2J_{ij} - K_{ij}) \quad (\text{II.5})$$

L'écriture  $\sum_{i=1}^m$ , signifie que l'on somme sur toutes les orbitales occupées.

$$h_{ii}^c = \langle \Psi_i(\mu) | h_{(\mu)}^c | \Psi_i(\mu) \rangle \quad (\text{II.6})$$

est l'intégrale mono électronique moléculaire de cœur, intégrale triple qui porte sur les coordonnées d'un seul électron : le  $\mu$  ème dans ce cas.

$$J_{ij} = \iint \Psi_i^*(\mu) \Psi_i(\mu) \frac{e^2}{r_{\mu\nu}} \Psi_j^*(\nu) \Psi_j(\nu) d\tau_\mu d\tau_\nu \quad (\text{II.7})$$

est l'intégrale mono électronique moléculaire coulombienne, parce qu'elle représente une somme de termes d'interactions coulombiennes, intégrale sextuple qui porte sur les coordonnées de deux électrons.

$$J_{ij} = \iint \Psi_i^*(\mu) \Psi_i^*(\nu) \frac{e^2}{r_{\mu\nu}} \Psi_j(\mu) \Psi_j(\nu) d\tau_\mu d\tau_\nu \quad (\text{II.8})$$

est l'intégrale biélectronique moléculaire d'échange, elle représente également une somme de répulsions entre charges élémentaires, l'électron occupant deux orbitales moléculaires  $\Psi_i$  et  $\Psi_j$ .

$r_{\mu\nu}$  représente la distance entre les deux électrons  $\mu$  et  $\nu$

### Remarques :

1)- Dans l'expression de l'énergie E, nous trouvons deux termes :

\*  $-E^c$ , qui est l'énergie de l'ensemble des électrons évoluant dans le champ des noyaux sans interactions les uns avec les autres.

\*  $-E^{RE}$ , qui est l'énergie de répulsion électronique.

$$E = E^c + E^{RE} \quad (\text{II.9})$$

Evidemment si l'on suppose qu'il n'existe pas d'interactions entre électrons, le second terme disparaît complètement.

2)- Si on a à traiter une molécule, il faut ajouter un terme supplémentaire de répulsion nucléaire.

$$E_T = E + \sum_{N < L} \frac{Z_K Z_L e^2}{R_{KL}} \quad (\text{II.10})$$

$Z_K$  et  $Z_L$  sont les charges des noyaux K et L et  $R_{KL}$  la distance entre ces noyaux. La relation (II.5) est équivalente à :

$$E = \sum_{i=1}^m \{ h_{jj}^c + (h_{jj}^c + \sum_{j=1}^m (2J_{ij} + K_{ij})) \} \quad (\text{II.11})$$

Le terme :

$$e_i = h_{ii}^c + \sum_{j=1}^m (2J_{ij} + K_{ij}) \quad (\text{II.12})$$

correspond à ce qu'on appelle l'énergie des orbitales moléculaires.

E se réduit donc à :

$$E = \sum_{i=1}^m 2e_i \quad (\text{II.14})$$

Dans la méthode de Hatree-Fock-Roothaan ceci n'est plus vrai : l'énergie des micro-systèmes n'étant pas égale à la somme des énergies des orbitales moléculaires.

Pour qu'il en soit ainsi, il faudrait que  $h_{ii} = e_i$  ce qui n'est pas vrai.

Les orbitales moléculaires ne sont pas connues. Le déterminant de Slater n'est connu que par rapport à un jeu de  $\{\Psi_i\}$  dont on ne sait rien, à part qu'elles sont orthogonales.

Le problème est de déterminer le jeu d'orbitales qui permet de construire le système de Slater.

### II.2.1.2. Détermination des Orbitales ou équations de Hartree-Fock

On construit le système de Slater à partir d'un jeu de  $\{\Psi_i\}$ .

Quelles propriétés doivent posséder les  $\Psi_i$  pour être acceptables au sens de la mécanique ondulatoire, et qu'elles puissent s'adapter au système particulier envisagé ?

Il faut que le déterminant de Slater soit une solution approchée de l'équation de

Schrödinger totale :

$$\mathbf{H}(\mathbf{1}, \mathbf{2}, \dots, \mathbf{n})\Psi(\mathbf{1}, \mathbf{2}, \dots, \mathbf{n}) = E\Psi(\mathbf{1}, \mathbf{2}, \dots, \mathbf{n}) \quad (\text{II.15})$$

La propriété la plus fondamentale des solutions de l'équation de Schrödinger est leur stabilité : c'est-à-dire que si on fait subir à la fonction d'onde déterminantale une perturbation du premier ordre, il s'ensuit une perturbation du premier ordre de l'énergie nulle. Il faut donc réaliser absolument cette condition

. Comme la variation du déterminant de Slater s'exprime par la variation du jeu des  $\{\Psi_i\}$ ,

il faudrait avoir, pour une variation première du jeu d'orbitales choisies, une variation première de l'énergie totale nulle, et pour cela il faut que les  $\Psi_i$  soient solutions des équations de Hartree-Fock [II.2-II.4]:

$$\{\delta\Psi_i\} \rightarrow \delta E^1 = 0 \quad (\text{II.16})$$

Ces deux conditions contiennent les équations de Hartree-Fock :

$$F_{(\mu)}\Psi_i(\mu) = e_i\Psi_i(\mu) \quad (\text{II.17})$$

L'équation de Hartree-Fock est une équation intégral-différentielle qui, contrairement à une équation de Schrödinger mono-électronique, fait intervenir un opérateur F qui dépend des fonctions inconnues  $\Psi_i$ .

Opérateur de Hartree-Fock :

$$F_{(\mu)} = [h_{(\mu)}^c + \sum_{i=1}^m 2J_i(\mu) - K_i(\mu)] \quad (\text{II.18})$$

$J_l$  et  $K_l$  sont, respectivement, les opérateurs coulombien et d'échange relatifs à chaque orbitale doublement occupée  $\Psi_i$ .

### II.2.1.3. Equations de Roothaan et Hall

Découlent de la méthode de Hartree-Fock lorsqu'on introduit la condition CLOA (Combinaison Linéaire d' Orbitales Atomiques).

Chaque orbitale moléculaire  $\psi_i$  se présentera sous la forme :

$$\Psi_i(\mu) = \sum_{p=1}^N C_{pi} \varphi_p(\mu) \quad (\text{II.19})$$

L'ensemble des orbitales atomiques  $\{\varphi_p\}$  étant supposé connu, la détermination des  $\Psi_i$  se ramène à la détermination des  $C_{pi}$ .

Les équations de Hartree-Fock prennent, en tenant compte de (II.19), une expression vectorielle assez simple :

$$\sum_{p=1}^N C_{pi} [F_{pq} - e_i S_{pq}] = 0 \quad , q \in [1, N] \quad (\text{II.20})$$

Les coefficients :

$$S_{pq} = \int \varphi_p^* \varphi_q d\tau \quad (\text{II.21})$$

$$F_{pq} = \int \varphi_p^* (F \varphi_q) d\tau$$

sont les intégrales de recouvrement sur la base des fonctions  $\varphi_p$  et les éléments matriciels de l'opérateur de Hartree-Fock  $F$ , et les valeurs propres sont les énergies orbitales  $e_i$

L'équation (II.20) est un système linéaire homogène (N équations à N inconnues) qu'on peut écrire sous la forme matricielle :

$$[F - e_i S] C_i = 0 \quad (\text{II.22})$$

Où  $F$  est la matrice  $[F_{pq}]$  ;  $S$  est la matrice  $[S_{pq}]$  ;  $C_i$  est la matrice  $[C_{pi}]$ .

$$F_{pq} = h_{pq}^c + \sum_{l=1}^N \sum_{m=1}^N p_{lm} \left[ \langle pq | lm \rangle - \frac{1}{2} \langle pm | lp \rangle \right] \quad (\text{II.23})$$

$h_{pq}^c$  = intégrale monoélectronique sur les orbitales atomiques de base.

$$h_{pq}^c = \langle \varphi_p(\mu) | h_{(\mu)}^c | \varphi_q(\mu) \rangle \quad (\text{II.24})$$

$${}^* \langle pm | lq \rangle = - \iint \varphi_p(\mu) \varphi_q(\mu) \frac{e^2}{r_{\mu\nu}} \varphi_l(\nu) \varphi_m(\nu) d\tau_\mu d\tau_\nu \quad (\text{II.25})$$

$${}^* \langle pm | lq \rangle = \int \varphi_p(\mu) \varphi_m(\mu) \frac{e^2}{r_{\mu\nu}} \varphi_l(\nu) \varphi_q(\nu) d\tau_\mu d\tau_\nu \quad (\text{II.26})$$

$${}^* p_{lm} = \sum_{i=1}^N 2C_{li} C_{mi} = \text{éléments de la matrice densité} \quad (\text{II.27})$$

$${}^* p = [p_{lm}] = \text{matrice densité} \quad (\text{II.28})$$

#### II.2.1.4. Quelques remarques sur les processus de résolution des équations de Hartree-Fock-Roothaan

L'équation de Hartree-Fock-Roothaan sous forme matricielle est :

$$[F - e_i S] C_i = 0 \quad (\text{II.29})$$

Löwdin [II.5] a proposé un procédé qui permet de se ramener dans tous les cas au calcul des valeurs propres et vecteurs propres d'une matrice moyennant une transformation de la base des orbitales atomiques (orthogonalisation de Löwdin).

Multiplions à gauche les deux membres de (II.22) par la matrice  $S^{-1/2}$ , qui n'est jamais singulière puisque S ne l'est pas ; il vient successivement:

$$S^{-1/2} F C_i = e_i S^{-1/2} S C_i$$

$$[S^{-1/2} F I S^{-1/2}] S^{-1/2} C_i = e_i S^{-1/2} C_i$$

Soit en posant :

$$[S^{-1/2} F I S^{-1/2}] = \bar{F} \text{ et } S^{-1/2} C_i = \bar{C}_i \quad (\text{II.29})$$

$$\bar{F} \bar{C} = e_i \bar{C}_i \text{ c'est } - \text{ à } - \text{ dire } [\bar{F} - e_i I] \bar{C}_i = 0 \quad (\text{II.30})$$

Les équations de Hartree-Fock-Roothaan sont résolues selon un procédé itératif qui se fait sur l'ensemble orthogonalisé.

$$\bar{F} \bar{C} = e_i \bar{C}_i \quad (\text{II.30})$$

On peut toujours initialiser le problème en choisissant a priori une matrice densité, obtenue en négligeant la matrice des interactions électroniques (problème d'ordre zéro). Le nombre d'itérations dépend du problème à résoudre.

### II.2.1.5. Détermination des intégrales de la méthode de Hartree-Fock-Roothaan (HFR)

Le très gros problème dans la méthode HFR est la détermination des intégrales.

#### ❖ Intégrales monoélectroniques atomiques de cœur :

$$h_{pq}^c = \langle \varphi_p(\mu) | h_{\mu}^c | \varphi_q(\mu) \rangle \quad (\text{II.31})$$

Il existe deux types d'intégrales de ce genre : **monocentres** lorsque  $\varphi_p$  et  $\varphi_q$  appartiennent au même atome R, **bicentres**, lorsque  $\varphi_p$  et  $\varphi_q$  appartiennent à des atomes différents.

Les intégrales monoélectroniques de cœur monocentres comprennent : les intégrales de cœur coulombiennes (même orbitale atomique des deux côtés) et les intégrales de cœur d'échange (les deux orbitales atomiques sont différentes)

$$h_{pq}^c = -\frac{\hbar^2}{2m} \int \varphi_p(\mu) \Delta_{(\mu)} \varphi_q(\mu) d\tau_{\mu} - \sum_K Z_K \int \varphi_p(\mu) \frac{e^2}{r_{K\mu}} \varphi_q(\mu) d\tau_{\mu} \quad (\text{II.32})$$

←  
intégrales cinétique

←  
intégrales d'attractions nucléaires

Les intégrales d'attractions nucléaires peuvent être monocentres, bicentres ou tricentres (très compliquées à calculer).

#### ❖ Intégrales bi-électroniques

$$G_{pq} = \sum_l \sum_m p_{lm} \left[ \langle pq | lm \rangle - \frac{1}{2} \langle pm | lq \rangle \right] \quad (\text{II.33})$$

$\langle pq | lm \rangle$ ,  $\langle pm | lq \rangle$  et  $p_{lm}$  sont respectivement définis par les relations (II.25), (II.26) et (II.27). On a plusieurs types d'intégrales :

- **monocentres**, lorsque,  $\varphi_p, \varphi_q, \varphi_l, \varphi_m$  appartiennent au même atome.
- **bicentres**, lorsque parmi,  $\varphi_p, \varphi_q, \varphi_l, \varphi_m$  il y en a qui appartiennent à deux atomes différents.
- **tricentres**, parmi,  $\varphi_p, \varphi_q, \varphi_l, \varphi_m$  il y en a qui appartiennent à trois atomes différents.
- **tétracentres**, chaque orbitale appartient à un atome différent.

Le calcul des intégrales biélectroniques prend le plus grand temps, et il n'est pas possible, en prenant des orbitales de Slater (II.34) d'en donner des expressions analytiques

$$\varphi_{n,l,m}(k, \vec{r}) = N r^{n-1} e^{-kr} y_{l,m}(\theta, \varphi) \quad (\text{II.34})$$

$y_{l,m}(\theta, \varphi)$  étant les harmoniques sphériques.

On décompose alors chaque orbitale de Slater en orbitales gaussiennes dont la partie radiale est de la forme  $e^{-kr^2}$ , ce qui permet de ramener un problème d'analyse numérique à un problème d'algèbre.

## II.2.2. Les méthodes semi-empiriques

Dans le cadre de cette théorie, une approche plus approximative est développée, ce qui permet d'éviter l'évaluation difficile de beaucoup d'intégrales et de sélectionner les valeurs de certaines autres en tenant compte des données expérimentales.

Les approches semi-empiriques, qui traitent des électrons de valence, sont désignées par des sigles dont les lettres correspondent aux approximations admises dans le recouvrement différentiel des orbitales.

### II.2.2.1. Définition du semi-empirisme

Une méthode est semi-empirique si elle admet le cadre de Hatree-Fock-Roothan, en y incorporant un certain nombre de simplifications.

On arrive ainsi à réduire considérablement le nombre d'intégrales. En particulier on élimine les intégrales biélectroniques à 3 et 4 centres, qui sont très faibles.

Une fois le cadre HFR simplifié, on évalue empiriquement les intégrales restantes en ajustant la méthode sur des molécules bien connues.

### II.2.2.2. Quelques théories semi-empiriques

La première théorie semi-empirique, ou théorie de Pople-Pariser-Parr (PPP), introduite en 1953 par Pariser et Parr [II.6-II.7], et utilisée la même année par Pople [II.8], permet d'étudier les systèmes conjugués sans tenir compte du squelette  $\sigma$ .

La première théorie des orbitales moléculaires semi-empirique tri-dimensionnelle est l'approximation au recouvrement différentiel nul (CNDO pour : Complete Neglect of Differential Overlap), introduite par Pople, Santry et Segal [II.9], pour être appliquée à tous les électrons de valence de molécules quelconques organiques ou minérales.

L'approximation utilisée dans CNDO, et dans de nombreuses approximations subséquentes, pour traiter des interactions électron-électron est connue comme :

- ✓ Approximation du champ moyen ;
- ✓ Théorie du champ auto-cohérent (SCF : Self Consistent Field)
- ✓ Théorie de Hartree- Fock (HF).

De ces appellations, l'approximation du champ moyen est probablement la plus expressive, mais c'est le terme SCF qui est le plus courant.

Comme le problème du calcul de l'énergie d'interaction électron- électron dans un système poly-électronique ne peut avoir de solution exacte, on doit utiliser des approximations. La théorie SCF traite chaque électron comme s'il interagissait (au cours du temps) avec le champ moyen de tous les autres électrons de la molécule. Ce qui signifie que les électrons restants de la molécule ne réagissent pas avec l'électron considéré dans sa position instantanée. Ainsi, le calcul de l'énergie de chaque électron pris individuellement devient un problème mono-électronique auquel nous avons à ajouter l'effet du champ causé par les électrons restants. Cette approximation néglige le fait que les mouvements des électrons sont corrélés de manière à réduire leurs répulsions mutuelles (c'est-à-dire que chaque électron réagit aux positions instantanées de tous les autres). Ainsi, la théorie SCF rend la tâche computationnelle gérable au prix d'une surestimation de l'énergie de répulsion électron- électron.

Cependant, en 1965, les ressources computationnelles nécessaires pour l'approche SCF complète n'étaient pas encore disponibles. La pratique des théories des orbitales moléculaires nécessitaient donc encore des approximations. Le principal problème réside dans le calcul et le stockage des intégrales tétracentres notées  $\langle uv | \lambda\sigma \rangle$ , nécessaires pour le calcul des interactions électron- électron dans le cadre de l'approximation SCF. Les indices  $\mu, \nu, \lambda$  et  $\sigma$  dénotent quatre centres d'orbitales atomiques de sorte que le nombre de telles orbitales à calculer croît proportionnellement à  $N^4$ , où  $N$  est le nombre d'orbitales atomiques. En fait, le nombre de telles intégrales n'est pas exactement égal à la puissance quatrième du nombre de fonctions de base parce que beaucoup d'entre elles sont reliées par symétrie. Ce qui était une tâche très difficile en 1965 ; ainsi Pople, Santry et Segal [II.9] ont introduit l'approximation que seules les intégrales pour lesquelles  $\mu = \nu$  et  $\lambda = \sigma$  c'est-à-dire :  $\langle \mu\mu | \nu\nu \rangle$  seront prises en compte et que, de plus, toutes les orbitales atomiques seront traitées de la même façon (comme si elles étaient des orbitales s), de sorte que l'équation (II.35) s'applique, où  $\mu$  est centrée sur l'atome A et  $\lambda$  sur l'atome B et ainsi  $\gamma_{AB}$  ne dépend que des identités de A et B, et peut être traité comme paramètre.

$$\langle \mu\mu | \nu\nu \rangle = \gamma_{AB} \quad (\text{II.35})$$

Une première approximation, due à Pariser et Parr [II.7] consiste à traiter le terme mono- centre  $\gamma_{AA}$  comme différence entre le potentiel d'ionisation  $PI_A$  et l'affinité électronique  $AE_A$  de A (Equation(II.36)):

$$Y_{AA} = PI_A - AE_A \quad (\text{II.36})$$

Les termes di-centres sont alors donnés par l'équation(II.37):

$$Y_{AA} = \frac{Y_{AA}+Y_{BB}}{2+r_{AB}(Y_{AA}+Y_{BB})} \quad (\text{II.37})$$

Ce qui conduit à  $Y_{AB} = (Y_{AA} + Y_{BB}) / 2$  pour une distance interatomique,  $r_{AB}$ , nulle et

$Y_{AB} \approx 1/r_{AB}$  pour des distances interatomiques plus grandes. Ces expressions (**Equations (II.35)**)

et (II.37)) montrent la simplicité de la technique CNDO, qui a été utilisée pour calculer des propriétés électroniques comme les moments dipolaires ou les énergies d'excitation, généralement à partir des géométries expérimentales. Il y a eu beaucoup de modifications des eqs. ((II.36) et (II.37)), mais elles restent d'une simplicité comparable. Pareillement, des expressions simplifiées ont aussi été utilisées pour les intégrales mono-électroniques.

Cependant, la méthode CNDO montra des insuffisances systématiques directement imputées aux simplifications ébauchées précédemment, aussi fut-elle remplacée par la méthode **INDO (Intermediale Neglect of Differential Overlap)**, introduite en 1967 par Pople, Beveridge et Dobosh [II.10]. L'approximation qui conduit à l'équation (II.35) s'étant avérée très sévère, elle fut remplacée par des valeurs individuelles pour les différents types d'interactions entre deux orbitales atomiques. Ces valeurs individuelles, souvent désignées par  $G_{ss}$ ,  $G_{sp}$ ,  $G_{pp}$  et  $G_{2pp}$  dans la littérature, peuvent être ajustées pour donner un accord avec l'expérience meilleur que celui obtenu avec la méthode CNDO. Cependant, en INDO les termes di-centres sont maintenus du même type que ceux apparaissant dans les équations. (II.36) et (II.37). Cette approximation conduit à des affaiblissements systématiques, comme par exemple dans le traitement des interactions entre doublets isolés.

Pour surmonter ces carences, Pople et collaborateurs revinrent à une approche plus complète que celle qu'ils proposèrent initialement en 1965[II.9]: l'approximation aurecouvrement différentiel diatomique nul (NDDO : Neglect of Diatomic Differential Overlap).

Dans la NDDO, toutes les intégrales tétracentres  $\langle \mu\nu | \lambda\sigma \rangle$  dans lesquelles  $\mu$  et  $\nu$  sont sur le même centre, comme le sont  $\lambda$  et  $\sigma$  (mais pas nécessairement sur le même comme le sont  $\mu$  et  $\nu$ ) sont prises en compte. De plus, les intégrales pour lesquelles les deux centres atomiques sont différents sont traitées de manière analogue que les intégrales mono-centres en INDO, entraînant, une amélioration de la description des interactions (doublet isolé)-(doublet isolé) par rapport aux méthodes précédentes. La NDDO forme la base de presque toutes les autres méthodes semi-empiriques qui, à quelques exceptions ont été développées par MJS Dewar et son école.

Les premières techniques semi-empiriques développées par Dewar et son groupe ont été désignées par MINDO/1-3 et ont été basées sur INDO. Beaucoup d'approximations d'intégrales de l'INDO originale ont été remplacées et les méthodes paramétrées pour reproduire un large intervalle de données expérimentales, particulièrement les énergies et les géométries.

Les méthodes MINDO sont maintenant largement obsolètes.

La méthode avantageuse pour la plupart des techniques modernes d'orbitales moléculaires semi-empiriques est la MNDO, qui a été publiée par Dewar et Thiel en 1977 [II.11-II.12]. La MNDO est une méthode NDDO dans laquelle Dewar et Thiel ont introduit un formalisme basé sur les multipôles pour le calcul des intégrales bi-électroniques. Elle a été paramétrée pour reproduire les chaleurs de formation expérimentales, les géométries, les moments dipolaires et les potentiels d'ionisation. Elle s'avéra très supérieure aux méthodes MINDO pour la plupart des grandeurs calculées. Cependant la MNDO présente une faiblesse qui limite sévèrement son utilité ; elle ne reproduit pas la liaison hydrogène. Cette faiblesse a été surmontée de façon pragmatique par Bustein et Isaev [II.13] qui modifièrent simplement le potentiel de répulsion cœur-cœur par addition de fonctions gaussiennes en vue d'obtenir des liaisons hydrogène. Ce « fixe » a été adopté par le groupe Dewar pour leur méthode suivante AM1 [II.13-II.15] qui est par ailleurs identique à la MNDO. AM1, en retour, s'avéra présenter une faiblesse dans le traitement des composés nitrosés et hypervalents. Ces faiblesses ont été abordées par Stewart dans une nouvelle paramétrisation nommée PM3 [II.16-II.17] qui est par ailleurs identiques à AM1. Cependant, MNDO, MNDO/H, AM1 et PM3 sont pour l'essentiel identiques du point de vue quanto-mécanique. Leur différences se limitent à la « correction » classique des potentiels entre atomes et pour laquelle les paramètres sont traités comme variables dans la procédure de paramétrisation.

### II.2.2.3. Limites et avantages des méthodes semi-empiriques [II.18]

La négligence de toutes les intégrales bi-électroniques tri et tétracentres réduit la matrice de Fock d'un ordre formel  $M^4$  à  $M^2$ . Toutefois, le temps requis pour la diagonalisation de la matrice  $\mathbf{F}$  croît comme le cube de la dimension de la matrice. La diagonalisation d'une matrice devient importante lorsque la dimension dépasse  $\sim 10\,000 \times 10\,000$ . De nombreuses itérations sont nécessaires pour la résolution des équations SCF, et habituellement la géométrie est également optimisée, nécessitant de nombreux calculs pour différentes géométries. Ce qui situe la limite actuelle des méthodes semi-empiriques à environ 1000 atomes. Il est à noter que la méthode classique de résolution des équations HF par diagonalisation de la matrice de Fock s'impose rapidement comme l'étape limitante réelle dans les méthodes semi-empiriques. Des développements ultérieurs se sont ainsi focalisés sur la formulation de méthodes alternatives pour l'obtention d'orbitales SCF sans passer par la diagonalisation [II.19-II.20]. De telles méthodes utilisent des

ajustements (combinaisons) linéaires avec le nombre d'atomes, ce qui permet d'effectuer des calculs pour des systèmes comprenant plusieurs milliers d'atomes.

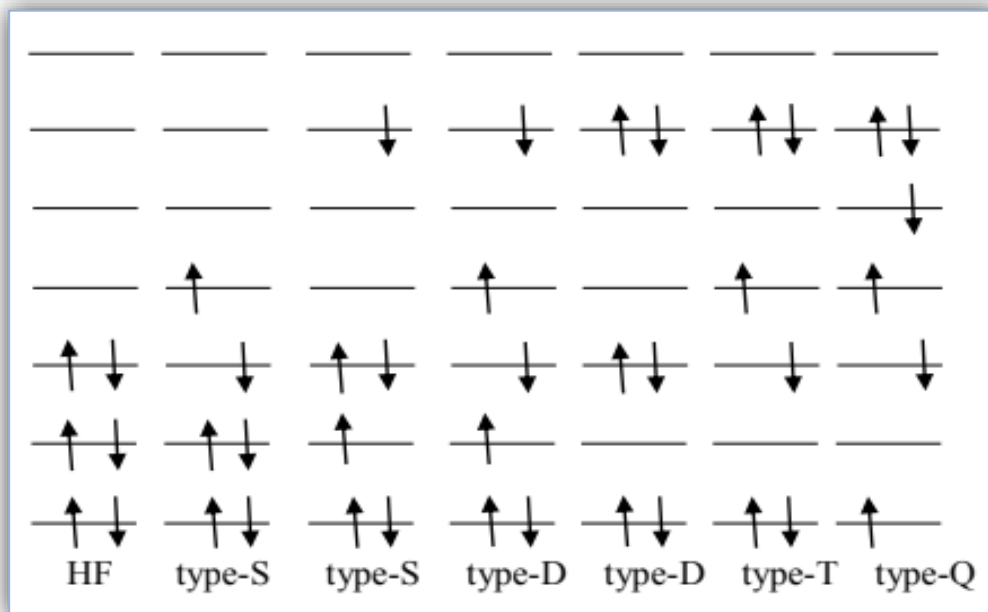
La paramétrisation de **MNDO/AM1/PM3** est réalisée en ajustant les constantes impliquées dans les différentes méthodes de façon à ce que les résultats des calculs HF ajustent les données expérimentales aussi près que possible. Ce qui est faux dans un sens. On sait que la méthode HF ne peut conduire au résultat correct, même à la limite d'un ensemble de base infini et sans approximations. Les résultats HF ne reproduisent pas la corrélation électronique, mais les données expérimentales impliquent naturellement de tels effets. Ceci peut être considéré comme un avantage, les effets de corrélation électronique sont implicitement pris en compte dans la paramétrisation, et il n'est pas besoin d'exécuter des calculs compliqués pour surmonter les déficiences de la procédure HF. Cependant, il y a réellement problème quand la fonction d'onde HF ne peut décrire le système correctement, même qualitativement, comme par exemple avec les bi-radicaux et les états excités.

Une flexibilité additionnelle peut être introduite dans la fonction d'onde d'essai en ajoutant davantage de déterminants de Slater, par exemple par l'intermédiaire d'une procédure d'interaction de configuration (CI : pour Configuration Interaction). Seulement la corrélation électronique est prise en compte deux fois, une première fois lors de la paramétrisation au niveau HF, et une seconde fois explicitement par le calcul CI.

**Remarque :** l'interaction de configuration CI résout le problème de la corrélation électronique en considérant plus d'un schéma d'occupation des orbitales moléculaires (OM) et en combinant les micro-états obtenus par permutation des positions électroniques sur toutes les OM disponibles. Dans sa forme la plus simple, un calcul CI consiste en un calcul SCF préliminaire qui fournit les OM qui seront utilisées telles quelles tout au long du reste du traitement. Des micro-états sont alors construits en déplaçant les électrons des orbitales occupées à celles vacantes selon des schémas pré-établis. La matrice CI est alors calculée, ses éléments diagonaux représentent les énergies des micro-états et les éléments non diagonaux leurs interactions. Cette matrice est diagonalisée en vue d'obtenir les énergies des différents états (fondamentaux et excités) de la molécule comme combinaisons linéaires des micro-états. De nouveau les énergies sont fournies par les valeurs propres et les coefficients de la combinaison linéaire par les vecteurs propres. Cette procédure conduit à la stabilisation de l'état fondamental, et fournit également les énergies et les fonctions d'onde des états excités. Le problème est que si l'on doit considérer chacun des arrangements possibles de tous les électrons dans toutes les OM (CI complète), les calculs deviennent pas trop importants même pour des molécules de taille moyenne avec un ensemble de base pas trop important (parce qu'il y a de trop nombreuses orbitales virtuelles).

Aussi, deux types de restrictions sont habituellement utilisées ; seul un nombre limité d'OM autour de l'intervalle des orbitales frontières (HOMO-LUMO) est inclus dans CI, et seuls certains types de réarrangements (excitations) des électrons sont utilisés.

La forme la plus économique est celle pour laquelle seuls les micro-états dans lesquels un électron est promu de l'état fondamental à une orbitale virtuelle (excitations simples) sont utilisés. Ce qu'on désigne, dans une forme abrégée, par CIS. En ajoutant toutes les excitations doubles (pour les quelles deux électrons sont promus) on est conduit à CISD, et ainsi de suite (Figure. II.1).



**Figure II.1:** Déterminants de Slater excités générés à partir d'une référence HF.

Les déterminants sont désignés par simples (S), doubles (D), Triples (T), quadruples (Q) etc...

La fonction d'onde avec interaction de configuration  $\Psi_{cl}$  peut être représentée par l'équation suivante :

$$\Psi_{cl} = a_0 \phi_{SCF} + \sum \text{simples}(s) a_s \phi_s + \sum \text{Doubles}(D) a_D \phi_D + \dots = \sum_{i=0} a_i \phi_i \quad (\text{II.38})$$

La méthode des multiplicateurs indéterminés de Lagrange [II.21] est ensuite appliquée pour minimiser l'énergie :

$$E = (\langle \Psi | \hat{H} | \Psi \rangle / \langle \Psi | \Psi \rangle) \quad (\text{II.39})$$

Les méthodes semi-empiriques partagent les avantages/désavantages des méthodes de champ de force (cf : II.3), elles sont davantage performantes avec les systèmes pour lesquels on dispose de données expérimentales en quantités, mais il leur est impossible de faire des prédictions pour des types de composés totalement inconnus. La dépendance des données expérimentales n'est pas aussi sévère que pour la méthode

du champ de force, à cause de la forme complexe de la fonctionnelle du modèle. Les méthodes NDDO nécessitent uniquement des paramètres atomiques, et nullement des paramètres di-, tri- et tétra-atomiques comme dans les méthodes de champ de force. Une fois un atome donné paramétré, tous les types de composés possibles contenant cet élément peuvent être traités. Le plus petit nombre de paramètres et la forme plus complexe de la fonctionnelle ont l'inconvénient, par rapport aux méthodes de champ de force, qu'il est très difficile de « réparer » un problème spécifique par métrisation.

Les méthodes semi-empiriques sont de dimension nulle, tout comme les méthodes de champ de force. Il n'y a aucun moyen d'évaluer la fiabilité d'un résultat donné dans les limites de la méthode. Cela est dû à la sélection d'un ensemble de base fixe (minimum). La seule façon de juger les résultats est de comparer la précision d'autres calculs sur des systèmes similaires avec des données expérimentales.

Les méthodes semi-empiriques fournissent une méthode de calcul de la fonction d'onde électronique, qui peut être utilisée pour la prévision d'une variété de propriétés. Il n'y a rien qui entrave le calcul, par exemple, de la polarisabilité d'une molécule, bien qu'il soit connu des calculs ab-initio que l'obtention de bons résultats nécessite un grand ensemble de base polarisé incluant des fonctions diffuses. Les méthodes semi-empiriques comme AMou PM3 n'ont qu'une base minimale (absence de polarisation et de fonctions diffuses), la corrélation électronique n'est qu'implicitement incluse par les paramètres et aucune donnée de polarisabilité n'a été utilisée pour dériver ces paramètres. Il est douteux que de tels calculs puissent conduire à des résultats comparables à ceux fournis par l'expérience, et ils nécessitent, pour le moins, un calibrage soigné [II.18]. Encore une fois, il convient de souligner que la capacité d'effectuer un calcul ne garantit pas la fiabilité des résultats obtenus.

### II.2.3. Analyse des distributions de charges

Plutôt que de décrire la distribution électronique d'une molécule par des cartes d'isodensité, on préfère caractériser cette distribution, dans le voisinage d'un atome ou d'une liaison, par des nombres simples ou indices. Cette procédure, qui entraîne une perte d'information, est avantageuse dans les études comparatives.

La caractérisation d'une molécule par un tel ensemble d'indices est appelée son **analyse de population**.

Il existe une famille d'analyses de population, parmi lesquelles nous citerons celles de Coulson et Longuet-Higgins [II.22], exprimée en termes de charges (ou « densités de charge ») et d'ordres de liaison, celle de Mulliken [II.23], que nous rappellerons brièvement, et qui fait intervenir les populations atomiques et de recouvrement.

### II.2.3.1. Analyse de population de Mulliken

Mulliken introduit le concept important de population de recouvrement, c'est-à-dire de population électronique non localisée sur un atome mais répartie dans la liaison entre deux atomes. Ce concept permet une représentation très nuancée de la liaison chimique.

Dans l'analyse de population électronique qu'il propose, Mulliken définit les grandeurs [II.23] :

$$p_v = \sum_k^{OM.occupées} N_k C_{kv}^* C_{kv} \quad (\text{II.40})$$

Où  $N_k$  est la population de l'O.M.  $\Psi_k$  ;  $p_v$  est la population électronique localisée dans l'O.A.  $\varphi_{\mu\nu}$ , que l'on appelle la population nette de l'O.A.  $\varphi_v$ , dans la molécule.

$$R_{\mu\nu} = 2 \sum_k^{OM.occupées} C_{k\mu}^* C_{kv} S_{\mu\nu} \quad (\text{II.41})$$

$R_{\mu\nu}$  est la population électronique localisée ni dans  $\varphi_\mu$ , ni dans  $\varphi_\nu$  mais répartie entre ces deux O.A, que l'on appelle population de recouvrement entre les O.A  $\varphi_\mu$  et  $\varphi_\nu$ .

En désignant par  $N$  le nombre total d'électrons, on a :

$$\sum_\mu R_{\mu\nu} = \sum_\mu \sum_\nu P_{\mu\nu} S_{\mu\nu} = N \quad [\text{Décomposition sur les OA}] \quad (\text{II.42})$$

$$\int \Psi^* \Psi d\tau = N \quad [\text{Décomposition sur les OM}] \quad (\text{II.43})$$

Posons :

$$q_\mu = \sum_\nu P_{\mu\nu} S_{\mu\nu} = \text{Quantité d'électricité qui peut être attribuée à la } \mu^{\text{ème}} \text{ orbitale atomique de base.}$$

Alors, la quantité d'électricité qui peut être attribuée à l'atome  $M$ , dans la molécule, est la somme des  $q_\mu$  ( $M$ )( $\mu \in M$ ), soit :

$$\delta_M = Z_M - Q_M \quad (\text{II.44})$$

$q_\mu$  = densité électronique de l'orbitale  $\mu$  ;

$Q_M$  = densité électronique de l'atome  $M$ .

On peut ainsi déterminer la **charge (formelle) de l'atome  $M$ , dans la molécule, soit  $\delta_M$**  :

$$\delta_M = Z_M - Q_M \quad (\text{II.45})$$

$Z_M$  = nombre d'électrons de l'atome isolé ;

quantité d'électricité qu'il possède dans la molécule.

$Q_M =$

### II.2.3.2. Calcul du moment dipolaire

Le moment dipolaire d'une molécule peut être décomposé, de façon unique, en trois composantes : une composante atomique ou d'hybridation, une composante de recouvrement, et une composante de transfert de charge (qui permet de définir les charges atomiques nettes), chacune étant définie de façon univoque dans le cadre du schéma OM-CLOA.

Dans ce schéma, l'expression en u.a du moment dipolaire d'une molécule, dans la convention des chimistes, est [II.24] :

$$\vec{\mu} = \sum_p \sum_q \sum_{r \in p} \sum_{s \in q} P_{rs}^{PQ} \int \varphi_r^* \vec{r} \varphi_s d_s d_r - \vec{\mu}_{nucl} \quad (\text{II.46})$$

Avec :

$$P_{rs}^{PQ} = \sum_i n_i C_{ir} C_{is} \quad (\text{II.47})$$

$n_i$  = **taux** d'occupation de l'OM  $\Psi_i$ ,  $C_{ir}$  et  $C_{is}$ , coefficients des orbitales  $\varphi_r$  et  $\varphi_s$  appartenant respectivement, aux atomes P et Q, dans l'approximation CLOA des  $\Psi_i$ . Le vecteur position d'un électron en général et le vecteur position d'un atome P (mesurés en u. a par rapport à la même origine arbitraire) seront notés  $\vec{r}$  et  $\vec{r}_p$ , alors que  $n_p$  désignera le nombre d'électrons de l'atome P engagés dans la formation de la molécule.

On peut alors faire les substitutions suivantes :

$$\vec{r} = \vec{r}_p + \vec{\epsilon}, \text{ dans les termes tels que } P = Q \quad (\text{II.48})$$

$$\vec{r} = \frac{1}{2}(\vec{r}_p + \vec{r}_q) + \vec{X}, \text{ dans les termes tels que } P \neq Q$$

Evidemment  $\vec{\xi}$  est le rayon vecteur qui a pour origine la position de l'atome P,  $\vec{X}$  est le rayon vecteur dont l'origine coïncide avec le milieu du segment PQ. En tenant compte de l'orthogonalité des deux orbitales  $\varphi_r$  et  $\varphi_s$  centrées sur le même atome P, en appelant  $S_{rs}^{PQ}$  l'intégrale de recouvrement des orbitales centrées sur des atomes P et Q différents, et en posant :

$$\vec{\xi}_{rr'}^p = \int \varphi_r^* \vec{\xi} \varphi_{r'} d_\tau; \vec{X}_{rs}^{PQ} = \frac{\int \varphi_r^* \vec{X} \varphi_s d_\tau}{S_{rs}^{PQ}} \quad (\text{II.49})$$

Le moment dipolaire (51) devient [II.23]:

$$\vec{\mu} = \sum_p \delta_p \vec{r}_p + \vec{\mu}_{hybrid} + \vec{\mu}_{recouvr} \quad (\text{II.50})$$

Avec :

$$\vec{\mu}_{\text{hybrid}} = \sum_p \sum_{r,r' \in p} P_{rr'}^{PP} \vec{\xi}_{rr'}^p \quad (\text{II.51})$$

ET :

$$\vec{\mu}_{\text{recouvr}} = \sum_p \sum_{r,r' \in p} \sum_Q \sum_{s \in Q} P_{rs}^{PQ} S_{rs}^{PQ} \vec{X}_{rs}^{PQ} \quad (\text{II.52})$$

### II.2.3.3. Application

Nous avons réuni dans la figure II.2 quelques applications [II.25] des indices électroniques de la méthode des orbitales moléculaires.

Sur la base des charges atomiques partielles on peut calculer des descripteurs électrostatiques simples qui peuvent servir pour le développement d'équations QSXR [Relations Quantitatives Structures –X ; où X= P (propriété) – A (activité) – R (rétention chromatographique) – T (toxicité)...].

- Les charges partielles minimale (la plus négative) et maximale (la plus positive) dans la molécule ( $q_{min}$ ,  $q_{max}$ ).
- Les charges partielles minimale et maximale pour les atomes particuliers (C, O etc...).
- Un paramètre de polarité simple ( $q_{max}$ ,  $q_{min}$ ) ou pondéré par une fonction de la distance  $r_{max}$  entre les atomes portant les charges partielles minimale et maximale.

$$P_f = \frac{q_{max} - q_{min}}{F(r_{max})} \quad (\text{II.53})$$

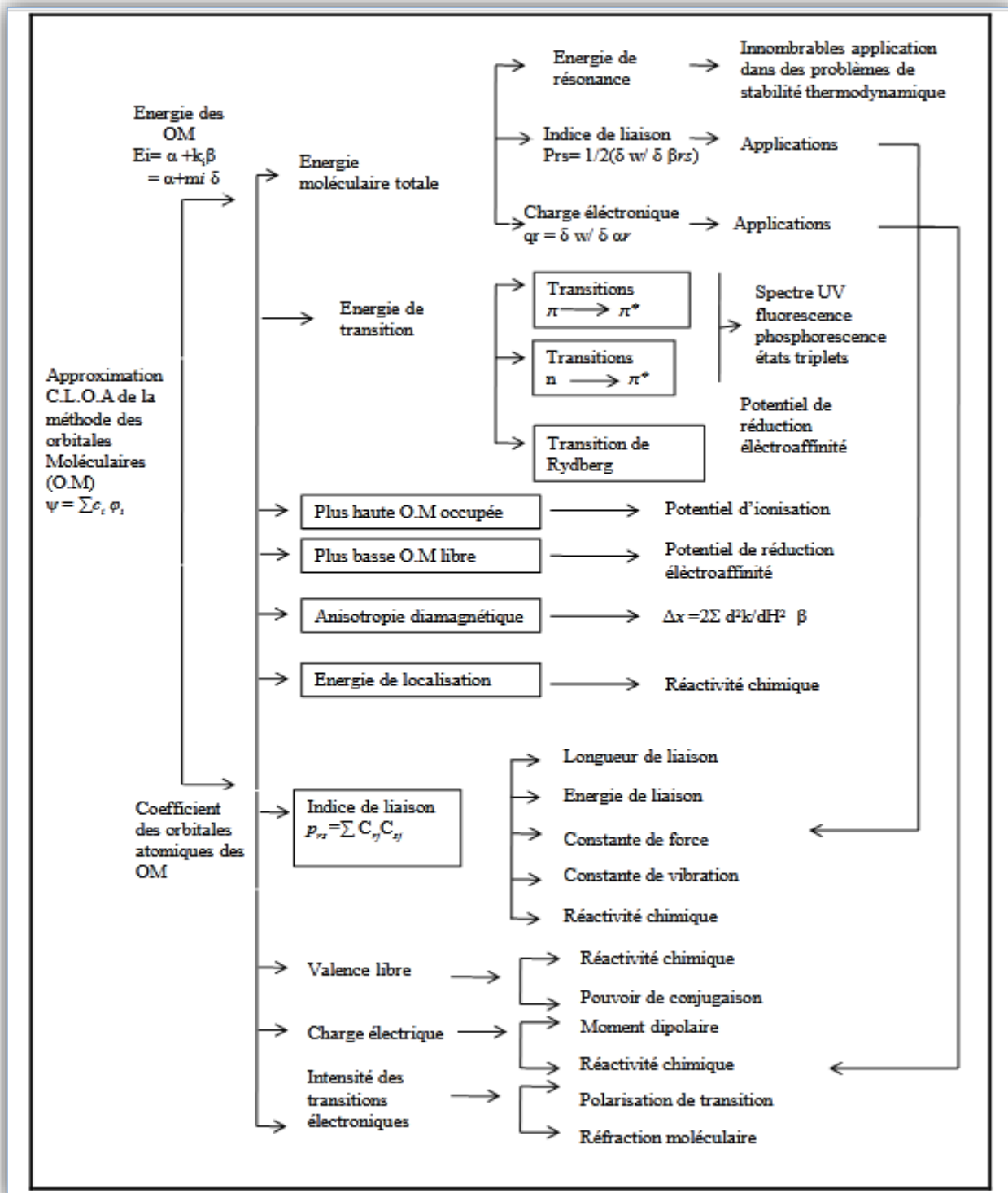


Figure II.2: Les indices électroniques de la méthode des orbitales moléculaires et leurs applications, d'après [II.25].

## II.3. La mécanique moléculaire

Si une molécule est trop grosse pour subir un traitement semi-empirique, il est toujours possible de modéliser son comportement en évitant complètement la mécanique quantique. Les méthodes désignées par mécanique moléculaire, établissent une expression algébrique simple de l'énergie d'un composé, sans avoir à calculer une fonction d'onde ou une densité électronique totale [II.26]. L'expression de l'énergie consiste en des équations classiques simples, comme l'équation de l'oscillateur harmonique, dans le but de décrire l'énergie associée à l'étirement de liaison, de flexion, de rotation, et aux forces intermoléculaires, telles que les interactions de Van Der Waals et de liaison hydrogène. Toutes les constantes apparaissant dans ces équations doivent être obtenues à partir de données expérimentales ou d'un calcul *ab initio*.

Dans une méthode de mécanique moléculaire, la base de données des composés utilisés pour paramétrer la méthode (un ensemble de paramètres et de fonctions est appelé un champ de force) est cruciale pour son succès. La méthode de mécanique moléculaire peut être paramétrée à partir d'une classe spécifique de molécules, telles que des protéines, des molécules organiques, organo-métalliques, etc...

La mécanique moléculaire permet la modélisation de très grosses molécules, comme les protéines et des segments de DNA, la faisant le premier outil de la biochimie computationnelle. Le défaut de cette méthode est qu'il y a beaucoup de propriétés chimiques qui n'y sont pas définies, comme par exemple les états électroniques excités. De plus, pour travailler avec des systèmes très grands et très compliqués, les logiciels doivent être très puissants et faciles dans l'utilisation des interfaces graphiques.

### II.3.1. Pas de calculs de champ de force sans définition préalable des types d'atomes

La géométrie de la molécule traitée (caractérisée par les coordonnées internes ou les coordonnées cartésiennes), le numéro atomique de chaque noyau, et l'état général de charge et de spin, constituent le nombre minimal d'entrées préalable à un calcul par mécanique moléculaire. Les informations concernant les distributions des électrons, en terme de densité électronique ou de fonction d'onde, ou les charges atomiques partielles, sont mieux interprétées sur la base de la géométrie moléculaire. Dans le contexte de la méthodologie du champ de force, l'entrée de la charge totale et du spin d'une molécule n'est pas obligatoire car ces types de calculs ne traitent pas des électrons. Pour représenter l'aspect électrostatique, il n'est même pas besoin des charges atomiques partielles si l'on utilise, par exemple, des dipôles de liaisons. Au contraire de la mécanique quantique, la mécanique moléculaire nécessite plus d'informations que le numéro atomique seul. En fait, chaque atome doit être décrit de manière plus détaillée.

Le concept de types d'atomes permet une différenciation en termes d'environnement local, d'état d'hybridation, ou de conditions spécifiques telles que la tension dans les systèmes comportant un petit anneau. Allinger et ses co-auteurs, qui ont développé les champs de force MM2, MM3, et MM4 pour les «

Petites molécules » (cf : III.3.3) ont défini dans la paramétrisation de MM3 plus de 15 types d'atomes différents pour le seul carbone. A savoir, alcanes  $sp^3$ , alcènes  $sp^2$ , cyclopropanes  $sp^2$ , carbonyles  $sp^2$ , alcynes  $sp$  etc..., tous nécessaires pour rendre MM3 applicable (ce qui signifie l'obtention de résultats raisonnables) pour un ensemble de molécules diverses. On peut constater immédiatement la difficulté de cette approche : le plus d'atomes définis, le plus de paramètres de contribution à la fonction énergie potentielle (liaisons, angles, dièdres...) doivent être développés. Des champs de force plus généraux affecteront donc, un seul type d'atome de carbone générique  $sp^2$ , sacrifiant en faveur d'une application générale. Une autre tendance consiste à utiliser pour les champs de force de classes spécifiques des types d'atomes plus importants en nombres, qu'on ne le ferait dans le cas de paramétrisations pour une application générale.

### II.3.2. Forme fonctionnelle des champs de force courants

Un champ de force ne consiste pas uniquement en une expression mathématique qui décrit l'énergie d'une molécule en fonction des coordonnées atomiques. La deuxième partie indispensable est le jeu de paramètres lui-même. Deux champs de force différents peuvent présenter la même forme fonctionnelle, mais utilisent un paramétrage complètement différent. D'un autre côté, différentes formes fonctionnelles peuvent conduire à des résultats presque identiques, en fonction des paramètres mis en jeu. Cette comparaison montre que les champs de force sont empiriques : il n'y a pas de forme « correcte ».

Parce que certaines formes fonctionnelles donnent de meilleurs résultats que d'autres, la plupart des implémentations dans les logiciels disponibles (académiques et commerciaux) sont très similaires.

Une hypothèse importante est que le champ de force déterminé à partir d'un ensemble de molécules est transférable à d'autres molécules.

Notons qu'un ensemble de paramètres développés et testés sur un nombre relativement petit de cas est encore applicable à une plus large gamme de problèmes. En outre, les paramètres développés à partir de données relatives à de petites molécules peuvent être utilisés pour étudier des molécules plus grandes telles que les polymères.

### II.3.3. Quelques exemples

Parmi les champs disponibles nous citerons les plus répandus largement utilisés pour le traitement de petites molécules.

- **MM2, MM3, et MM4** : (<http://europa.chem. Uga. edu/allinger/mm2mm3.html>). Introduit par Allinger et al. [II.27-II.30], largement utilisé pour le traitement de petites molécules.

- **AMBER** : (Assisted Method Building and Energy Refinement) (<http://amber.scripps.edu>) Introduit par Mac kerell et al., [II.31], très largement utilisé dans le traitement des protéines et des acides nucléiques.
- **CHARMM** : (Chemistry at Harvad Molecular Modeling) (<http://yuri.harvard.edu>) Développé par Mac kerell et al., 1995/1998 ; Brook et al., 1983; [II.31-II.33] qui est largement utilisé pour la simulation de petites molécules et jusqu'aux complexe solvatés de grandes macromolécules biologiques. CHARMM est une version commerciale disponible de CHARMM qui est également applicable aux petits composés organiques [II.34].
  - **MMFF**: (Merck Molecular Force Field) Développé par Halgren [II.35-II.39], il est similaire à MM3 dans la forme, mais en diffère par son application focalisée sur les processus de condensation de phases (en dynamique moléculaire). Il reproduit l'exactitude MM3 pour les petites molécules, et est applicable aux protéines et autres systèmes d'importance biologique.

Les différents champs de force se distinguent par trois aspects principaux :

1. La forme de la fonction de chaque terme énergétique.
2. Le nombre de termes croisés (qui reflètent le couplage entre coordonnées internes) inclus.
3. Le type d'information utilisé pour ajuster les paramètres.

## II.4. Génération des descripteurs moléculaires

La représentation numérique de la structure chimique (descripteurs moléculaires) est une étape importante de l'investigation QSPR. Les performances du modèle élaboré et la précision des résultats sont étroitement liées au mode de détermination de ces descripteurs.

Nous avons utilisé le logiciel de modélisation moléculaire Hyperchem 6.03 [II.40] pour représenter et optimiser les molécules. Tous les calculs ont été menés dans le cadre du formalisme RHF sans interaction de configuration. Les structures moléculaires ont été optimisées à l'aide de l'algorithme Polak- Ribiere avec pour critère une racine du carré moyen du gradient égale à 0,01 kcal/mol. Les géométries finales ont été obtenues à l'aide de la méthode semi-empirique PM3 (Parametric Method 3). Les géométries ainsi optimisées ont été transférées dans le logiciel DRAGON [II.41] pour le calcul de plus de 1600 descripteurs appartenant aux 20 classes différentes suivantes :

1. Descripteurs constitutionnels
2. Descripteurs topologiques
3. Descripteurs Walk and path counts (dénombrements de pas et chemins)
4. Descripteurs de connectivité
5. Descripteurs indices d'information
6. Descripteurs autocorrélation 2D

7. Descripteurs indices d'adjacence
8. Descripteurs valeur propre de Burden
9. Descripteurs indices de charge topologique
10. Descripteurs indices basés sur des valeurs propres
11. Descripteurs profil moléculaire de Randic
12. Descripteurs géométriques
13. Descripteurs RDF
14. Descripteurs Morse - 3D
15. Descripteurs WHIM
16. Descripteurs GETAWAY
17. Descripteurs nombre de groupes fonctionnels
18. Descripteurs fragments des atomes centraux
19. Descripteurs de charge
20. Propriétés moléculaires

En utilisant les options correspondantes du logiciel DRAGON, nous avons d'abord éliminé les descripteurs à valeurs constantes (écarts types inférieurs à 0,001) qui n'apportent aucune information, ensuite ceux qui sont hautement corrélés ( $R \geq 0,95$ ) et qui véhiculent une information redondante. Pour chaque paire de descripteurs corrélés, est éliminé automatiquement celui qui présente les plus hautes corrélations croisées avec les autres descripteurs.

## II.5. Méthodes appliquées pour la sélection d'échantillons

La sélection d'échantillons représentatifs est une étape importante dans une procédure d'élaboration de modèles QSAR/ QSPR. En effet, si les jeux d'étalonnage et de validation ne couvrent pas les mêmes domaines de variation, la validation du modèle ne sera pas correcte. Les échantillons d'étalonnage doivent donc répondre à certains critères ; on identifié 3 règles d'optimalité pour les échantillons de calibrage:

- les échantillons retenus doivent présenter une variabilité maximale;
- la plage de variation des valeurs doit être la plus grande possible, mais limitée aux valeurs rencontrées dans la pratique;
- les échantillons doivent être uniformément répartis.

Différentes méthodes de sélection d'échantillons sont appliquées dans les études QSAR/QSPR, chacune avec ses avantages et ses inconvénients. Parmi ces méthodes de sélection d'échantillons, nous avons appliqué les deux méthodes suivantes :

### II.5.1. Sélection aléatoire des échantillons

L'échantillonnage aléatoire simple (au hasard) est la méthode la plus courante pour le fractionnement des données dans le développement des modèles, où les données sont sélectionnées avec une probabilité uniforme. L'échantillonnage au hasard simple est facile à réaliser et peut être efficacement exécuté dans un seul passage sur les données en utilisant des algorithmes tels que l'algorithme de Knuth [II.42].

Cependant, le problème avec cette approche est qu'il y'a une chance que la sélection de données souffre de la variance, ou de partialité, en particulier lorsque les données ne sont pas réparties uniformément [II.43].

Nous avons appliqué le choix aléatoire pour la sélection des ensembles de calibrage et de test lors de la modélisation des points d'éclair des hydrocarbures non-saturés et de n-alcane.

### II.5.2. Algorithme CADEX pour la sélection d'échantillons

L'algorithme de Kennard et Stone [II.44] est une méthode séquentielle d'échantillonnage qui maximise les distances euclidiennes entre les nouveaux échantillons sélectionnés. Elle commence par situer les deux échantillons les plus éloignés l'un de l'autre, qui sont retirés de la base de données initiale et affectés à l'ensemble de calibrage.

Pour chaque échantillon non sélectionné (éch i), l'algorithme :

- calcule la distance vers chaque échantillon déjà sélectionné ;
- attribue à (éch i) la plus petite des distances.

L'échantillon (éch i) associé à la plus grande distance est donc le plus éloigné de tous les échantillons ; c'est donc lui qui est sélectionné.

La procédure est répétée jusqu'à l'obtention du nombre d'échantillons désirés pour l'ensemble de calibrage. Le fait de sélectionner les échantillons les plus éloignés les uns des autres introduit une grande diversité dans l'ensemble de calibrage ; l'obtention d'une répartition uniforme est un autre avantage de cette technique. L'algorithme de Kennard et Stone est parmi les meilleures méthodes de construction des ensembles de calibrage et de test pour les différentes bases des données [II.44].

Cette méthode a été appliquée pour sélectionner les deux sous ensembles de calibrage et de test lors de la modélisation de la température d'ébullition des différentes classes de solvants.

### II.6. Sélection d'un sous-ensemble de descripteurs significatifs

Dans toutes les méthodes QSPR, les descripteurs (variables explicatives) sont disposés en colonnes et les molécules en lignes, ce qui aboutit à une matrice. L'objectif consiste à trouver une corrélation entre un

ensemble réduit de descripteurs et la propriété physico-chimique ou bien l'activité biologique des molécules. Les algorithmes génétiques constituent une méthode de choix pour la sélection de sous-ensembles de variables explicatives.

### II.6.1. Principe de sélection par algorithme génétique

Dans la terminologie des algorithmes génétiques, le vecteur binaire  $I$ , appelé "chromosome", est un vecteur de dimension  $p$  où chaque position (un "gène") correspond à une valeur (1 si elle figure dans le modèle, 0 sinon). Chaque chromosome représente un modèle basé sur un ensemble de variables explicatives.

On commence par définir le paramètre statistique à optimiser (par exemple maximiser  $Q_2$  en utilisant la validation croisée par "leave-one-out"), avec la taille  $P$  de la population du modèle (par exemple,  $P = 100$ ), et le nombre maximum de variables  $L$  permises pour le modèle (par exemple,  $L = 10$ ) ; le minimum de variables permises est généralement supposé égal à 1. De plus, une probabilité de croisement  $p_c$  (habituellement élevée,  $p_c > 0,9$ ), et une probabilité de mutation  $p_M$  (habituellement faible,  $p_M < 0,1$ ) doivent être également définies.

Après définition des principaux paramètres, la mise en œuvre de l'algorithme génétique est démarrée, son évolution comprend trois étapes principales.

### II.6.2. Initialisation aléatoire du modèle

La population est constituée au départ de modèles aléatoires avec des variables comprises entre 1 et  $L$ , puis les modèles sont ordonnés eu égard au paramètre statistique sélectionné – la qualité du modèle – (le meilleur modèle est en première position, le plus mauvais en position  $P$ ).

### II.6.3. Étape de croisement

A partir de la population, on sélectionne des paires de modèles (aléatoirement, ou avec une probabilité proportionnelle à leur qualité). Puis, pour chaque paire de modèles on conserve les caractéristiques communes, c'est-à-dire les variables exclues dans les 2 modèles restent exclues, et les variables intégrées dans les 2 modèles sont conservées. Pour les variables sélectionnées dans un modèle et éliminées dans l'autre, on en essaye un certain nombre au hasard que l'on compare à la probabilité de croisement  $p_c$  : si le nombre aléatoire est inférieur à la probabilité de croisement, la variable exclue est intégrée au modèle et vice versa. Finalement, le paramètre statistique du nouveau modèle est calculé : si la valeur de ce paramètre est meilleure que la plus mauvaise pour la population, le modèle est intégré à la population, à la place correspondant à son rang ; dans le cas contraire, il n'est pas pris en compte. Cette procédure est répétée pour de nombreuses paires (100 fois par exemple)

### II.6.4. Étape de mutation

Pour chaque modèle de la population (c'est-à-dire pour chaque chromosome)  $p$  nombres aléatoires sont éprouvés, et, un à la fois, chacun est comparé à la probabilité de mutation,  $p_M$ , définie : chaque gène demeure inchangé si le nombre aléatoire correspondant excède la probabilité de mutation, dans le cas contraire, on le change de 0 à 1 ou vice versa.

Les faibles valeurs de  $p_M$  permettent uniquement peu de mutations, conduisant à de nouveaux chromosomes peu différents des chromosomes générateurs.

Après obtention du modèle transformé, on en calcule le paramètre statistique : si cette valeur est meilleure que la plus mauvaise de la population, le modèle est intégré à la population, à la place correspondant à son rang ; dans le cas contraire il n'est pas pris en compte. Cette procédure est répétée pour tous les chromosomes, c'est-à-dire  $P$  fois.

### II.6.5. Conditions d'arrêt

Les étapes précédentes (croisement et mutation) sont répétées jusqu'à la rencontre d'une condition d'arrêt (par exemple un nombre maximum d'itérations défini par l'utilisateur), ou qu'il est mis fin arbitrairement au processus.

Une caractéristique importante de la sélection d'un ensemble réduit de variables par algorithme génétique est qu'on n'obtient pas nécessairement un modèle unique, mais le résultat consiste habituellement en une population de modèles acceptables ; cette caractéristique, parfois considérée comme un désavantage, fournit une opportunité pour procéder à une évaluation des relations avec la réponse à différents points de vue.

La sélection des descripteurs a été réalisée par algorithme génétique, dans la version MOBYDIGS de Todeschini [II.45], en maximisant le coefficient de prédiction  $Q_{LOO}^2$ .

## II.7. Développement des modèles QSPR/QSAR

Les techniques les plus courantes pour établir des modèles QSPR/QSAR utilisent l'analyse de régression (dans les cas où la propriété étudiée est disponible sur une échelle continue), comme la régression linéaire multiple : RLM et la régression par les moindres carrés partiels (MCP ou PLS). Les méthodes de classification, ou en utilisant la régression non-linéaire telle que la régression par les Machines à Vecteurs Supports (SVM) sont également utilisées en général.

### II.7.1. La régression linéaire multiple (RLM)

La régression linéaire multiple (RLM) est largement utilisée dans les méthodes QSPR pour sa simplicité, sa transparence, sa reproductibilité, et sa facilité d'interprétation. Une RLM consiste en une équation (multi) linéaire obtenue par régression des données expérimentales en fonction d'un ensemble de descripteurs pré-sélectionnés, en utilisant la méthode des moindres carrés ordinaires (MCO) [II.46].

$$Y = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + a_3 \cdot X_3 + \dots + a_n \cdot X_n \quad (\text{II.54})$$

Y est la réponse (propriété physique ou activité biologique d'intérêt),  $X_1, X_2, X_3, \dots, X_n$  sont les descripteurs ou variables explicatives présentes dans le modèle,  $a_1, a_2, a_3, \dots, a_n$  sont les coefficients de la régression et  $a_0$  est le terme constant du modèle.

Supposons qu'on ait mesuré sur n individus et p variables représentées par des vecteurs de :  $\mathfrak{R}^n: y, x_1, x_2, \dots, x_k$  Y est la variable dépendante ou à expliquer et les j x les variables

explicatives ou encore prédicteurs (descripteurs moléculaires). On cherche alors à reconstruire y au moyen des  $x_j$  par une formule linéaire.

On pose :

$$Y = \beta_0 \mathbf{1} + X(j)\beta(j) + \varepsilon(j) \quad (\text{II.55})$$

Y est un vecteur de dimension n contenant la propriété étudiée des composés chimiques considérés,  $\mathbf{1}$  est un vecteur unité, c'est-à-dire une matrice colonne formée d'éléments égaux à 1,  $X(j)$  indique la matrice ( $n \times j$ ), et  $\varepsilon(j)$  correspond aux résidus qui doivent suivre une distribution normale ; ils peuvent être définis comme la différence entre les valeurs observées et prédites de y notées  $\hat{y}_i$ .

$$\varepsilon_i = y_i - \hat{y}_i \quad (\text{II.56})$$

Les estimateurs  $\{\beta\}$  sont calculés en utilisant la technique des moindres carrés ordinaires.

### II.7.2. Machines à vecteurs supports SVM

Plusieurs générations de machines d'apprentissage ont vu le jour dans le but de classifier, de catégoriser ou de prédire des structures particulières dans les données. Mais la plupart de ces techniques éprouvent de grandes difficultés à traiter les données de très haute dimension. Pour surmonter ce problème, on procède souvent par la sélection d'une partie des attributs des données pour réduire la dimension de l'espace d'entrée. Mais dans ce cas on aura besoin d'utiliser des hypothèses simplificatrices qui ne se vérifient pas toujours en pratique.

Par ailleurs, une méthode issue récemment d'une formulation de la théorie de l'apprentissage statistique due en grande partie à l'ouvrage de Vapnik en 1995 intitulé « The nature of learning statistical theory » [II.47] surmonte ce problème en imposant que le nombre de paramètres soit linéairement lié au nombre des données d'apprentissage. Cette technique est appelée machines à vecteurs support (SVM : Support Vector Machines, en anglais)

SVM est un modèle discriminant qui tente de minimiser les erreurs d'apprentissage tout en maximisant la marge séparant les données des classes. La maximisation de la marge est une méthode de régularisation qui réduit la complexité du classifieur [II.48].

Cela revient à ne pas considérer les erreurs inférieures à  $\varepsilon$  et à interdire celles supérieures à  $\varepsilon$  [II.49].

Maximiser la platitude de la fonction permet de minimiser la complexité du modèle qui influe sur ses performances en généralisation. En effet, la théorie de l'apprentissage [II.48] permet de borner l'erreur de généralisation par une somme de deux termes : l'un dépendant de la complexité du modèle et l'autre dépendant de l'erreur sur les données d'apprentissage [II.50].

Les méthodes SVM sont basées sur le contrôle de la complexité du modèle lors de l'apprentissage.

Dans la méthode SVM, différents hyperparamètres apparaissent :  $C$ , qui représente le compromis entre la complexité du modèle et l'erreur sur les données d'apprentissage;  $\varepsilon$ , qui correspond à la largeur du tube d'insensibilité ; les éventuels paramètres de la fonction noyau  $k(\alpha, \gamma, \dots)$ . Ces hyperparamètres sont en général réglés en fonction d'une estimation de l'erreur de généralisation qui peut être évaluée sur un jeu indépendant de données de validation ou par validation croisée [II.51].

Cela implique de réaliser l'apprentissage pour différentes valeurs et d'estimer leur performance. Dans le cas d'une estimation de l'erreur de généralisation par validation croisée, cette procédure peut se révéler très coûteuse en temps de calcul.

### II.7.3. Paramètres statistiques d'évaluation d'un modèle QSAR/QSPR

L'analyse de régression linéaire multiple (RLM) et la sélection des variables ont été effectuées avec le logiciel MobyDigs [II.45] en utilisant la méthode de régression par des moindres carrés ordinaire (OLS) et l'algorithme génétique (GA-VSS) pour la sélection des sous-ensembles [II.52].

Deux paramètres sont couramment utilisés comme critères pour le choix du modèle linéaire :

- **Le coefficient de détermination multiple:**

Le coefficient de détermination  $R^2$  est une mesure de la précision de l'ajustement de la droite de régression. Il augmente de façon monotone avec l'introduction de nouvelles variables même si celles-ci sont peu

corrélées avec la variable expliquée Y. Son utilisation n'est donc pas recommandée sauf dans le cas de modèle à nombres de variables identiques. Il est donné par l'équation suivante :

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{SCE}{SCT} \quad (\text{II.57})$$

Où

$\hat{y}_i$  est la valeur estimée de la propriété étudiée de l'ensemble de calibrage, et

$\bar{y}$  la moyenne des valeurs observées de cet ensemble.

SCE est la somme des carrés des résidus ( $e_i = y - \hat{y}_i$ ) la somme des carrés totale ( $= \sum (y_i - \bar{y}_i)^2$ )

- **La racine de l'erreur quadratique moyenne de prédiction** (désignée également par SDEP) :

$$\sigma_N = EQMP = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} = \sqrt{\frac{PRESS}{n}} \quad (\text{II.58})$$

PRESS (Predictive Residual Sum of Squares) dans l'équation précédente est une mesure de la dispersion des estimations.

- La racine de l'écart quadratique moyen (RMSE) calculée sur les ensembles de calibrage (EQMC), et sur l'ensemble de validation externe (EQMP<sub>ext</sub>), sont deux paramètres à considérer, également. Ces deux paramètres sont définis par les équations suivantes :

$$EQMC = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (\text{II.59})$$

$$EQMP_{ext} = \sqrt{\frac{\sum_{i=1}^{n_{ext}} (y_i - \hat{y}_{(i)})^2}{n_{ext}}} \quad (\text{II.60})$$

- La stabilité des modèles a été explorée en utilisant la "validation croisée par omission d'une observation" (LOO : cross-validation by leave-one-out) [II.53]. Elle consiste à recalculer le modèle sur (n-1) composés, le modèle obtenu servant alors à estimer la valeur de la propriété du composé écarté noté  $\hat{y}_{(i)}$ . On répète le procédé pour chacun des composés étudiés.

$$Q_{LOO}^2 = \frac{SCT - PRESS}{SCT} \quad (\text{II.61})$$

Contrairement à  $R^2$  qui augmente avec le nombre de paramètres du modèle, le facteur  $Q_{LOO}^2$  affiche une courbe avec maximum (ou avec palier) obtenu pour un certain nombre de descripteurs, puis décroît de façon monotone. Ce fait confère une grande importance au coefficient  $Q_{LOO}^2$ . Une valeur de  $Q_{LOO}^2 > 0.5$  est considérée comme satisfaisante, et une valeur supérieure à 0,9 est excellente [II.53,II.54].

Si de petites valeurs de  $Q_{LOO}^2$  indiquent des modèles peu robustes, caractérisés par de faibles capacités prédictives internes, le contraire n'est pas nécessairement vrai. En fait, si une forte valeur de  $Q_{LOO}^2$  est une condition nécessaire de robustesse et d'une possible capacité prédictive élevée d'un modèle, cette condition seule n'est pas suffisante, et peut conduire à une surestimation de la capacité prédictive du modèle lorsqu'il est appliqué à des composés réellement externes.

- La technique de validation du bootstrap est une validation interne. Elle consiste à simuler k échantillons de taille n, la même que celle de l'échantillon initial. Le tirage est fait au hasard. Le modèle obtenu pour l'ensemble des objets sectionnés sera utilisé pour prédire l'ensemble exclu. Le bootstrapping a été répété 500 fois pour la validation de chaque modèle [II.55].

Obtenir un modèle robuste ne donne aucune information réelle sur son pouvoir prédictif. Cela sera évalué par la prédiction du modèle en utilisant l'ensemble de test. Le paramètre  $Q_{ext}^2$  [II.56] qui a pour principe de valider un nombre d'observations qui n'ont pas contribué à la construction du modèle, est déterminé par l'équation suivante :

$$Q_{ext}^2 = 1 - \frac{\sum_{i=1}^{n_{ext}} (\hat{y}_{i/i} - y_i)^2}{\sum_{i=1}^{n_{tr}} (y_i - \bar{y}_{tr})^2 / n_{tr}} \quad \text{(II.62)}$$

Où  $y_i$  et  $\hat{y}_{i/i}$  sont respectivement les valeurs mesurées et prédites (sur l'ensemble de prédiction) de la variable dépendante, et  $\bar{y}$  est la valeur moyenne de la variable dépendante pour l'ensemble de calibrage.  $n_{tr}$  et  $n_{ext}$  sont respectivement le nombre des observations dans l'ensemble de calibrage et l'ensemble de test.

- D'autres paramètres utiles sont : l'écart-type de prédiction externe (SDEP<sub>ext</sub>), défini comme suit:

$$SDEP_{ext} = \sqrt{\frac{1}{n_{ext}} \sum_{i=1}^{n_{ext}} (y_i - \bar{y})^2} \quad \text{(II.63)}$$

Où la somme porte sur l'ensemble des objets tests ( $n_{ext}$ ).

- Selon Golbraikh et Tropsha [II.57], un bon modèle QSPR doit satisfaire aux critères suivants:

$$R_{CV_{ext}}^2 > 0.5 \quad \text{(II.64)}$$

$$r^2 > 0.6 \quad (\text{II.65})$$

$$(r^2 - r_0^2)/r^2 < 0.1 \text{ or } (r^2 - r_0'^2)/r^2 < 0.1 \quad (\text{II.66})$$

$$0.85 \leq k \leq 1.15 \text{ or } 0.85 \leq k' \leq 1.15 \quad (\text{II.67})$$

$$Ab = |r^2 - r_0'^2| < 0.3 \quad (\text{II.68})$$

Où:

$$r = \frac{\Sigma(y_i - \bar{y}_i)(\tilde{y}_i - \bar{y})}{\sqrt{\Sigma(y_i - \bar{y}_i)^2 \Sigma(\tilde{y}_i - \bar{y})^2}} \quad (\text{II.69})$$

$$r_0^2 = 1 - \frac{\Sigma(\tilde{y}_i - \tilde{y}_i^{t_0})}{\Sigma(\tilde{y}_i - \bar{y})^2} \quad (\text{II.70})$$

$$r_0'^2 = 1 - \frac{\Sigma(y_i - y_i^{t_0})^2}{\Sigma(y_i - \bar{y})^2} \quad (\text{II.71})$$

$$k = \frac{\Sigma y_i \tilde{y}_i}{\Sigma y_i^2} \quad (\text{II.72})$$

$$k' = \frac{\Sigma y_i \tilde{y}_i}{\Sigma \tilde{y}_i^2} \quad (\text{II.73})$$

$$T1 = \frac{(r^2 - r_0^2)}{r^2} \quad (\text{II.74})$$

$$T2 = \frac{(r^2 - r_0'^2)}{r^2} \quad (\text{II.75})$$

Ces critères de validation externe sont appliqués uniquement à l'ensemble de test et recommandés pour évaluer la prédictivité du modèle QSPR.

$R^2$  est le coefficient de corrélation entre les valeurs calculées et expérimentale de l'ensemble de test;  $r_0^2$  (valeurs calculées par rapport à celles observées) et  $r_0'^2$  (valeurs observées par rapport à celles calculées) sont les coefficients de détermination;  $k$  et  $k'$  sont les pentes des droites de régressions passant par l'origine pour les valeurs calculées par rapport aux valeurs observées et observées par rapport aux calculées, respectivement;  $y_i^{t_0}$  et  $\tilde{y}_i^{t_0}$   $y$  et  $0$  sont donnés respectivement par :  $y_i^{t_0} = k\tilde{y}$  et  $\tilde{y}_i^{t_0} = k'y$  ; et les sommations portent sur tous les échantillons dans l'ensemble de test.

### II.7.4. Analyse des résidus pour la détection des observations aberrantes : (Commande « regstats » du logiciel Matlab [II.58])

Les tests statistiques utilisés pour la détection des observations aberrantes en analyse de régression sont réunis dans les tableaux III.5 et III.6 pour les hydrocarbures non saturés, les tableaux III.9 et III.10 pour les n-alcanes et dans les tableaux III.16 et III.17 pour la série des solvants.

- Les résidus ordinaires  $e_i$ , différences entre les valeurs observées ( $y_i$ ) et estimées par le Modèle ( $\hat{y}_i$ ).

$$e_i = y_i - \hat{y}_i \quad (\text{II.76})$$

Les résidus de prédiction standardisés  $e_i$  st d ou bien  $d_i$ , obtenus en divisant les  $e_i$  par l'écart type estimé(s).

- Les leviers,  $h_{ii}$ , permettent de juger de l'influence d'une observation  $i$  dans la détermination de l'équation de régression.

Les valeurs de  $h_{ii}$ ,  $i^{\text{ème}}$  terme diagonale de la matrice de projection (ou matrice chapeau) :

$H = X(X'X)^{-1} X'$  ou  $X$  est la matrice des valeurs observées des variables explicatives et  $X'$  sa transposée, ces valeurs sont utiles pour le calcul des résidus caractéristiques.

Les observations avec des valeurs supérieures à  $\frac{2(k+1)}{n}$  sont considérées comme potentiellement très influentes, où  $k$  est le nombre de prédicteurs et  $n$  est le nombre d'observations.

- Le résidu studentisé interne  $r_i$ , est le résidu d'une prédiction divisé par son écart type propre.

$$r_i = \frac{e_i}{\sqrt{1-h_{ii}}} \quad (\text{II.77})$$

- Les estimations  $S_{(i)}^2$  de  $\sigma^2$  calculées selon l'équation suivante :

$$S_{(i)}^2 = \frac{(n-p) \text{CME} - \frac{e_i^2}{1-h_{ii}}}{n-p-1} \quad (\text{II.78})$$

Pour  $n-1$  observations, la  $i^{\text{ème}}$  étant exclue ; CME est le carré moyen des écarts, et  $k = (p - 1)$  le nombre de variables explicatives.

### II.7.5. Diagnostics d'influence

D'autres mesures d'influence utiles dont l'étude complète la recherche des observations aberrantes [II.59].

- La statistique représentée par le symbole DFFITS ou DFITS permet de mesurer l'influence d'une observation  $i$  sur la valeur ajustée ou prédite ;

$$DFFITS = \frac{1}{p} \sqrt{\left(\frac{h_{ii}}{1-h_{ii}}\right)} t_i \quad (\text{II.79})$$

Les DFFIS combinent le bras de levier  $h_{ii}$  et le résidu studentisé en une mesure globale, qui renseigne jusqu'où une observation est inhabituelle.

Les valeurs absolues de  $DFFITS_i$  supérieures à la valeur critique  $\sqrt{p/n}$  avec  $(p = k + 1)$  sont inhabituelles.

- La distance de Cook,  $D_i$ , permet d'étudier l'influence d'une observation  $i$  sur les coefficients de régression estimés par les moindres carrés. Les valeurs supérieures à  $\left(\frac{4}{n}\right)$  sont considérées comme très influentes.

$$D_i = \frac{1}{p} \frac{h_{ii}}{1-h_{ii}} d_i^2 \quad (\text{II.80})$$

Cette relation montre que la distance  $D_i$  est une fonction croissante du carré du résidu standardisé  $d_i$  et de  $h_{ii}$ . Pour une valeur fixée de  $p$  et, pour une régression avec terme indépendant,  $D_i$  sera d'autant plus grand que  $e_i$  est grand, en valeur absolue et que le vecteur  $X_i$  est éloigné du vecteur  $\bar{X}_i$ .

Les valeurs de  $D_i$  peuvent être comparées à une valeur  $F_{1-\alpha}$  relative à la variable  $F$  de SNEDECOR à  $p$  et  $(n-p)$  degrés de liberté, bien qu'il ne s'agisse pas d'un test statistique rigoureux [II.60].

Sur cette base, WEIBERG considère qu'une attention particulière doit être donnée aux observations ayant une valeur de  $D_i$  supérieure à l'unité.

En fait, il est suggéré de contrôler les observations pour lesquelles :

$$COOK > F(p; n - p) \cong 1$$

- La disposition des points dans l'espace des x (régresseurs) est importante pour la détermination des propriétés du modèle. En particulier, les observations éloignées ont, potentiellement, des influences disproportionnées sur les paramètres estimés, les valeurs prédites, et les statistiques élémentaires. La somme pondérée des carrés des distance du point i au centre des données WSSDi (Weighted of the Sum Squared Distance of the center of data) peut être utilisée pour localiser les points éloignés dans l'espace des x [II.60] soit:

$$WSSD_i = \sum_{j=1}^n \left[ \frac{\hat{\beta}_j(X_{ji} - \bar{X}_j)}{\sqrt{CME}} \right]^2, \quad i = 1, 2, \dots, n \quad (\text{II.81})$$

### II.7.6. Statistique DFBETAS<sub>j,i</sub>.

L'ampleur du changement, en unité d'écart-type, du  $j^{\text{ème}}$  coefficient de régression  $\hat{\beta}_j$  estimé, est indiquée par la statistique :

$$DFBETAS_{i,j} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{\sqrt{s_{(j)}^2 C_{jj}}} \quad (\text{II.82})$$

Où  $C_{jj}$  est le jème élément diagonal de la matrice et  $\hat{\beta}_{j(i)}$  le  $j^{\text{ème}}$  coefficient de régression calculé sans utiliser l'observation i. L'examen de la colonne DFBETAS<sub>i,j</sub> en tenant compte de la valeur critique  $\frac{2}{\sqrt{n}}$  fait ressortir les points influents, avec n le nombre d'individus de calibrage.

### II.7.7. COVRATIO<sub>i</sub>

Le COVRATIO<sub>i</sub>, que l'on calcule pratiquement à partir de la relation (II.83) fait ressortir le rôle de la  $i^{\text{ème}}$  observation sur la précision de l'estimation

$$COVRATIO_i = \frac{(S_{(i)})^p}{(CME)^p} \left( \frac{1}{1-h_{ii}} \right), \quad i = 1, 2, \dots, n \quad (\text{II.83})$$

En résumé si COVRATIO<sub>i</sub> > 1, la  $i^{\text{ème}}$  observation améliore la précision de l'estimation, alors que si COVRATIO<sub>i</sub> < 1 l'inclusion de la  $i^{\text{ème}}$  observation détériore la précision.

### II.7.8. Test de randomisation

Ce test permet de mettre en évidence des corrélations dues au hasard. Il consiste à générer un vecteur « propriété considérée » par permutation aléatoire des composantes du vecteur réel. On calcule alors sur le vecteur obtenu (considéré comme vecteur expérimental réel) un modèle QSAR, selon la méthode habituelle. Ce procédé est répété plusieurs fois (100 dans notre cas).

### . II.7.9. Le domaine d'application du modèle QSPR

Le domaine d'application (AD) [II.56-II.54] est une région théorique dans l'espace défini par les descripteurs du modèle et la réponse à modéliser, pour laquelle un QSPR donné devrait faire des prédictions fiables. Dans ce travail, le domaine d'application structurale a été vérifié par l'approche du levier ( $h_{ii}$ ) [II.60].

Le levier critique est en général fixé à  $(K + 1) / n$ , où  $n$  est le nombre total des objets dans l'ensemble de calibrage et  $k$  est le nombre de descripteurs impliqués dans la corrélation.

La présence des valeurs aberrantes en réponse (Y outliers) et des composés structurellement influents (X outliers) a été, en plus, vérifiée par le diagramme de Williams [II.61], représentant les résidus standardisés en fonction des valeurs des leviers  $h_{ii}$ .

# **PARTIE PRATIQUE**

# CHAPITRE III:

## Résultats et discussions



### III.1. Modélisation du point d'éclair d'un ensemble de n-alcanes

#### III.1.1. Introduction

Les alcanes sont des hydrocarbures saturés. Ils ne sont constitués que d'atomes de carbone (C) et d'hydrogène (H), liés par des liaisons simples, les atomes de carbone sont reliés à un nombre maximal d'atomes d'hydrogène. Ils peuvent être linéaires ou ramifiés et possèdent alors une formule brute  $C_nH_{2n+2}$  ou cycliques de formule générale  $C_nH_{2n}$ , où n est un nombre entier.

Les alcanes sont très répandus dans la nature. On les trouve sous forme de gisements de gaz naturel (composés principalement de méthane) et de pétrole (il existe une très grande variété de pétroles qui diffèrent selon les mélanges d'hydrocarbures qu'ils contiennent). La séparation de ces alcanes est réalisée par distillation dans la tour de distillation de la raffinerie.

Le risque d'incendie et d'explosion est l'un des risques majeurs lors de l'utilisation des solvants pétroliers. Les vapeurs de ces substances peuvent former avec l'air des mélanges explosifs. En effet, ils sont tous inflammables à l'exception des solvants de classe A3 ( $Tec > 55\text{ °C}$ ) qui sont non inflammables à température ambiante.

Au cours des deux dernières décennies de nombreuses études QSPR ont été consacrées aux températures d'éclair des n-alcanes. Nous citerons :

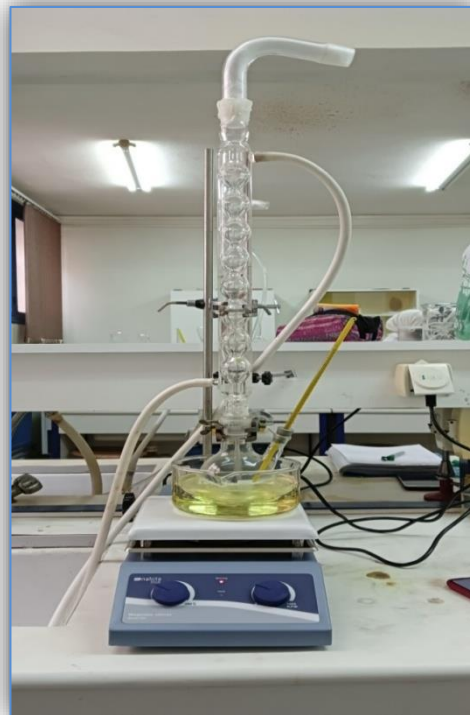
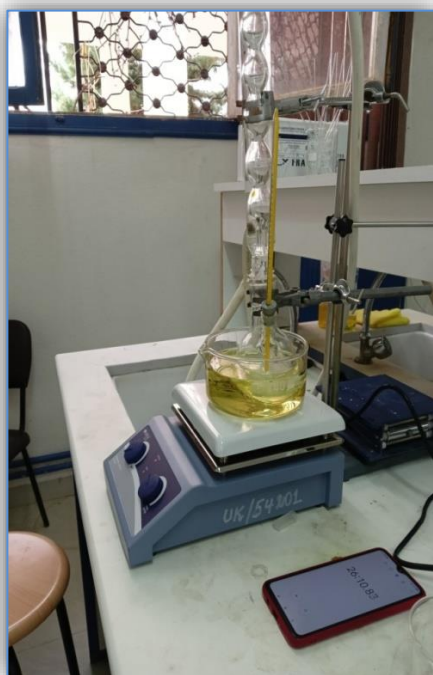
- Une méthode de contribution de groupes proposée par Wang en 1999 pour la prédiction du point d'éclair [III.1].
- Une prédiction de Tec des alcanes par la méthode de contribution de groupes en utilisant les réseaux de neurones artificiels [III.2].
- L'application d'une approche inductive dans les relations structure-propriétés pour le développement d'un modèle QSAR pour la prédiction des points d'éclair d'une série d'alcanes [III.3].

Dans ce travail nous nous sommes intéressés à la modélisation des points d'éclair d'une série de 92 n-alcanes. Nous avons développé un modèle QSPR en utilisant la technique de régression linéaire multiple en vue de prédire cette propriété à partir des descripteurs moléculaires théoriques calculés à l'aide du logiciel DRAGON [III.4].

#### III.1.2. Données et méthodes de recherches

Les valeurs expérimentales des points d'éclair des 92 alcanes étudiés varient entre 169 K (Propane) et 408 K (Hexadécane), ont été prélevées de la littérature [III.2].

Le nombre d'atomes de carbone dans les structures chimiques des alcanes étudiés varie de 3 à 16 atomes de carbone.



**Figure III.1:** Le montage à reflux utilisé

**Tableau III.1:** Composés étudiés et valeurs des températures d'éclair expérimentales.

N°	Composés	T <sub>exp</sub> (K)	N°	Composés	T <sub>exp</sub> (K)
1	Propane	169	14	3-Éthylhexane	278
2	Butane	213	15	3-Méthylheptane	279
3	2,2-Diméthylpropane	208	16	2,3-Diméthylhexane	283
4	Hexane	250	17	2,4-Diméthylhexane	283
5	2,2-Diméthylbutane	225	18	2,5-Diméthylhexane	271
6	2,3-Diméthylbutane	244	19	3,4-Diméthylheptane	288
7	Heptane	269	20	3-Éthyl-2,3-diméthylpentane	288
8	3,3-Diméthylpentane	245	21	2,2,5-Triméthylhexane	286
9	2,4-Diméthylpentane	261	22	3-Méthyloctane	297
10	2,2-Diméthylpentane	250	23	2,2,3,4-Tetraméthylpentane	284
11	2,2,4-Triméthylpentane	261	24	4-Éthylheptane	288
12	2,2,3-Triméthylpentane	270	25	3,3,4-Triméthylhexane	288
13	3-Éthyl-3-méthylpentane	276	26	2,3,5-Triméthylhexane	288

Tableau III.1 (suite)

N°	Composés	Tec <sub>exp</sub> (K)	N°	Composés	Tec <sub>exp</sub> (K)
27	2,2,3-Triméthylhexane	288	48	Dodécane	344
28	3,5-Diméthylheptane	288	49	Tétradécane	372
29	Tetraéthylméthane	294	50	Hexadécane	408
30	Décane	319	51	2,2,4,4,6,8,8- Heptaméthylnonane	368
31	4-Éthyloctane	314	52	2-Méthylbutane	216
32	2,4,5-Triméthylheptane	304	53	2-Méthylpentane	250
33	2,3-Diméthylloctane	314	54	2,3,3-Triméthylpentane	273
34	3,3-Diméthylloctane	314	55	2,2-Diméthylhexane	296
35	3,5-Diméthylloctane	314	56	3-Éthyl-2,2-diméthylpentane	286
36	2,6-Diméthylloctane	314	57	2,3,3-Triméthylhexane	288
37	3,3,4,4-Tétraméthylhexane	304	58	2,4,6-Triméthylheptane	304
38	2,2,5,5-Tétraméthylhexane	304	59	3-Éthyl-2,3,4-triméthylpentane	304
39	2,3,3,4-Tétraméthylhexane	304	60	2,3,4,4-Tétraméthylhexane	304
40	2,3,4,5-Tétraméthylhexane	304	61	3,4,5-Triméthylheptane	304
41	2,2,4,4-Tétraméthylhexane	304	62	3-Éthyl-5-méthylheptane	304
42	3,3,5-Triméthylheptane	304	63	2-Méthylpropane	186
43	2,3,5-Triméthylheptane	304	64	2,3-Diméthylpentane	258
44	3-Éthyl-3-méthylheptane	314	65	3-Méthylhexane	258
45	4-Éthyl-3-méthylheptane	314	66	2,2,3-Triméthylbutane	247
46	2-Méthylnonane	314	67	Octane	286
47	Undécane	333	68	2,2,3,3-Tétraméthylbutane	273

Tableau III.1 (suite et fin)

N°	Composés	Tec <sub>exp</sub> (K)	N°	Composés	Tec <sub>exp</sub> (K)
69	2,3,4-Triméthylpentane	273	81	2,2-Diméthylheptane	297
70	3,4-Diméthylhexane	277	82	2,4-Diméthylheptane	288
71	2,7-Diméthylheptane	314	83	2,3,4-Triméthylhexane	288
72	Tridécan	352	84	3-Éthyl-2,2-diméthylhexane	311
73	Pentadécane	388	85	2-Méthylheptane	277
74	Nonane	304	86	2,4,4-Triméthylhexane	288
75	Pentane	224	87	4,4-Diméthylheptane	288
76	3-Éthylpentane	255	88	5-Méthylnonane	312
77	3,3-Diméthylhexane	272	89	3-Méthylpentane	241
78	2,3-Diméthylheptane	288	90	3-Éthyl-4-méthylhexane	288
79	2,6-Diméthylheptane	299	91	2,2,4-Triméthylhexane	288
80	2,2,4,4-Tétraméthylpentane	276	92	2,2,3,4-Tétraméthylhexane	304

Certains modèles, parmi les 100 modèles obtenus par le logiciel MOBYDIGS [III.5], peuvent présenter des performances similaires. Dans ce cas nous avons opté pour les modèles avec les valeurs de  $\Delta k$  les plus élevées tout en tenant compte des autres paramètres statistiques.

### III.1.3. Résultats et discussions

#### III.1.3.1. Calcul du modèle

Après l'exécution du programme, le meilleur modèle MLR à deux dimensions obtenu, est donné par l'équation de régression suivante.

$$Tec = 65,212(\pm 4,191) + 1.898(\pm 0,251)POL + 63,993(\pm 1,631)DP0 \quad (III.5)$$

$$n_{cal} = 75 \quad n_{test} = 18 \quad R^2 = 98,26\% \quad R^2_{(ajst)} = 98,21\%$$

$$Q^2 = 97,99\% \quad Q^2_{(ext)} = 98,88\%$$

$$S = 5,39 \quad F=2006,91 \quad kx= 70,56 \quad kxy= 82,57 \quad SDEC = 5,28 \quad SDEP=5,67 \quad SDEP_{ext} = 4,24$$

les valeurs élevées et proches de  $R^2$  et  $R^2$  (aj) confirment le bon ajustement du modèle, la petite différence en valeur absolue entre  $R^2$  et  $Q^2(=0,27)$  montre sa robustesse.

La grande valeur du F de Fisher est un gage de la bonne prédiction des n (=74) valeurs de Tec de l'ensemble de calibrage.

La bonne capacité prédictive externe du modèle est confirmée par la valeur élevée de  $Q^2$  (ext).

Ici on a:  $K_{xy} - K_x = 82,57 - 70,56 = 12,01 > 5$  (une valeur positive). Cette valeur vérifie la condition de la règle QUICK (Q Under Influence K) [III.6], basée sur l'indice de corrélation multivariable K.

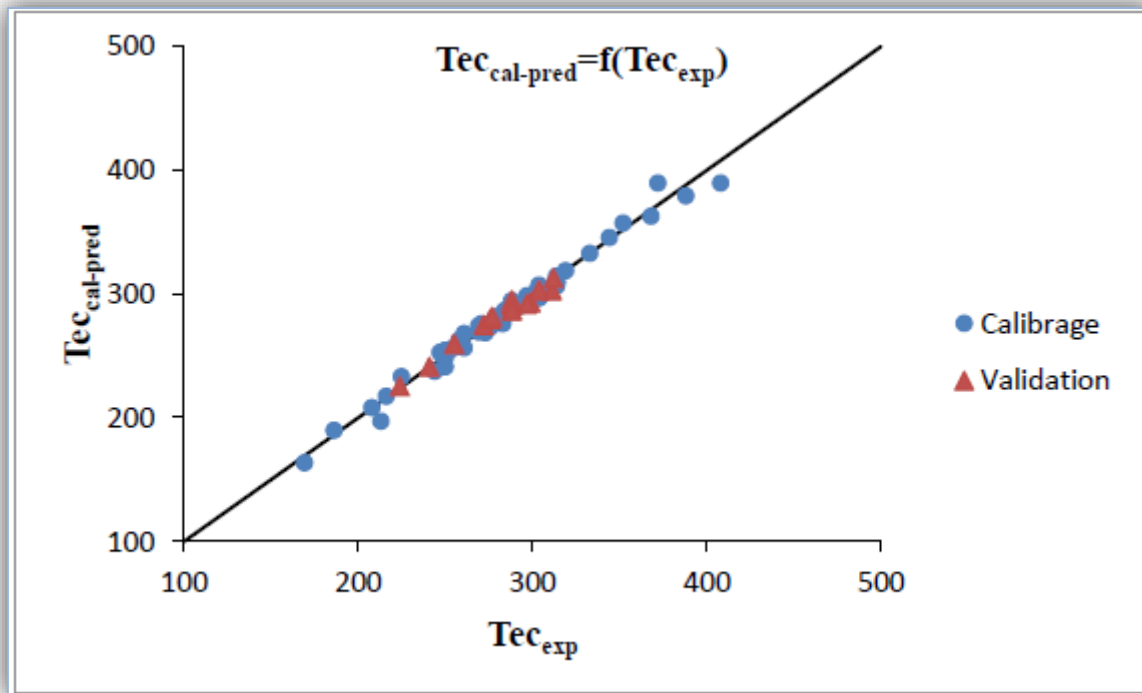
Cette règle est déduite de l'hypothèse, que la corrélation totale dans l'ensemble formé par les prédicteurs X du modèle plus la réponse Y ( $K_{xy}$ ) doit toujours être plus grande que celle uniquement mesurée dans l'ensemble des prédicteurs ( $K_x$ ). Le calcul de  $K_{xy}$  est réalisé en considérant la réponse Y comme une variable et en calculant la matrice de corrélation correspondante.

Pol et DP01 sont des descripteurs moléculaires calculés en utilisant le logiciel MOBYDIGSn. La définition des descripteurs utilisés pour le calcul du modèle est donnée au tableau III.2 (page suivante).

**Tableau III.2:** Classe et définition des deux descripteurs moléculaires

Descripteur	Classe	Définition
Pol	Descripteur topologique (bloc 2)	Les descripteurs topologiques sont basés sur une représentation graphique de la molécule. Ce sont des quantificateurs numériques de la topologie moléculaire obtenus par l'application des opérateurs algébriques à des matrices représentant des graphes moléculaires et dont les valeurs sont indépendantes de la numérotation ou de l'étiquetage des sommets (atomes). Ils peuvent être sensibles à une ou plusieurs caractéristiques structurales de la molécule telle que la taille, la forme, la symétrie, la ramification et la cyclicité et peuvent également encoder des informations chimiques concernant le type d'atome et la multiplicité des liaisons. Pol est le nombre de polarité, il est calculé sur la matrice de distance comme le nombre de paires de sommets à une distance topologique égale à trois [III.5]. On suppose généralement que le nombre de polarité tient compte de la flexibilité des structures acycliques, le nombre de polarité étant égal au nombre de liaisons autour desquelles les rotations libres peuvent avoir lieu. De plus, il concerne les propriétés stériques des molécules.
DP01	Profils moléculaires de Randic (bloc 11)	Les profils moléculaires sont des séquences de descripteurs moléculaires proposés par Randic et dérivés des distances géométriques interatomiques d'une molécule [III.8-III.9-III.10]. Le logiciel DRAGON fournit deux profils moléculaires dont l'un est beaucoup plus lié à la structure 3D moléculaire globale: (DP01, DP02, ..., DPk, ..., DP20) et l'autre lié à la forme moléculaire où DP01 est le profil moléculaire N° :01.

Une comparaison graphique entre les valeurs expérimentales et prédites du point d'éclair des deux ensembles est présentée dans la figure III.1. On remarque que la faible dispersion des données autour de la droite d'ajustement vérifie le bon ajustement du modèle obtenu.



**Figure III.2:** Droite d'ajustement des Tec prédites en fonction des Tec expérimentales pour les ensembles de calibration et de test.

### III.1.3.2. Analyse des résidus

Les résultats obtenus sont condensés dans le tableau III.3, dont les colonnes sont numérotées de (1) à (9). La première rassemble les résidus ordinaires  $e_i$ , Notons que 20 résidus ordinaires dépassent en valeur absolue l'erreur standard,  $S = 5,39$ . Remarquons que 3 valeurs absolues des résidus ordinaires sont importantes. Ce sont dans l'ordre décroissant:  $e_{50}$ ,  $e_{49}$  et  $e_2$ . Notons que les observations 2, 49, et 50 désignent respectivement : le butane, le tétradécane et l'hexadécane.

Notons aussi que les observations  $e_{50}$ ,  $e_{49}$  ont des résidu ordinaires, en valeurs absolues, supérieures à 3 fois l'erreur standard ( $|e_i| > 3S$ ), soit  $3 \times 5,39 = 16,17$ .

Tous les résidus standardisés  $d_i$  de la colonne (2) sont compris entre les limites  $\pm 3$ , à l'exception des points déjà signalés 2, 49 et 50

La colonne (3) rassemble les résidus studentisés internes  $r_i$  qui sont du même ordre de grandeur que les  $d_i$  correspondants. On a ici  $p = 3$  et  $n = 74$ , et on constate que tous les  $t_i$  exceptés  $r_2$ ,  $r_{49}$ , et  $r_{50}$  sont inférieurs en valeur absolue à  $t_{(0,025;n-p)} [= 1,9945]$  qui est le 0,975 quantile d'une loi de Student avec  $(n-p)$  degrés de liberté

La colonne (4) donne les valeurs de  $h_{ii}$ ,  $i^{\text{ème}}$  terme diagonal de la matrice de projection (ou matrice chapeau) :  $H = X (X' X)^{-1} X'$  où  $X$  est la matrice des valeurs observées des variables explicatives et  $X'$  sa transposée, ces valeurs sont utiles pour le calcul des résidus caractéristiques . Nous observons la plus grande valeur de  $h_{ii}$  pour le point (1) avec  $h_{ii}=0,1475 > h^*$ , avec  $h^*= 3(k+1)/n_r = 3(2+1) /74 = 0,1216$  . Ce point est dit point influent. L'observation (1) désigne le propane.

La colonne (5) contient les résidus prédits, qui sont du même ordre de grandeur que les résidus ordinaires correspondants.

La colonne (6) montre le calcul de la somme des carrés des erreurs de prédiction (statistique PRESS) obtenue pour ce modèle. La valeur du PRESS = 2378,6162, et la valeur de SCE = 2059,6254, obtenue pour ce modèle

La colonne (7) condense les estimations  $S_{(i)}^2$  de  $\sigma^2$  calculées selon l'équation .

$S_{(i)}^2$  intervient dans le calcul des résidus studentisés externes, rassemblés dans la colonne (8); tous les  $t_i$  sont du même ordre de grandeur que les  $r_i$  correspondants.

Comme les  $t_i$  sont inférieurs en valeur absolue à  $t(0,025;n-p-1) [=1,995]$ , à l'exception encore une fois de ceux des points 2, 49 et 50 l'analyse des résidus studentisés internes et externes permet de détecter ces 3 observations comme aberrantes.

Enfin, en désignant par SCT la somme des carrées totale (= 11495,959), la statistique PRESS (=2378,6162) conduit à un  $R^2$  de prédiction égal à :

$$R_{pred}^2 = 1 - \frac{PRESS}{SCT} = 1 - \frac{2378,6162}{118495,959} = 97,99\%$$

Ainsi, le modèle permettrait d'expliquer environ 98 % de la variabilité de nouvelles observations estimées.

Tableau III.3: Résidus caractéristiques et valeurs estimées des points d'éclair.

	Colonne1	Colonne2	Colonne3	Colonne4	Colonne5	Colonne6	Colonne7	Colonne8	Colonne9
<i>Observation i</i>	$e_i$	$d_i$	$r_i$	$t_i$	$h_{ii}$	$e_i/(1-h_{ii})$	$(e_i/(1-h_{ii}))^2$	$(s_i)^2$	Tec estimé
<b>1</b>	4,9836	1,0021	1,0021	1,0022	0,1475	5,8457	34,1723	29,0070	164,0164
<b>2</b>	15,2171	2,9534	2,9534	3,1311	0,0848	16,6278	276,4838	25,8086	197,7829
<b>3</b>	-0,7473	-0,1450	-0,1450	-0,1440	0,0840	-0,8158	0,6656	29,4145	208,7473
<b>4</b>	0,9381	0,1778	0,1778	0,1766	0,0404	0,9777	0,9558	29,4101	249,0619
<b>5</b>	-8,7036	-1,6516	-1,6516	-1,6724	0,0427	-9,0914	82,6541	28,2928	233,7036
<b>6</b>	6,4144	1,2130	1,2130	1,2171	0,0360	6,6538	44,2733	28,8135	237,5856
<b>7</b>	-0,6459	-0,1222	-0,1222	-0,1213	0,0362	-0,6702	0,4492	29,4170	269,6459
<b>8</b>	-1,9083	-0,3584	-0,3584	-0,3562	0,0228	-1,9527	3,8130	29,3700	255,9083
<b>9</b>	4,0246	0,7590	0,7590	0,7567	0,0306	4,1518	17,2373	29,1845	256,9754
<b>10</b>	-4,8636	-0,9171	-0,9171	-0,9160	0,0304	-5,0161	25,1608	29,0747	254,8636
<b>11</b>	-7,6644	-1,4403	-1,4403	-1,4515	0,0238	-7,8512	61,6410	28,5636	268,6644
<b>12</b>	-2,5671	-0,4813	-0,4813	-0,4787	0,0194	-2,6180	6,8537	29,3272	272,5671
<b>13</b>	1,7907	0,3371	0,3371	0,3350	0,0271	1,8405	3,3873	29,3761	274,2093
<b>14</b>	-0,0921	-0,0172	-0,0172	-0,0171	0,0145	-0,0935	0,0087	29,4231	278,0921
<b>15</b>	-2,7613	-0,5179	-0,5179	-0,5152	0,0201	-2,8180	7,9411	29,3121	281,7613

Tableau III.3 (suite)

	Colonne1	Colonne2	Colonne3	Colonne4	Colonne5	Colonne6	Colonne7	Colonne8	Colonne9
<i>Observation i</i>	$e_i$	$d_i$	$r_i$	$t_i$	$h_{ii}$	$e_i/(1-h_{ii})$	$(e_i/(1-h_{ii}))^2$	$(s_i)^2$	Tec estimé
<b>16</b>	5,5478	1,0377	1,0377	1,0382	0,0146	5,6300	31,6971	28,9770	277,4522
<b>17</b>	6,8061	1,2753	1,2753	1,2810	0,0182	6,9319	48,0512	28,7492	276,1940
<b>18</b>	-5,4076	-1,0180	-1,0180	-1,0183	0,0273	-5,5596	30,9088	28,9937	276,4076
<b>19</b>	-6,7510	-1,2629	-1,2629	-1,2683	0,0149	-6,8534	46,9689	28,7623	294,7510
<b>20</b>	-2,5906	-0,4953	-0,4953	-0,4926	0,0568	-2,7465	7,5433	29,3216	290,5906
<b>21</b>	-0,8807	-0,1655	-0,1655	-0,1643	0,0231	-0,9016	0,8129	29,4119	286,8807
<b>22</b>	-2,1457	-0,4030	-0,4030	-0,4006	0,0227	-2,1956	4,8206	29,3559	299,1457
<b>23</b>	-2,6663	-0,5019	-0,5019	-0,4992	0,0270	-2,7402	7,5085	29,3189	286,6663
<b>24</b>	-6,4526	-1,2067	-1,2067	-1,2106	0,0142	-6,5458	42,8477	28,8198	294,4526
<b>25</b>	-3,1242	-0,5911	-0,5911	-0,5883	0,0369	-3,2438	10,5220	29,2785	291,1242
<b>26</b>	-2,1651	-0,4048	-0,4048	-0,4024	0,0136	-2,1949	4,8176	29,3553	290,1651
<b>27</b>	-1,9515	-0,3653	-0,3653	-0,3630	0,0161	-1,9834	3,9340	29,3679	289,9515
<b>28</b>	-5,7487	-1,0749	-1,0749	-1,0761	0,0141	-5,8307	33,9976	28,9444	293,7487
<b>29</b>	3,2174	0,6150	0,6150	0,6123	0,0565	3,4100	11,6282	29,2665	290,7826
<b>30</b>	-0,3674	-0,0700	-0,0700	-0,0695	0,0492	-0,3865	0,1494	29,4212	319,3674

Tableau III.3 (suite)

	Colonne1	Colonne2	Colonne3	Colonne4	Colonne5	Colonne6	Colonne7	Colonne8	Colonne9
<i>Observation i</i>	$e_i$	$d_i$	$r_i$	$t_i$	$h_{ii}$	$e_i/(1-h_{ii})$	$(e_i/(1-h_{ii}))^2$	$(s_i)^2$	Tec estimé
<b>31</b>	4,3387	0,8129	0,8129	0,8109	0,0180	4,4182	19,5204	29,1494	309,6613
<b>32</b>	-1,6081	-0,3013	-0,3013	-0,2994	0,0182	-1,6379	2,6828	29,3856	305,6081
<b>33</b>	2,6109	0,4894	0,4894	0,4868	0,0190	2,6614	7,0832	29,3240	311,3891
<b>34</b>	4,5947	0,8608	0,8608	0,8592	0,0179	4,6782	21,8856	29,1162	309,4053
<b>35</b>	4,8506	0,9087	0,9087	0,9076	0,0177	4,9382	24,3854	29,0810	309,1494
<b>36</b>	3,5492	0,6670	0,6670	0,6644	0,0240	3,6365	13,2242	29,2388	310,4508
<b>37</b>	1,9480	-0,3840	-0,3840	-0,3817	0,1127	-2,1954	4,8200	29,3621	305,9480
<b>38</b>	6,7101	1,2593	1,2593	1,2646	0,0212	6,8553	46,9952	28,7661	297,2899
<b>39</b>	-1,9696	-0,3818	-0,3818	-0,3795	0,0824	-2,1465	4,6075	29,3628	305,9696
<b>40</b>	0,0989	0,0188	0,0188	0,0186	0,0407	0,1031	0,0106	29,4231	303,9011
<b>41</b>	6,1134	1,1435	1,1435	1,1460	0,0148	6,2049	38,5011	28,8813	297,8866
<b>42</b>	-0,0083	-0,0016	-0,0016	-0,0016	0,0183	-0,0085	0,0001	29,4232	304,0083
<b>43</b>	-0,4563	-0,0855	-0,0855	-0,0849	0,0183	-0,4648	0,2160	29,4202	304,4563
<b>44</b>	7,1976	1,3537	1,3537	1,3618	0,0254	7,3851	54,5392	28,6639	306,8024
<b>45</b>	6,6217	1,2452	1,2452	1,2501	0,0252	6,7927	46,1401	28,7807	307,3783

Tableau III.3 (suite)

	Colonne1	Colonne2	Colonne3	Colonne4	Colonne5	Colonne6	Colonne7	Colonne8	Colonne9
<i>Observation i</i>	$e_i$	$d_i$	$r_i$	$t_i$	$h_{ii}$	$e_i/(1-h_{ii})$	$(e_i/(1-h_{ii}))^2$	$(s_i)^2$	Tec estimé
<b>46</b>	-1,0159	-0,1927	-0,1927	-0,1914	0,0420	-1,0605	1,1246	29,4078	315,0159
<b>47</b>	-0,1683	-0,0322	-0,0322	-0,0320	0,0576	-0,1785	0,0319	29,4228	333,1683
<b>48</b>	-1,9452	-0,3739	-0,3739	-0,3716	0,0670	-2,0849	4,3469	29,3653	345,9452
<b>49</b>	-17,8217	-3,5149	-3,5149	-3,8401	0,1138	-20,1099	404,4092	24,3033	389,8217
<b>50</b>	18,1783	3,5852	3,5852	3,9338	0,1138	20,5123	420,7523	24,0964	389,8217
<b>51</b>	4,6721	0,9067	0,9067	0,9056	0,0847	5,1044	26,0549	29,0825	363,3279
<b>52</b>	-1,9190	-0,3673	-0,3673	-0,3650	0,0590	-2,0393	4,1586	29,3673	217,9190
<b>53</b>	8,3613	1,5846	1,5846	1,6020	0,0402	8,7114	75,8886	28,3827	241,6387
<b>54</b>	-0,3134	-0,0590	-0,0590	-0,0586	0,0280	-0,3225	0,1040	29,4218	273,3134
<b>55</b>	-5,3598	-1,0084	-1,0084	-1,0085	0,0261	-5,5037	30,2903	29,0018	274,3598
<b>56</b>	-2,3301	-0,4383	-0,4383	-0,4358	0,0256	-2,3912	5,7177	29,3436	288,3301
<b>57</b>	-2,4419	-0,4589	-0,4589	-0,4563	0,0239	-2,5017	6,2586	29,3360	290,4419
<b>58</b>	-0,7555	-0,1416	-0,1416	-0,1407	0,0192	-0,7703	0,5933	29,4149	304,7555
<b>59</b>	-0,6682	-0,1319	-0,1319	-0,1310	0,1153	-0,7552	0,5704	29,4160	304,6682
<b>60</b>	-0,2634	-0,0505	-0,0505	-0,0501	0,0600	-0,2802	0,0785	29,4222	304,2634

Tableau III.3 (suite et fin)

	Colonne1	Colonne2	Colonne3	Colonne4	Colonne5	Colonne6	Colonne7	Colonne8	Colonne9
<i>Observation i</i>	$e_i$	$d_i$	$r_i$	$t_i$	$h_{ii}$	$e_i/(1-h_{ii})$	$(e_i/(1-h_{ii}))^2$	$(s_i)^2$	Tec estimé
<b>61</b>	-3,7407	-0,7079	-0,7079	-0,7054	0,0376	-3,8867	15,1060	29,2155	307,7407
<b>62</b>	-1,3522	-0,2534	-0,2534	-0,2517	0,0182	-1,3772	1,8968	29,3966	305,3522
<b>63</b>	-3,9974	-0,7822	-0,7822	-0,7800	0,0997	-4,4400	19,7135	29,1697	189,9974
<b>64</b>	-0,7879	-0,1479	-0,1479	-0,1468	0,0210	-0,8048	0,6477	29,4142	258,7879
<b>65</b>	-4,8411	-0,9093	-0,9093	-0,9082	0,0229	-4,9545	24,5466	29,0806	262,8411
<b>66</b>	-6,0926	-1,1455	-1,1455	-1,1481	0,0249	-6,2479	39,0360	28,8794	253,0926
<b>67</b>	-1,9262	-0,3646	-0,3646	-0,3623	0,0376	-2,0015	4,0060	29,3681	287,9262
<b>68</b>	3,7181	0,7019	0,7019	0,6994	0,0328	3,8441	14,7770	29,2190	269,2819
<b>69</b>	-1,2949	-0,2427	-0,2427	-0,2410	0,0182	-1,3190	1,7397	29,3988	274,2949
<b>70</b>	-0,8145	-0,1525	-0,1525	-0,1514	0,0162	-0,8280	0,6855	29,4136	277,8145
<b>71</b>	3,5276	0,6668	0,6668	0,6642	0,0353	3,6567	13,3713	29,2390	310,4724
<b>72</b>	-5,9542	-1,1510	-1,1510	-1,1537	0,0775	-6,4545	41,6607	28,8742	357,9542
<b>73</b>	8,2675	1,6187	1,6187	1,6377	0,1007	9,1932	84,5154	28,3374	379,7325
<b>74</b>	-0,3508	-0,0666	-0,0666	0,0423	-0,3663	0,1341	0,0180	29,4214	304,3508

### III.1.3.3. Diagnostics d'influence

Le tableau III.4 condense d'autres mesures d'influence utiles, [III.11] dont l'étude complète la recherche des observations aberrantes.

**Tableau III.04:** Diagnostics d'influence.

	Colonne 1	Colonne 2	Colonne 3	Colonne 4	Colonne 5	Colonne 6	Colonne 7
<i>Observation i</i>	$WSSD_i$	$D_i$	DIFITS	$DFBETAS_{0,1}$	$DFBETAS_{0,2}$	$DFBETAS_{0,3}$	$Covratio_i$
1	418,3525	0,0579	0,4168	0,3726	-0,0053	-0,2777	1,1728
2	211,7425	0,2695	0,9534	0,7283	-0,2120	-0,4513	0,7695
3	150,7769	0,0006	-0,0436	-0,0220	0,0232	0,0066	1,1381
4	33,5899	0,0004	0,0363	0,0095	-0,0235	0,0038	1,0860
5	73,2523	0,0405	-0,3530	-0,2052	0,1349	0,0881	0,9691
6	66,1136	0,0183	0,2351	0,1621	-0,0483	-0,0930	1,0165
7	6,1622	0,0002	-0,0235	0,0013	0,0181	-0,0095	1,0820
8	28,7128	0,0010	-0,0544	-0,0381	-0,0020	0,0259	1,0620
9	21,3864	0,0061	0,1345	0,0349	-0,0797	0,0129	1,0505
10	24,9999	0,0088	-0,1622	-0,0509	0,0900	-0,0063	1,0384
11	7,7977	0,0169	-0,2266	-0,0321	0,1320	-0,0444	0,9779
12	8,5669	0,0015	-0,0674	-0,0426	-0,0270	0,0371	1,0538
13	8,9883	0,0011	0,0559	0,0369	0,0351	-0,0377	1,0673
14	2,5061	0,0000	-0,0021	-0,0006	0,0003	0,0002	1,0589
15	0,7290	0,0018	-0,0738	0,0042	0,0420	-0,0262	1,0529
16	2,8882	0,0053	0,1264	0,0405	-0,0140	-0,0124	1,0115
17	2,8620	0,0100	0,1742	0,0213	-0,0803	0,0311	0,9914
18	2,4056	0,0097	-0,1707	0,0127	0,1195	-0,0690	1,0265

Tableau III.04 (suite)

	Colonne 1	Colonne 2	Colonne 3	Colonne 4	Colonne 5	Colonne 6	Colonne 7
<i>Observation i</i>	<b>WSSD<sub>i</sub></b>	<b>D<sub>i</sub></b>	<b>DIFITS</b>	<b>DFBETAS<sub>0,1</sub></b>	<b>DFBETAS<sub>0,2</sub></b>	<b>DFBETAS<sub>0,3</sub></b>	<b>Covratio<sub>i</sub></b>
<b>19</b>	0,8464	0,0081	-0,1562	-0,0233	-0,0442	0,0175	0,9895
<b>20</b>	3,0354	0,0049	-0,1208	-0,0608	-0,1051	0,0806	1,0948
<b>21</b>	0,6536	0,0002	-0,0253	0,0051	0,0163	-0,0124	1,0670
<b>22</b>	5,6794	0,0013	-0,0611	0,0230	0,0348	-0,0368	1,0604
<b>23</b>	1,5590	0,0023	-0,0831	-0,0413	-0,0580	0,0475	1,0610
<b>24</b>	1,2924	0,0070	-0,1455	0,0101	0,0211	-0,0326	0,9947
<b>25</b>	1,4589	0,0045	-0,1151	-0,0538	-0,0914	0,0687	1,0675
<b>26</b>	0,1165	0,0008	-0,0472	-0,0043	0,0012	-0,0026	1,0505
<b>27</b>	0,1442	0,0007	-0,0464	-0,0144	-0,0185	0,0133	1,0546
<b>28</b>	1,0124	0,0055	-0,1285	0,0056	0,0162	-0,0254	1,0075
<b>29</b>	2,9661	0,0076	0,1498	0,0751	0,1302	-0,0996	1,0884
<b>30</b>	37,4998	0,0001	-0,0158	0,0104	0,0109	-0,0133	1,0973
<b>31</b>	13,1586	0,0040	0,1098	-0,0388	-0,0124	0,0466	1,0332
<b>32</b>	6,7957	0,0006	-0,0407	0,0005	-0,0155	0,0013	1,0587
<b>33</b>	15,5763	0,0016	0,0677	-0,0270	-0,0103	0,0320	1,0529
<b>34</b>	12,8179	0,0045	0,1159	-0,0401	-0,0124	0,0483	1,0295
<b>35</b>	12,4818	0,0050	0,1219	-0,0413	-0,0123	0,0483	1,0295
<b>36</b>	12,4818	0,0050	0,1219	-0,0413	-0,0123	0,0499	1,0257
<b>37</b>	16,8675	0,0037	0,1042	-0,0529	-0,0476	0,0689	1,0492
<b>38</b>	4,1715	0,0062	-0,1360	-0,0550	-0,1275	0,0858	1,1687
<b>39</b>	5,9191	0,0044	-0,1137	-0,0427	-0,1036	0,0673	1,1302

Tableau III.04 (suite)

	Colonne 1	Colonne 2	Colonne 3	Colonne 4	Colonne 5	Colonne 6	Colonne 7
<i>Observation i</i>	WSSD <sub>i</sub>	D <sub>i</sub>	DIFITS	DFBETAS <sub>0,1</sub>	DFBETAS <sub>0,2</sub>	DFBETAS <sub>0,3</sub>	Covratio <sub>i</sub>
40	4,2488	0,0000	0,0038	0,0012	0,0031	0,0018	1,0878
41	2,1629	0,0065	0,1403	0,0051	0,0279	0,0014	1,0017
42	5,3970	0,0000	-0,0002	0,0000	-0,0001	0,0000	1,0629
43	5,7709	0,0001	-0,0116	-0,0003	-0,0047	0,0008	1,0626
44	6,7964	0,0159	0,2198	0,0287	0,1382	-0,0556	0,9899
45	7,3151	0,0133	0,2009	0,0231	0,1244	-0,0476	1,0018
46	28,2717	0,0005	-0,0401	0,0248	0,0272	-0,0324	1,0875
47	69,3031	0,0000	-0,0079	0,0059	0,0048	-0,0069	1,1072
48	107,1532	0,0034	-0,0996	0,0797	0,0525	-0,0880	1,1118
49	294,9627	0,5288	-1,3760	1,1785	0,2908	-1,0970	0,6635
50	294,9627	0,5501	1,4095	-1,2073	-0,2978	1,1237	0,6467
51	131,4060	0,0254	0,2755	-0,0842	0,1665	0,0171	1,1009
52	124,4626	0,0028	-0,0914	-0,0609	0,0298	0,0318	1,1025
53	50,7295	0,0351	0,3278	0,1392	-0,1724	-0,0229	0,9759
54	10,0055	0,0000	-0,0100	-0,0067	-0,0063	0,0068	1,0734
55	3,4295	0,0091	-0,1652	0,0033	0,1113	-0,0584	1,0261
56	1,0280	0,0017	-0,0706	-0,0328	-0,0481	0,0379	1,0622
57	0,6289	0,0017	-0,0715	-0,0300	-0,0471	0,0351	1,0596
58	9,3000	0,0001	-0,0197	0,0076	0,0073	-0,0107	1,0630
59	6,5248	0,0008	-0,0473	-0,0197	-0,0444	0,0303	1,1786
60	4,6200	0,0001	-0,0127	-0,0045	-0,0111	0,0070	1,1099

Tableau III.04 (suite et fin )

	Colonne 1	Colonne 2	Colonne 3	Colonne 4	Colonne 5	Colonne 6	Colonne 7
<i>Observation i</i>	WSSD <sub>i</sub>	D <sub>i</sub>	DIFITS	DFBETAS <sub>0,1</sub>	DFBETAS <sub>0,2</sub>	DFBETAS <sub>0,3</sub>	Covratio <sub>i</sub>
<b>61</b>	6,8688	0,0065	-0,1394	-0,0320	-0,1078	0,0559	1,0614
<b>62</b>	6,5601	0,0004	-0,0343	0,0002	-0,0132	0,0014	1,0599
<b>63</b>	246,1438	0,0226	-0,2595	-0,1882	0,0754	0,1093	1,1293
<b>64</b>	23,3152	0,0002	-0,0215	-0,0138	0,0006	0,0087	1,0649
<b>65</b>	14,5763	0,0065	-0,1390	-0,0436	0,0647	-0,0024	1,0310
<b>66</b>	34,5432	0,0112	-0,1833	-0,1374	-0,0175	0,0992	1,0118
<b>67</b>	2,0327	0,0017	-0,0716	0,0233	0,0570	-0,0448	1,0782
<b>68</b>	15,2676	0,0056	0,1287	0,0927	0,0851	-0,0954	1,0565
<b>69</b>	6,7920	0,0004	-0,0329	-0,0196	-0,0122	0,0167	1,0602
<b>70</b>	3,8130	0,0001	-0,0194	-0,0099	-0,0058	0,0079	1,0596
<b>71</b>	20,0297	0,0054	0,1271	-0,0721	-0,0837	0,0977	1,0615
<b>72</b>	149,9057	0,0371	-0,3344	0,2788	0,1484	-0,2926	1,0690
<b>73</b>	244,1858	0,0978	0,5480	-0,4703	-0,1563	0,4542	1,0365
<b>74</b>	14,0004	0,0001	-0,0139	0,0073	0,0105	-0,0106	1,0894

Les résultats obtenus sont condensés dans le tableau III.4, dont les colonnes sont numérotées de (1) à (7). Les valeurs de la somme pondérée des carrés des distance du point  $i$  au barycentre des données [III.4] (WSSD<sub>i</sub>: Weighted of Squared Distance of the center of data), données dans la colonne (1) du tableau III.4 montrent la présence de 6 observations parmi 92 composés sont les plus éloignées. L'observation (1) qui désigne le propane est le composé le plus éloigné avec une valeur de WSSD= 418,3524.

La colonne (2) rassemble les valeurs de la distance de Cook,  $D_i$ . Les composés indexés par les numéros (1, 2, 49, 50,73) dont les noms sont respectivement : propane, butane, tétradécane, hexadécane, pentadécane, ont des valeurs de  $D_i$  supérieures à la valeur critique ( $4/74=0,0541$ ). Ces observations sont considérées comme très influentes.

La colonne (3) rassemble les valeurs de la statistique DIFITS, qui permet de mesurer l'influence d'une observation  $i$  sur la valeur ajustée ou prédite. Cinq observations (1, 2, 49, 50, 73) ont des valeurs absolues de DIFITS, supérieures à la valeur critique  $\sqrt{3/74} = 402774$ . Ces observations ont une influence sur la valeur prédite.

Les colonnes (4), (5) et (6) rassemblent les valeurs de la statistique  $DFBETAS_{j,i}$ . L'examen des colonnes  $DFBETAS_{j,i}$  en tenant compte de la valeur critique

$$\frac{2}{\sqrt{n}} = 2/\sqrt{74} = 0,2325$$

fait ressortir les points influents 1, 2, 49, 50, 72, 73.

La colonne (7) rassemble les valeurs du  $COVRATIO_i$  où toutes les valeurs sont supérieures à 1. Ces observations améliorent la précision de l'estimation à l'exception les observations 2, 49, et 50 qui ont des valeurs inférieures à 1 et détériorent la précision de l'estimation.

### III.1.3.4. Domaine d'applicabilité

Le domaine d'application a été vérifié à l'aide du diagramme de Williams. La figure III.3 fait ressortir les points aberrants et les points influents du modèle obtenu.

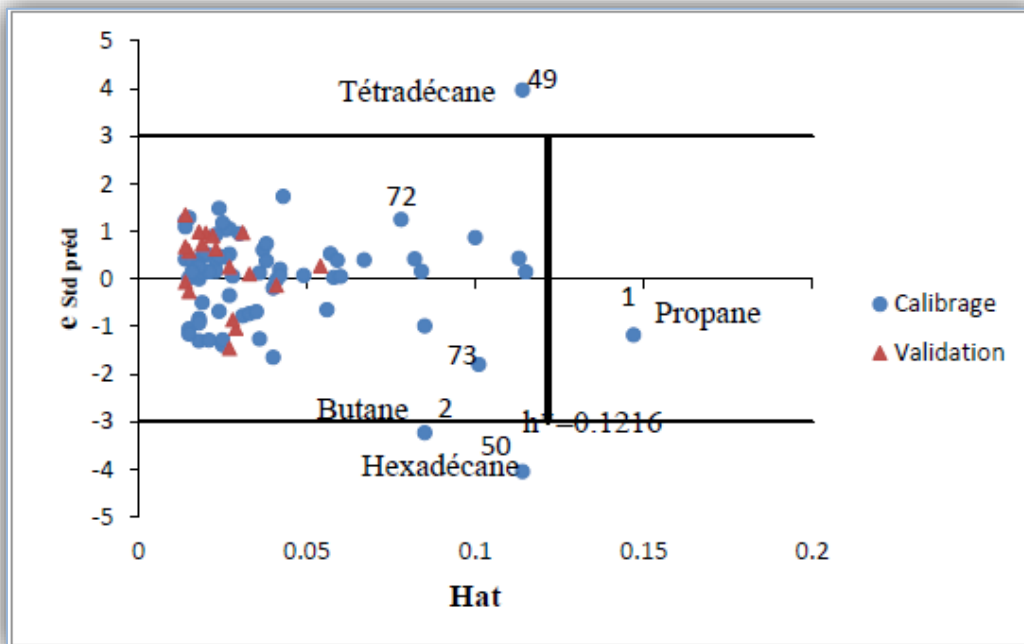


Figure III.3: Diagramme de Williams

On note 3 points aberrants (Butane, Tétradécane et Héxadécane) appartenant à l'ensemble de calibrage avec des valeurs des résidus standardisées hors de l'intervalle  $[-3, +3]$  unité et un seul point influent (Propane) de l'ensemble de calibrage avec une valeur de  $h^* > h_i$ , avec  $h^* = 3(m+1)/n_{tr} = 3(2+1)/74 = 0,1216$  et  $h_i_{Propane} > 0,1216$ .

### III.1.3.5. Test de randomisation

Le test de randomisation est représenté par la figure III.4 suivante:

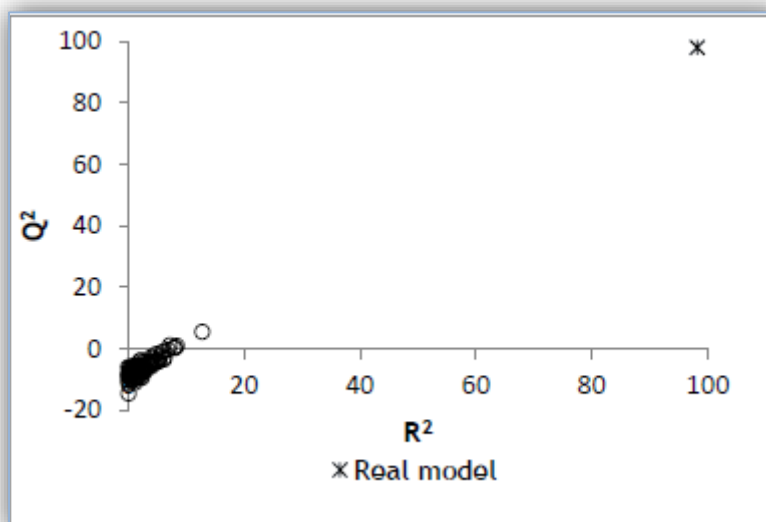


Figure III.4: Test de randomisation

Il est visiblement clair que les statistiques obtenues des 100 vecteurs modifiés du point d'éclair Tec sont plus petites que celles de modèle réel symbolisé par un astérisque (\*). Les valeurs de  $Q^2$  sont toutes inférieures à 5,49 % et la majorité des valeurs sont négatives.

En conclusion, une relation QSPR a été établie et le modèle obtenu n'est pas dû au hasard. On conclut que le modèle établi n'est pas dû au hasard.

### III.1.3.6. Validation statistique externe

La valeur élevée du coefficient de prédiction externe ( $Q_{ext}^2 = 98,88\%$ ), illustre la fiabilité du modèle obtenu. Ainsi que la valeur de la somme des erreurs de prédiction externe ( $SDEP_{ext} = 4,24$ ) est proche de celle pour la prédiction de l'ensemble de calibrage ( $SDEP = 5,67$ ).

Les conditions générales pour un pouvoir prédictif du modèle réel sont vérifiées sur l'ensemble de test soient :

$$R_{CV_{ext}}^2 = 0,9661 > 0,5 \quad ; \quad r^2 = 0,9685 > 0,6 \quad ; \quad r_0^2 = 0,9963$$

$$r_0'^2 = 0,9969 \quad ; \quad T1 = -0,0286 < 0,1 \quad ; \quad T2 = -0,0293 < 0,1$$

$$0,85 < k = 0,9954 < 1,15 ; 0,85 < k' = 1,0044 < 1,5$$

$$|r^2 - r_0'^2| = 0,0284 < 0,3$$

- Nous avons également modélisé l'ensemble des données, les paramètres caractéristiques sont condensés dans le tableau III.5.

**Tableau III.5:** Valeurs des paramètres statistiques pour tout l'ensemble étudié.

R2	Q2	R2adj	SDEC	SDEP	Kx	Kxy	F	SE
98,14	97,91	98,1	5,081	5,394	70,06	82,31	2352,16	5,1655

L'étude du modèle fait sortir cinq points influents dont deux points sont à la fois des points aberrants, ainsi que la présence d'un autre point aberrant.

- Un travail inverse a été réalisé en utilisant 18 composés pour le calcul et 74 composés pour sa validation statistique externe.

Les résultats représentés dans le tableau III.6 montrent une légère baisse des valeurs de R<sup>2</sup>, Q<sup>2</sup> par rapport à celles du modèle étudié et à celles du modèle calculé sur l'ensemble des donnés.

**Tableau III.6:** Valeurs des paramètres statistiques pour le modèle inverse.

R <sup>2</sup>	Q <sup>2</sup>	Q <sup>2</sup> <sub>boot</sub>	R <sup>2</sup>	SDEC	SDEP
96,85	95,47	93,8	96,43	3,949	4,738
Kx	Kxy	Kxy-Kx		F	SE
66,19	80,18	13,99		230,87	4,3261

Remarquons que le nombre des points influents obtenus par la validation croisée est de 15. Cette valeur est très élevée en comparaison de celles des deux modèles précédents (le modèle obtenu au début et le modèle calculé sur la totalité des données).

### III.1.4. Conclusion

La modélisation des points d'éclair d'un ensemble de n-alcanes, a conduits à un bon modèle dont les paramètres statistiques sont élevés. Notons aussi que la bonne qualité de l'ajustement, la robustesse, la bonne capacité prédictive et la fiabilité du modèle ont été vérifiées par les graphes et les tests réalisés. L'analyse des résidus appliquée sur l'ensemble de calibrage fait ressortir un seul composé (le propane) de l'ensemble de calibrage avec un bras de levier important ( $h_i > h^*$ ), et 3 points (le butane, l'hexane et le tétradécane) de l'ensemble de calibrage aussi avec des erreurs standardisées hors les limites  $\pm 3$ . Ces résultats sont confirmés graphiquement par diagramme de Williams.

# CONCLUSION GÉNÉRALE

### Conclusion générale

Les produits chimiques tout au long de leur production, manipulation, transport et utilisation représentent un véritable danger pour la santé et l'environnement. Les gens de tous âges, des plus jeunes au plus âgés, parlant différentes langues et utilisant différents alphabets, de conditions sociales très différentes, éventuellement illettrés, sont chaque jour confrontés aux produits dangereux.

Face à ce danger, et étant donné l'importance du commerce mondial des produits chimiques et la nécessité de mettre au point des programmes nationaux pour assurer l'utilisation, le transport et l'élimination de ces produits en toute sécurité.

Le SGH (Système Générale Harmonisé) décrit la classification des produits chimiques par types de danger et propose des éléments de communication correspondant à ces dangers, y compris des étiquettes et des fiches de données de sécurité.

Les dangers les plus évidents sont les dangers d'incendie ou d'explosion. Plusieurs critères techniques et méthodes d'essai spécifiques pour l'identification des liquides inflammables et combustibles.

Dans le cas des liquides inflammables, la classification s'effectue en fonction de la valeur du point d'éclair. Cette propriété a fait le sujet de notre travail de recherche .

Le point d'éclair d'un liquide est la température la plus basse à laquelle ce liquide libère assez de vapeur pour s'enflammer (commencer à brûler) à la surface de ce liquide. On trouve parfois plusieurs valeurs de point d'éclair pour un produit chimique donné selon la méthode de mesure utilisée.

Nous avons utilisé la méthodologie QSPR pour relier cette propriétés des n-alcane à des descripteurs moléculaires théoriques, calculés à l'aide de logiciels spécialisés du commerce.

**RÉFÉRENCES  
BIBLIOGRAPHIQUES**

## RÉFÉRENCES BIBLIOGRAPHIQUES

---

- [1] Gramatica, P., Applicatione delle metodologie QSAR a problematiche ambientali di inquinanti organici, Università degli Studi di Bologna - Dottorato in Chimica Ind –2004.
- [2] Journal officiel de l'Union européenne, REGLEMENT (CE) N° 1272/2008 DU PARLEMENT EUROPÉEN ET DU CONSEIL du 16 décembre 2008.
- [3] JOURNAL OFFICIEL DE LA REPUBLIQUE ALGERIENNE N° 07. 28 Rabie Ethani 1437. 7 février 2016. pp21-22 [4] Aurélie, G.S.A., Thèse de doctorat, Une nouvelle méthode d'apprentissage de données structurées : applications à l'aide à la découverte de médicaments, Université de Paris 6, 2008.
- [I.1] INRS (Institut national de la recherche scientifique). Ventilation des ateliers d'encollage de petits objets (chaussures). Guide pratique de ventilation n° 5. ED 672, 1987 : 28 .
- [I.2] Durrans, T. H., (1938), Solvents Ed4., P238. Van Nostrand (New-York).
- [I.3] Guinot, H., (1948), solvants et plastifiants ouvrage de le collection « Matériaux et synthèse ».
- [I.4] Cohr, K.H., (1985), "Definition and Practical Limitation of the Concept of OrganicSolvents ". In: Chronic Effects of Organic Solvents on the Central Nervous System and Diagnostics Criteria. World Health Organization - Regional Office for Europe, Copenhagen, pp. 43-55.
- [I.5] Environnement Canada, (1992). PCE-12-89 : Peintures : Peintures à base de solvant. Environnement Canada; Programme Choix environnemental, Ottawa.
- [I.6] De Lanty, P., (2005), " Paramètres de solubilité ", OCL (Oilseeds and fats, Crops and Lipids). 12 (4). pp. 299-301.
- [I.7] INERIS (Institut National de l'Environnement Industriel et des RISques), DRA-09- 103185-12091D, 22 décembre 2009.
- [I.8] CUSSTR (Commission Universitaire de Sécurité et Santé au Travail Romande), Danger incendie, Version 1, 2005. p. 3.
- [I.9] Florian, M., Benoit, S., (2013), " Point des connaissances ", INRS, ED 5001, 2013 : pp.1-2.
- [I.10] Thomson, N, J., (1929), Auto-Ignition Température of Flammable Liquids,(1929), Ind. Eng.Chem.21(2),pp 133-139.
- [I.11] Centre d'expertise en analyse environnementale du Québec. Détermination de la température du point d'éclair selon la technique Pensky-Martens (vase clos). MA. 108 – P.E.
- 1.1, Ministère du Développement durable, de l'Environnement, de la Faune et des Parcs du Québec, 2012, p 10.
- [I.12] Règlement (CE) n° 1272/2008 du Parlement européen et du Conseil du 16 décembre 2008 relatif à la classification, à l'étiquetage et à l'emballage des substances et des mélanges, modifiant et abrogeant les directives 67/548/CEE et 1999/45/CE et modifiant le règlement (CE) n° 1907/2006,

## RÉFÉRENCES BIBLIOGRAPHIQUES

---

- Annexe I : Prescriptions relatives à la classification et à l'étiquetage des substances et mélanges dangereux, Partie 2 : Dangers physiques, Paragraphe 2.6 : Liquides inflammables.
- [I.13] Boust, C., (2011), Fiche solvants " les solvants particuliers", INRS, ED 4229/Ed4224.
- [I.14] Palmade-le Dantec, N., Picot, A., " La prévention du risque : le remplacement des solvants les plus toxiques par des solvants moins toxiques ". Actes du colloque " Conservation-Restauration et Sécurité des personnes ", 3-5 février 2010, Draguignan; ISB 978-2-9531978-1-5.
- [I.15] Outemzabt, M., Article publié dans le Matin d'Algérie, le 12 mai 2017.
- [I.16] Guerzi, C., mémoire de Magister, Scénarii d'Incendie –Explosion au niveau du complexe pétrochimique de SKIKDA, Etude de cas: Unité Ethylène, Université d'Oran, 2011.
- [I.17] Séminaire retour d'expérience – IMPEL (European Union Network for the Implementation and Enforcement of Environmental Law) / Inspecteurs des installations classées– Bordeaux, N°ARIA 19979, juin 2002.
  
- [II.1] Pople, J. A., Beveridge, D. L., Approximate Molecular Theory, Mc Graw- Hill (196).
- [II.2] Hartree, D.R., (1928), "The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods". Proc. Cambridge. Phil. Soc.24, 328.
- [II.3] Fock, V., (1930), "Näherungsmethode zur lösung des quantenmechanischen mehrkörper problems Z. Physik.61, 126 .
- [II.4] Slater, J. C., Phys. Rev. The self consistent field and the structure of atoms, 1928. 32, 339 ; atomic shielding constants. 1930.35, 1210.
- [II.5] Löwdin, P. O., (1950), "On the Non Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals". J. Chem. Phys, 18, 365.
- [II.6] Pariser, R., Parr, R. G., (1953), "A Semi Empirical Theory of the Electronic Spectra and Electronic Structure of Complex Unsaturated Molecules .I", J. Chem. Phys. 21, 466.
- [II.7] Pariser, R., Parr, R. G., (1953), "A Semi Empirical Theory of the Electronic Spectra and Electronic Structure of Complex Unsaturated Molecules. II", J. Chem. Phys. 21, 767.
- [II.8] Pople, J. A., (1953), Electron interaction in unsaturated hydrocarbons, Trans. Faraday Soc. 49, 1375.
- [II.9] Pople, J. A., Santry, D. P., Segal, G. A., (1965), "Approximate Self Consistent Molecular Orbital Theory. I. Invariant Procedures", J. Chem. Phys., 43, S129-S135.
- [II.10] Pople, J.A., Beveridge, D. L., Dobosh, P. A., (1967), "Approximate SelfConsistent Molecular Orbital Theory. V. Intermediate Neglect of Differential Overlap". J. Chem. Phys, 47, 2026.

## RÉFÉRENCES BIBLIOGRAPHIQUES

---

- [II.11] Dewar, M. J. S., Thiel, W., (1977), " Ground states of molecules. 38. The MNDO method. Approximations and parameters", J. Am. Chem. Soc., 1977, 99, 4899- 4907, 4907- 4914.
- [II.12] Thiel, W., Encyclopedia of Computational Chemistry, Vol. 3, P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III, P. R. Schreiner (Eds.), Wiley, Chichester, 1998, p. 1599.
- [II.13] Burstein, K. Y., Isaev, A.N., Theor. Chim. Acta 1984, 64, 397- 401.
- [II.14] Dewar, M. J. S., Zoebisch, E. G., Healy, E. F., Stewart, J. J. P., J. Am. Chem. Soc. 1985, 107, 3902- 3909 ; AM1
- [II.15] Holder, A. J., AM1, in Encyclopedia of Computational Chemistry. Vol.1 (Eds.: P. V. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer, III, P. R. Schreiner), Wiley, Chichester, 1998, pp. 8-11.
- [II.16] Stewart, J. J. P., J. Comput. Chem., 10, 1989, 209- 220 ; 221- 264; PM3 :
- [II.17] Stewart, J. J. P., Encyclopedia of Computational Chemistry; Vol. 3, Schleyer, P. V. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer III, H. F., Schreiner, P. R., (Eds.), Wiley, Chichester, 1998, pp. 2080-2086.
- [II.18] Jensen, F., Introduction to Computational Chemistry, Wiley, pp. 94- 96.
- [II.19] Stewart, J. J. P., Application of Localized Molecular Orbitals to the Solution of Semiempirical Self-Consistent Field Equations, Int. J. Quantum. Chem. 1996. 58 , 133.(page
- [II.20] Daniels, A. D., Millam, J. M., Scuseria, G. E., J. Chem. Phys., 1997.107 , 425.
- [II.21] Ramachandran, K. I., Deepa, G., Namboori, K., "Computational Chemistry and Molecular Modeling, Principles and Applications ", Springer-Verlag Berlin Heidelberg, 2008.
- [II.22] Coulson, C. A., Longuet-Higgins, H. C., "The Electronic Structure of Conjugated Systems". I. R Proc. Roy. Soc. (London) A .1947.191, 39 .
- [II.23] Mulliken, R. S., (1926), "Criteria for the Construction of Good Self Consistent Field Molecular Orbital Wave Functions, and the Significance of LCAOMO Population Analysis" , J. Chem. Phys., 36, 3428.
- [II.24] Kutzelnigg, W., Del Re, G., Berthier, G., (1971), "  $\pi$  and  $\sigma$  Electrons in Theoretical Organic Chemistry", Springer Verlag, Berlin.
- [II.25] Pullman, B., (1969), "La Biochimie Electronique", Collection Que sais-je ? PUF, n°1075, Deuxième édition, Paris.
- [II.26] Boyd, D. B., Lipkowitz, K. B., eds. Reviews in Computational Chemistry, History of the Gordon Conferences on Computational Chemistry, Wiley- VCH, New York, 2000.399- 439.
- [II.27] Allinger, N. L., Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms, J. Am. Chem. Soc., 1977, 99, 8127.

## RÉFÉRENCES BIBLIOGRAPHIQUES

---

- [II.28] Burkert, U., Allinger, N. L., *Molecular Mechanics*, ACS Monograph No. 177, American Chemical Society, Washington, DC, 1982, 1986.
- [II.29] Allinger, N. L., Yuck, Y. H., Lii, J. H., *Molecular Mechanics. The MM3 Force Field for Hydrocarbon* 3. 1, *J. Am. Chem. Soc.*, 1989, 111, 8551.
- [II.30] Allinger, N. L., Chem, K., Katzenellbogen, J. A., Wilson, S. R., Anstead, G. M., (1996), *Hyperconjugative Effects on Carbon-Carbon Bond Lengths in Molecular Mechanics (MM4)*, *J. Comput. Chem.*, 17, 747.
- [II.31] MacKerell, A. D., Jr., Bashford, D., Bellott, M., Dumbrack, R. L., Evaseck, J. D., Field, M. J., Fischer, S., Gao, J., Gao, H., He, S., Joseph-MacCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michmick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E. III., Roux, B., Schlemkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., Karplus, M., *J. Phys. Chem. B*, 1998, 102, 3586-3616.
- [II.32] MacKerell, A. D. Jr., Wiorkiewicz-Kuczera, J. and Karplus, M., (1995), "An all-atom empirical energy function for the simulation of nucleic acids", *Journal of the American Chemical Society*, 117 (48), 11946-11975
- [II.33] Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States David, J., Swaminathan, S., Karplus, M., *CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations* *J. Comput. Chem.* 1983.4, 187,
- [II.34] Momany, F.A., Rone, R., (1992), "Validation of the General Purpose QUANTA 3.2/CHARMm Force Field". *J. Comput. Chem.*, 13 (7):888-900.
- [II.35] Halgren, T. A., (1996d), "Merck Molecular Force Field: I. Basis, Form, Scope, Parameterization and Performance of MMFF94", *J. Comp. Chem.* 17, 490-519.
- [II.36] Halgren, T. A., (1996a), "Merck Molecular Force Field. II. MMFF94 van der Waals and Electrostatic Parameters for Intermolecular Interactions", *J. Comp. Chem.* 17, 520-552.
- [II.37] Halgren, T. A. (1996b), "Merck Molecular Force Field. III. Molecular Geometries and Vibrational Frequencies", *J. Comp. Chem.* 17, 553-586.
- [II.38] Halgren, T. A., (1996c), "Merck Molecular Force Field. V. Extension of MMFF94 Using Experimental Data, Additional Computational Data and Empirical Rules", *J. Comp. Chem.* 17, 616-641.
- [II.39] Halgren, T. A., Nachbar, R. B., (1996). "Merck Molecular Force Field. IV. Conformational Energies and Geometries for MMFF94", *J. Comp. Chem.* 17, 587-615.
- [II.40] HyperChemTM. 2000. Release 6,03 for windows. Molecular Modeling System.
- [II.41]. Todeschini, R., Consonni, V., Mauri, A. and Pavan, M. (2006), "DRAGON Software for the calculation of molecular descriptors". Release 5.3 for Windows, Milano. Italy.

## RÉFÉRENCES BIBLIOGRAPHIQUES

---

- [II.42] Donald, E. K., (1997), "The art of computer programming", Vol,2 (3rd ed), Boston: Addison Wesley.
- [II.43] Tourassi, G.A., Frederick, E. D., Markey, M. K., Floyd, C.E., (2001), "Application of the mutual information criterion for feature selection in computer-aided diagnosis", *Medical Physics*, Vol, 28(12), pp, 2394-2402.
- [II.44] Kennard, R., Stone, L. A., (1969), "Computer aided design of experiments *Technometrics*", Vol. 11 No. 1, pp. 137-148.
- [II.45] Todeshini, R., Ballabio, D., Consonni, V., Mauri, A., Pavan, M., (2009), *MOBY DIGS Software for Multilinear Regression Analysis and Variable Subset Selection by Genetic Algorithm*, Release 1,1 for windows, Milano.
- [II.46] Snedecor, G. W., Cochran, W.G., (1967), "Statistical methods", 6th Edition, Oxford and IBH Publishing Co., Bombay/ New Delhi.
- [II.47] Vapnik, V.,(1995), " The nature of Statistical Learning Theory" , Springer-Verlag, New York.
- [II.48] Vladimir, N., Vapnik, V., (1995), "The nature of statistical learning theory", Springer- Verlag, New York, USA.
- [II.49] Smola Alex, J., Bernhard Schölkop, A.,(2004), "Tutorial on support vector regression", *Statistics and Computing*, Vol 14(3), pp, 199-222.
- [II.50] Cristianini, N., Shawe-Taylor, J., "An Introduction to Support Vector Machines and other Kernel Based Learning Methods", Cambridge University, Press, 2000.
- [II.51] Bishop, C.M., (1995), "Neural Networks for Pattern Recognition", Oxford University Press.
- [II.52] Leardi, R., Boggia, R., Terrile, M., (1999), "Genetic algorithms as a strategy for feature selection", *J, Chemom*, Vol, 6, pp, 267- 281.
- [II.53] Eriksson, L., Jaworska, J., Worth, A. P., Cronin, M. T. D., Mc Dowell, R. M., Gramatica, P., (2003), "Methods for Reliability and Uncertainty Assessment and for Applicability Evaluations of Classification and Regression Based QSARs", *Health Perspectives*, Vol. 111 No. 10, pp.1361-1375.
- [II.54] Tropsha, A., Gramatica, P., Grombar, V. K., (2003), "The importance of being earnest validation in the absolute essential for successful application and interpretation of QSPR models", *QSAR and Combinatorial Science*, Vol 22 No.1, pp. 69-76.
- [II.55] Efron, B., (1994), "The jackknife, the Bootstrap and Other Resampling Planes", Society for Industrial and Applied Mathematics, Philadelphia, PA.
- [II.56] Shi, L.M., Fang, H., Tong, W., Wu, J., Perkins, R., Blair, R.M., Branham, W.S., Dial, S.L., Moland, C.L., Sheehan, D.M., (2001), " QSAR models using a large diverse set of estrogens", *Journal of Chemical Information and Computer Science*, Vol. 41, pp. 186-195.

## RÉFÉRENCES BIBLIOGRAPHIQUES

---

- [II.57] Golbraikh, A., Tropsha, A., (2002), "Beware of  $q^2$ !", Journal of Molecular Graphics and modelling, Vol. 20 No 4, pp. 269-276.
- [II.58] MATLAB, Version 7.8.0.347 (R2009a), The Language of Technical Computing, The Math Works, February 12 (2009).
- [II.59] Montgomery, D.C., Peck, E.A., (1992) Introduction to Linear Regression Analysis, Second Edition, Wiley-Interscience, New York, pp.527,
- [II.60] Weisberg, S. (2005), "Applied Linear Regression", 3rd edn, John Wiley and sons, Inc, New Jersey.
- [II.61] SCAN - Software for Chemometric Analysis, (1995), Version1.1 - for Windows, Minitab USA.
  
- [III.1] Wang, K. Q., (1999), A new method for predicting the densities of alkanes from the information of molecular structure-group bond contribution method, Chin. J. Org. Chem, 19: 04 - 308.
- [III.2] Pan, Y., Jiang, J., Wang, Z., (2007), Prediction of the flash points of alkanes by group bond contribution method using artificial neural networks, Front. Chem. Eng. China, 1(4), 390– 394.
- [III.3] Mathieu, D., (2010), Inductive modeling of physico-chemical properties: Flash point of alkanes. Journal of Hazardous Materials, 179: 1161–1164.
- [III.4] Todeschini, R., Consonni, V., Mauri, A., and Pavan, M., DRAGON Software for the Calculation of Molecular Descriptor. Version. 5.3 for Windows, Talete S. r. l., Milan. Italy. 2005.
- [III.5] Todeschini, R., Ballabio, D., Consonni, V., Mauri, A., and Pavan, M., MOBYDIGS Software for Multilinear Regression Analysis and Variable Subset Selection by Genetic Algorithm. Release 1.1 for windows, Milano. 2009.
- [III.6] Todeschini, R., Maiocchi, A., Consonni, V., (1999), Chemom. Intell. Lab. Syst. 46,13-29.
- [III.7] Platt, J. R., (1947), influence of the neighbor bonds on additive bond properties in paraffins, J. Chem. Phys. 15, 419-420.
- [III.8] Randic, M., (1995), Molecular shape profile, J. Chem. Inf. Comput. Sci., 35, 373-382.
- [III.9] Randic, M., Razinger, M. (1995), On Characterization of Molecular Shapes J. Chem. Inf. Comput., Sci, 35, 594-606.
- [III.10] Randic, M., (1995), Molecular Profiles Novel Geometry-Dependent Molecular Descriptors, New. J. Chem., 19, 781-791.
- [III.11] Montgomery, D.C., Peck, E.A., 1992. Introduction to Linear Regression Analysis, Second Edition, Wiley-Interscience, New York, pp.527

## RÉFÉRENCES BIBLIOGRAPHIQUES

---

- [III.12] Weisberg, S., (2005), Applied Linear Regression, 3rd edn, John Wiley and sons, Inc, New Jersey.

