

الجمهورية الجزائرية الديمقراطية الشعبية  
وزارة التعليم العالي والبحث العلمي

UNIVERSITE BADJI MOKHTAR - ANNABA  
BADJI MOKHTAR – ANNABA UNIVERSITY



جامعة باجي مختار – عنابة

Faculté : .....TECHNOLOGIE.....

Département : .....GENIE DES PROCEDES.....

Domaine : SCIENCES ET TECHNOLOGIES

Filière : .GENIE DES PROCEDES

Spécialité : Génie Des Procédés De L'Environnement

## Mémoire

Présenté en vue de l'obtention du Diplôme de Master

Thème:

**Développement des modèles d'intelligence artificielle de type RQSP-RNA pour estimer les propriétés critiques des hydrocarbures purs**

Présenté par : *AOUADI Farouk*

*HARIKI Yanisse*

Encadrant : *Roubehie Fissa Mohamed* Docteur Université BADJI MOKHTAR-ANNABA

### Jury de Soutenance :

LAKROUT	MAB	UBMA	Président
DERADJI	Professeur	UBMA	Examineur
ROUBEHIE FISSA	MAB	UBMA	Encadrant

Année Universitaire : 2023/2024

---

الحمد لله كما ينبغي لجلال  
وجهه وعظيم سلطانه

---



## *Dédicace*

*Après de nombreuses années d'études, et connaissant l'importance d'aider ceux qui m'aiment, je voudrais humblement leur dédier cet humble acte sachant qu'ils méritent le meilleur.*

*Je dédie ce travail :*

*À mes premiers professeurs de la vie, ma mère attentionnée et mon cher père, qui m'ont comblé d'amour et d'affection, et qui m'ont toujours encouragé et soutenu tout au long de mes années scolaires et dans les moments les plus difficiles.*

*À mes chers frères.*

*À mes chères sœurs,*

*À toute ma famille de ma mère et de mon père.*

*À mon compagnon Yanisse et sa famille.*

*À tous mes chers amis.*

*À ma promotion de 2023/2024 sans exception.*

*À tous ceux qui m'ont aidé de près ou de loin ne serait-ce qu'un mot d'encouragement et de gentillesse.*

*AOUADI Farouk*



## *Dédicace*

*Je dédie ce modeste travail à : À mes chers parents, que  
nulle dédicace ne puisse exprimer mes sincères sentiments,  
Pour leur patience illimitée, leur encouragement continu,  
leur aide, en témoignage de mon profond amour et respect  
pour leurs grand sacrifices*

*À mon frère*

*À mon binôme «Farouk»*

*Et à tous ceux qui ont contribué de près ou de loin pour  
que ce travail soit possible, je vous dis merci*

*HARIKI Yanisse*



## *Remerciements*

*Tout d'abord, nous remercions Dieu qui nous a donné la santé, la patience et les raisons qui nous ont permis d'accomplir ce travail.*

*Un grand merci à notre encadreur Monsieur ROUBEHIE FISSA Mohamed, qui a rendu ce travail plus agréable grâce à sa gentillesse et ses encouragements, ses efforts pour le bon déroulement de ces travaux, et son aide précieuse.*

*Un grand merci aux personnels de Sonatrach Skikda (raffinerie RA1K).*

*Nous tenons à exprimer nos sincères remerciements aux membres du jury de nous avoir honorés de leur présence et d'avoir accepté d'examiner notre humble travail. Nous leur devons nos sincères remerciements.*

*A tous les enseignants du Département génie des procédés, nous vous remercions pour la qualité de l'enseignement qui restera gravé dans nos mémoires.*

*Enfin, nous adressons nos sincères remerciements à toutes les personnes qui ont contribué directement ou indirectement à la préparation de cette mémoire.*



---

**Résumé : Développement des modèles d'intelligence artificielle de type RQSP-RNA pour estimer la propriété critique des hydrocarbures purs.**

---

Le but de cette étude est de développer des modèles mathématiques pour prédire les propriétés critiques des hydrocarbures purs (température, pression et volume critiques) en utilisant un Réseau de Neurone Artificiel-Perceptron Multi Couches (RNA-PMC) basé sur les Relations Quantitatives Structure-Propriétés (RQSP). Trois modèles RQSP-RNA ont été créés, utilisant une base de données de 420, 565 et 506 points pour Tc, Pc et Vc respectivement. Après une réduction statistique des descripteurs moléculaires, l'algorithme BFGS a été appliqué pour entraîner les RNA. Les meilleurs modèles avaient des architectures (20-14-1), (20-28-1) et (20-16-1) pour Tc, Pc et Vc. Les résultats ont montré une précision élevée, confirmée par des coefficients de régression (R) et de détermination (R<sup>2</sup>) élevés, et de faibles erreurs quadratiques moyennes (REQM) et relatives. Comparés à d'autres études, les modèles RQSP-RNA ont démontré de meilleures performances, et cette approche pourrait être utile en ingénierie pétrolière pour prédire précisément les propriétés des hydrocarbures.

**Mots clés :** optimisation, Propriétés critiques, Hydrocarbures, Modélisation, RQSP-RNA.

---

**Abstract : Development of RQSP-RNA type artificial intelligence models to estimate the critical property of pure hydrocarbons.**

---

The aim of this study is to develop mathematical models to predict critical properties of pure hydrocarbons (critical temperature, pressure, and volume) using a Multi-Layer Perceptron Artificial Neural Network (MLP-ANN) based on Quantitative Structure-Property Relationships (QSPR). Three QSPR-ANN models were created using a database of 420, 565, and 506 data points for Tc, Pc, and Vc respectively. After statistical reduction of molecular descriptors, the BFGS algorithm was applied to train the ANN. The best models had architectures (20-14-1), (20-28-1), and (20-16-1) for Tc, Pc, and Vc. The results showed high accuracy, confirmed by high regression (R) and determination (R<sup>2</sup>) coefficients, and low mean square errors (RMSE) and relative errors. Compared to other studies, QSPR-ANN models demonstrated superior performance, and this approach could be useful in petroleum engineering for accurately predicting the properties of hydrocarbons.

**Keywords:** optimization, Critical properties, Hydrocarbons, Modeling, RQSP-RNA.

---

**المخلص : تطوير نماذج الذكاء الاصطناعي من نوع RQSP-RNA لتقدير الخاصية الحرجة للهيدروكربونات النقية.**

---

يهدف هذا البحث إلى تطوير نماذج رياضية للتنبؤ بالخصائص الحرجة للهيدروكربونات النقية (درجة الحرارة الحرجة، الضغط الحرج، والحجم الحرج) باستخدام شبكة عصبية اصطناعية متعددة الطبقات (MLP-ANN) بناءً على علاقات الكم الهيكلية-الخاصية (QSPR). تم إنشاء ثلاثة نماذج QSPR-ANN باستخدام قاعدة بيانات تضم 420 و 565 و 506 نقطة بيانات لـ Tc و Pc و Vc على التوالي. بعد تقليص إحصائي للمؤشرات الجزيئية، تم تطبيق خوارزمية BFGS لتدريب الشبكة العصبية. كانت أفضل النماذج لها هيكل (20-14-1)، (20-28-1)، و (20-16-1) لـ Tc و Pc و Vc. أظهرت النتائج دقة عالية، تم تأكيدها من خلال معاملات الارتباط (R) ومعاملات التحديد (R<sup>2</sup>) العالية، وأخطاء متوسطة مربعة منخفضة (RMSE) وأخطاء نسبية. مقارنة بالدراسات الأخرى، أظهرت نماذج QSPR-ANN أداءً أفضل، ويمكن أن تكون هذه الطريقة مفيدة في الهندسة البترولية للتنبؤ بدقة بخصائص الهيدروكربونات.

**الكلمات المفتاحية :** التحسين، الخواص الحرجة، الهيدروكربونات، النمذجة، RQSP-RNA.

---



## Tableau de matière

Introduction générale.....	Erreur ! Signet non défini.
I.1. Introduction.....	Erreur ! Signet non défini.
I.2. Généralités sur le pétrole bruts et ses différentes coupes : .....	Erreur ! Signet non défini.
I.3. Composition chimique des pétroles bruts et des fractions pétrolières : .....	Erreur ! Signet non défini.
I.3.1. Les hydrocarbures : .....	Erreur ! Signet non défini.
I.3.1.1. Les paraffines : .....	Erreur ! Signet non défini.
I.3.1.2. Les naphtènes : .....	Erreur ! Signet non défini.
I.3.1.3. Les oléfines : .....	Erreur ! Signet non défini.
I.3.1.4. Les aromatiques : .....	Erreur ! Signet non défini.
I.3.1.5. Autres hydrocarbures : .....	Erreur ! Signet non défini.
I.4. Propriétés usuelles des hydrocarbures et des fractions pétrolières : .....	Erreur ! Signet non défini.
I.4.1. Propriétés liées aux fluides.....	Erreur ! Signet non défini.
I.4.1.1. Masse volumique .....	Erreur ! Signet non défini.
I.4.1.2. Densité relative .....	Erreur ! Signet non défini.
I.4.1.3. Gravité de l'API .....	Erreur ! Signet non défini.
I.4.1.4. Viscosité.....	Erreur ! Signet non défini.
I.4.2. Autres propriétés physicochimiques.....	Erreur ! Signet non défini.
I.4.2.1. Poids moléculaire .....	Erreur ! Signet non défini.
I.4.2.2. Température d'ébullition .....	Erreur ! Signet non défini.
I.4.2.3. Pression de vapeur .....	Erreur ! Signet non défini.
I.4.2.4. Point d'aniline.....	Erreur ! Signet non défini.
I.4.3. Propriétés thermodynamiques .....	Erreur ! Signet non défini.
I.4.3.1. Enthalpie standard de formation .....	Erreur ! Signet non défini.
I.4.3.2. Enthalpie de vapeur .....	Erreur ! Signet non défini.
I.4.4. Facteurs et indices spécifiques.....	Erreur ! Signet non défini.
I.4.4.1. Facteur de caractérisation de Watson .....	Erreur ! Signet non défini.
I.4.4.2. Facteur acentrique .....	Erreur ! Signet non défini.
I.4.4.3. Facteur de compressibilité .....	Erreur ! Signet non défini.
I.4.4.4. Indice de réfraction .....	Erreur ! Signet non défini.
I.4.4.5. Indice d'octane .....	Erreur ! Signet non défini.
I.4.5. Propriétés liées à la sécurité et le risque d'explosion .....	Erreur ! Signet non défini.
I.4.5.1. Point éclair.....	Erreur ! Signet non défini.
I.4.5.2. Température d'auto-inflammation .....	Erreur ! Signet non défini.
I.5. Propriétés critiques : .....	Erreur ! Signet non défini.

## *Tableau de matière*

---

---

I.5.1. Définition : .....	Erreur ! Signet non défini.
I.6. Mesures expérimentales : .....	Erreur ! Signet non défini.
I.6. Raffinage du pétrole : .....	Erreur ! Signet non défini.
I.7. Conclusion .....	Erreur ! Signet non défini.
II.1. Introduction.....	Erreur ! Signet non défini.
II.2. Approche relations quantitatives structures propriétés/activités RQSP/RQSA .	<b>Erreur ! Signet non défini.</b>
II.2.1. Domaines d'application et utilité des méthodes RQSP/RQSA .....	Erreur ! Signet non défini.
II.2.2. Principe et méthodologie de base des études RQSP/RQSA.....	Erreur ! Signet non défini.
II.3. Codage et représentation des structures chimiques .....	Erreur ! Signet non défini.
II.3.1.1. Représentations d'une structure par quelques éléments basés sur la théorie des graphes chimiques .....	Erreur ! Signet non défini.
II.4. Notions de base sur les descripteurs.....	Erreur ! Signet non défini.
II.4.1. Définition .....	Erreur ! Signet non défini.
II.4.2. Classification.....	Erreur ! Signet non défini.
II.4.2.1. Descripteurs 0-D.....	Erreur ! Signet non défini.
II.4.2.2. Descripteurs 1-D .....	Erreur ! Signet non défini.
II.4.2.3. Descripteurs 2-D.....	Erreur ! Signet non défini.
II.4.2.4. Descripteurs 3-D .....	Erreur ! Signet non défini.
II.4.2.5. Descripteurs 4-D.....	Erreur ! Signet non défini.
II.4.3. Généralisation des descripteurs.....	Erreur ! Signet non défini.
II.5. Méthodes de sélection des descripteurs .....	Erreur ! Signet non défini.
II.5.1. Sélection objective .....	Erreur ! Signet non défini.
II.5.2. Sélection subjective.....	Erreur ! Signet non défini.
II.6. Méthodes d'apprentissage statistiques .....	Erreur ! Signet non défini.
II.6.1. Régression linéaire multiple .....	Erreur ! Signet non défini.
II.6.2. Réseaux de neurones artificiels.....	Erreur ! Signet non défini.
II.6.2.1. Du neurone biologique au neurone artificiel formel .....	Erreur ! Signet non défini.
II.6.2.2. Classification des réseaux de neurones selon leurs architectures.....	<b>Erreur ! Signet non défini.</b>
II.6.2.3. Classification des réseaux de neurones selon leurs modes d'apprentissage.	<b>Erreur ! Signet non défini.</b>
II.6.2.4. Phénomènes de sous-apprentissage et sur-apprentissage.....	Erreur ! Signet non défini.
II.6.2.5. Algorithme d'apprentissage : .....	Erreur ! Signet non défini.
II.6.2.6. Réseaux de neurones perceptron multicouche : .....	Erreur ! Signet non défini.
II.7. Analyses et validation des modèles .....	Erreur ! Signet non défini.
II.8. Conclusion : .....	Erreur ! Signet non défini.

## *Tableau de matière*

---

---

III.1. Introduction.....	<b>Erreur ! Signet non défini.</b>
III.2. Matériels et méthodes .....	<b>Erreur ! Signet non défini.</b>
III.2.1. Outils informatiques utilisés.....	<b>Erreur ! Signet non défini.</b>
III.2.2. Méthodologie générale de la modélisation .....	<b>Erreur ! Signet non défini.</b>
III.2.2.1. Collection et préparation de la base des données .....	<b>Erreur ! Signet non défini.</b>
III.2.2.2. Méthode de génération de descripteurs moléculaires.....	<b>Erreur ! Signet non défini.</b>
III.2.2.3. Procédure de prétraitement et sélection des descripteurs .....	<b>Erreur ! Signet non défini.</b>
III.2.2.4. Choix des paramètres pour le développement des modèles de type RQSP-RNA.....	<b>Erreur ! Signet non défini.</b>
III.2.2.5. Analyses statistiques et calculs des erreurs des meilleurs modèles développés	<b>Erreur ! Signet non défini.</b>
III.2.2.6. Mesure de la contribution des descripteurs des modèles RQSP-RNA développés	<b>Erreur ! Signet non défini.</b>
III.3. Résultats et discussions .....	<b>Erreur ! Signet non défini.</b>
III.3.1. Évolution du nombre de descripteurs au cours de la procédure de sélection ...	<b>Erreur ! Signet non défini.</b>
III.3.2. Architectures des meilleurs modèles RQSP-RNA obtenus .....	<b>Erreur ! Signet non défini.</b>
III.3.3. Application des modèles obtenus dans la prédiction des propriétés critiques des.....	<b>Erreur ! Signet non défini.</b>
III.3.4. Estimation de la performance des résultats d'estimation issus des modèles RQSP-RNA développés .....	<b>Erreur ! Signet non défini.</b>
III.3.4.1. Comparaison graphique entre les valeurs expérimentales et calculées à partir des modèles RQSP-RNA .....	<b>Erreur ! Signet non défini.</b>
III.3.4.2. Calculs des paramètres de performance et analyse statistique des modèles de type RQSP-RNA obtenus.....	<b>Erreur ! Signet non défini.</b>
III.3.4.3. Analyse de sensibilité des descripteurs dans chaque modèle RQSP-RNA par la méthode de poids .....	<b>Erreur ! Signet non défini.</b>
III.4. Conclusion .....	<b>Erreur ! Signet non défini.</b>
Conclusion générale et perspectives.....	<b>Erreur ! Signet non défini.</b>
Références bibliographiques .....	<b>Erreur ! Signet non défini.</b>

---

---

## Liste des figures

---

<b>Chapitre I. Généralités sur les hydrocarbures et les fractions pétrolières</b>		
<b>Figure I.1</b>	Indices d'octane des différentes familles chimiques d'hydrocarbures	<b>16</b>
<b>Figure I.2</b>	Diagramme de coordonnées (P-T) des trois équilibres d'un corps pur.	<b>18</b>
<b>Figure I.3</b>	Représentation de la chaleur de vaporisation pour un mélange de corps purs.	<b>19</b>

<b>Chapitre II. Techniques de modélisation neuronal et méthodes de RQSPRQSA</b>		
<b>Figure II.1</b>	Principe et méthodologie de base des modèles RQSP/RQSA	<b>25</b>
<b>Figure II.2</b>	Graphe moléculaire ( $G$ ) correspondant à la structure du 2.3.4-triméthylhexane.	<b>28</b>
<b>Figure II.3</b>	Conception générale de neurones biologiques et artificiels.	<b>37</b>
<b>Figure II.4</b>	Schématisation distinctive d'un neurone artificiel formel	<b>38</b>
<b>Figure II.5</b>	Classification des réseaux de neurones artificiels selon les architectures	<b>42</b>
<b>Figure II.6</b>	Phénomènes d'apprentissage optimale, sous-apprentissage et sur-apprentissage.	<b>44</b>
<b>Figure II.7</b>	Structure d'un perceptron multi couches non-bouclé	<b>47</b>

<b>Chapitre III. Modélisation des propriétés des hydrocarbures par RQSP_RNA</b>		
<b>Figure III.1</b>	Capture d'écran des logiciels utilisés dans cette étude : Microsoft Office Excel, ChemDraw Office, MATLAB, OCHEM Dragon V. 7 et, TIBCO Statistica	<b>50</b>
<b>Figure III.2</b>	Organigramme générale de la méthodologie de développement suiviez dans cetravail.	<b>51</b>
<b>Figure III.3</b>	Capture d'écran du fichier texte de la molécule azulène	<b>53</b>
<b>Figure III.4</b>	Représentation graphique de division de la base des données.	<b>57</b>
<b>Figure III.5</b>	Capture d'écran pour le choix des fonctions d'activation sur STATISTICA	<b>58</b>
<b>Figure III.6</b>	Représentation schématique d'un réseau du modèle de type RQSP-RNA.	<b>64</b>
<b>Figure III.7</b>	Comparaison entre les valeurs expérimentales et calculées par le modèle RQSP-RNA de la température critique $T_c$ pour la base des données entier.	<b>73</b>
<b>Figure III.8</b>	Comparaison entre les valeurs expérimentales et calculées par le modèle RQSP-RNA de la pression critique $P_c$ pour la base des données entier.	<b>74</b>
<b>Figure III.9</b>	Comparaison entre les valeurs expérimentales et calculées par le modèle RQSP-RNA du volume critique $V_c$ pour la base des données entier.	<b>75</b>
<b>Figure III.10</b>	Contributions relatives des descripteurs pertinents d'entrée pour le modèle de RQSP-RNA de la température critique $T_c$ .	<b>78</b>
<b>Figure III.11</b>	Contributions relatives des descripteurs pertinents d'entrée pour le modèle de RQSP-RNA de la pression critique $P_c$ .	<b>78</b>
<b>Figure III.12</b>	Contributions relatives des descripteurs pertinents d'entrée pour le modèle de RQSP-RNA du volume critique $V_c$ .	<b>79</b>

---

---

## Liste des tableaux

---

### Chapitre I. Généralités sur les hydrocarbures et les fractions pétrolières

<b>Tableau I.1</b>	Classification des hydrocarbures en fonction de leurs familles (Dupuy 2015).	7
<b>Tableau I.2</b>	Structures de quelques hydrocarbures aromatiques ( Rebas 2016 ).	10
<b>Tableau I.3</b>	Exemples de quelques structures d'autres hydrocarbures insaturés	11

### Chapitre II. Techniques de modélisation neuronal et méthodes de RQSPRQSA

<b>Tableau II.1</b>	Notions correspondant à la théorie des graphes et la théorie des graphes chimiques	27
<b>Tableau II.2</b>	Exemple de calcul des matrices et vecteur basé sur la théorie des graphes chimique correspond à la structure du 2.3.4-triméthylhexane	28
<b>Tableau II.3</b>	Fonctions de transfert (d'activation) usuelles dans logiciel Statistica	39
<b>Tableau II.4</b>	Fonctions de transfert usuelles dans le logiciel Matlab	40

<b>Chapitre III. Modélisation des propriétés des hydrocarbures par RQSP_RNA</b>		
<b>Tableau III.1</b>	Partie d'une 20 molécule de la base des données utilisées dans ce travail	52
<b>Tableau III.2</b>	Description quantitatives et qualitatives de la base des données	53
<b>Tableau III.3</b>	Nombres de descripteurs initialement calculés par chaque classe correspondante	54
<b>Tableau III.4</b>	Désignation de grandeurs de performances	58
<b>Tableau III.5</b>	Évolution du nombre de descripteurs après chaque étape de prétraitement pour les trois paramètres étudiés.	61
<b>Tableau III.6</b>	Caractéristique des descripteurs retenus pour les trois paramètres étudiés	62
<b>Tableau III.7</b>	Architectures des meilleurs modèles obtenus pour chaque paramètre	63
<b>Tableau III.8</b>	Fonctions d'activation apparues dans cette étude	64
<b>Tableau III.9</b>	Poids et biais du meilleur modèle RQSP-RNA pour Tc (20-14-1)	67
<b>Tableau III.10</b>	Poids et biais du meilleur modèle RQSP-RNA pour Pc (20-28-1).	68
<b>Tableau III.11</b>	Poids et biais du meilleur modèle RQSP-RNA pour Vc (20-16-1).	70
<b>Tableau III.12</b>	Paramètres statistiques des meilleurs modèles RQSP-RNA pour les trois propriétés.	76

---

# *Introduction Générale*

---

Le pétrole, en tant que source d'énergie non renouvelable, a vu sa demande mondiale augmenter de manière significative au cours des cinquante dernières années (**Apergis 2012 ; Pao 2013 ; Amri 2017 ; Dudley 2018**) en raison de son rôle essentiel en tant que matière première utilisée dans de nombreux secteurs industriels. Il joue un rôle central dans la fabrication d'une variété de produits, notamment les lubrifiants, les polymères, les plastiques, les textiles, les engrais et même les médicaments. (**Guthrie 1960 ; Speight 2015**) En tant que principal concurrent sur le marché de l'énergie, le pétrole demeure la ressource la plus consommée dans le monde. (**Rebas 2006 ; Belghit 2019**)

L'industrie pétrolière est un secteur majeur qui a joué un rôle clé dans le développement économique de nombreux pays à travers le monde. (**Masseron 1990 ; Dudley 2018**) Au sein de notre propre économie, elle constitue la principale source de revenus. Les hydrocarbures et les produits pétroliers figurent parmi les principaux matériaux énergétiques traités dans ce domaine. La production de pétrole naturellement piégé dans les réservoirs rocheux est réalisée de diverses manières en fonction de la profondeur, des caractéristiques du pétrole et de la pression des fluides dans le bassin. (**Wauquier 1998 ; Arab et al. 2015**) Le pétrole brut et ses différentes fractions sont composés d'un mélange complexe de molécules et de structures, dont les principaux constituants sont les hydrocarbures (**Wauquier 1998**).

La compréhension des propriétés physico-chimiques et thermodynamiques revêt une importance cruciale dans divers domaines tels que l'ingénierie, l'industrie pétrochimique, et d'autres secteurs connexes. Ces éléments représentent des variables clés essentielles pour de nombreuses opérations et procédés industriels, allant de la fabrication à la transformation, en passant par le traitement et le stockage des matières énergétiques comme les hydrocarbures (**Chitour 1999 ; Riazi 2005 ; Fahim et al. 2009 ; Rebas 2016**). Ils permettent d'appréhender un grand nombre de phénomènes, offrant ainsi aux industriels et aux scientifiques un aperçu approfondi de la nature des matériaux et des méthodes adaptées pour leur transformation et leur transfert. Ceci, dans le but de minimiser les coûts tout en garantissant des résultats de qualité et une sécurité accrue.

Les caractéristiques d'un mélange pétrolier sont influencées par les propriétés de ses composants individuels. Alors que certaines caractéristiques des mélanges peuvent être facilement mesurées, d'autres, comme les propriétés critiques, sont plus difficiles à évaluer (**Constantinou 1994 ; Riazi 2005 ; Roubehie fissa et al. 2019**). Malgré les avancées dans ce domaine, les données sur ces propriétés restent souvent insuffisantes en raison du coût et de la complexité des tests en laboratoire (**Roubehie fissa et al. 2019**). Dans cette étude, nous avons proposé des modèles mathématiques pour estimer ces propriétés de manière plus efficace.

La modélisation moléculaire est une méthode essentielle pour représenter les structures en deux ou trois dimensions, ainsi que pour assister les chimistes théoriques dans des études approfondies et précises des systèmes moléculaires. Elle permet d'exprimer explicitement ces systèmes à l'aide de modèles mathématiques (**Jensen 2007 ; Engel & Gasteiger 2018**). L'utilisation croissante de méthodes computationnelles d'intelligence artificielle dans la modélisation moléculaire est répandue dans de nombreux domaines, notamment dans l'industrie pétrolière, pour développer des modèles fiables capables de prédire diverses propriétés et activités liées aux phénomènes physiques, chimiques et biologiques (**Minsky 1961 ; Duch et al. 2007 ; Ghahramani 2015 ; Roubehie fissa et al. 2019**). Une méthode clé de cette approche est la modélisation de Relation Quantitative Structure

Propriété/Activités (RQSP/RQSA) (QSPR/QSAR en anglais), qui vise à prédire les propriétés et/ou les activités des systèmes chimiques et/ou biologiques en fonction de leurs structures moléculaires (**Charton 2002**). Ces relations sont généralement établies à l'aide de méthodes statistiques d'apprentissage telles que les Réseaux de Neurone Artificiels (RNAs) (ANN en anglais) (**Hamadache 2014 ; Roubehie fissa et all. 2019**).

L'objectif principal de cette étude est de créer des modèles mathématiques basés sur la structure des molécules d'hydrocarbures purs en utilisant des techniques d'intelligence artificielle. Ces modèles visent à prédire trois propriétés critiques : la température critique, la pression critique et le volume critique. Pour ce faire, trois modèles ont été développés en combinant la Relation Quantitative Structure-Propriété (RQSP) et les Réseaux Neuronaux Artificiels à Perceptron Multi-Couches (RNA-PMC). Ces modèles, connus sous le nom de modèles RQSP-RNA, permettent une approche innovante dans la prédiction de ces propriétés cruciales.

Le développement des modèles RQSP-RNA a suivi plusieurs étapes essentielles, dont la(1) collecte et le traitement des données, (2) le calcul et la sélection des caractéristiques moléculaires, (3) la création et la sélection des modèles RQSP-ANN les plus performants. Pour évaluer l'efficacité et la précision de ces modèles, divers paramètres statistiques tels que le coefficient de détermination ( $R^2$ ), la Racine d'Erreur Quadratique Moyenne (REQM) et l'Erreur Relative Moyenne (ERM) ont été utilisés. Une analyse de la sensibilité des poids a permis d'évaluer l'impact de chaque caractéristique d'entrée sur les modèles RQSP-RNA respectifs. Les résultats ont indiqué que les modèles RQSP-RNA sont capables de prédire avec une précision satisfaisante les propriétés critique  $T_c$ ,  $P_c$  et  $V_c$  des hydrocarbures par rapport aux modèles existants dans la littérature.

Le présent thème s'articule principalement autour de trois chapitres et une conclusion générale :

- **Le premier chapitre:** réalisé une analyse littéraire portant sur le pétrole brut, ses diverses fractions et les hydrocarbures purs. et abordé des concepts fondamentaux liés à ce sujet, tels que la signification, les origines et la composition du pétrole, les critères utilisés pour classer le pétrole brut, les propriétés des hydrocarbures, leurs applications variées, les principaux paramètres industriels.
- **Le deuxième chapitre:** traite de la modélisation, en se concentrant sur les Relations Quantitatives Structure-Propriété/Activités (RQSP/RQSA) et sur les méthodes basées sur les apprentissages statistiques tels que les Réseaux de Neurones Artificiels (RNAs) qui relèvent de l'intelligence artificielle. Une vue d'ensemble est donnée sur ces approches, les types variés disponibles, les éléments essentiels à la construction des modèles et leurs applications, ainsi que sur les descripteurs moléculaires, les outils de calcul, les étapes de sélection et de traitement. En outre, des méthodes statistiques d'analyse et de validation de la modélisation sont abordées de manière générale.
- **Le dernier chapitre:** inclut l'ensemble des résultats de modélisation des trois propriétés principales des hydrocarbures purs, avant, pendant et après leur évolution. Nous avons examiné et discuté en détail les résultats générés par les modèles de type RQSP-ARN développés dans le cadre de ce projet, notamment les calculs de descripteurs moléculaires, l'évolution du nombre de ces descripteurs au cours des étapes de prétraitement et l'analyse qualitative des descripteurs pertinents. De plus, nous avons mis

en œuvre une méthodologie d'ingénierie pour optimiser les meilleurs modèles de type RQSP-ARN obtenus. Une approche basée sur un examen statistique complet des modèles, utilisant différents types d'erreurs et de coefficients statistiques, a été utilisée pour évaluer avec précision chaque modèle lors de l'estimation des propriétés critiques des hydrocarbures. Nous avons également classé et hiérarchisé les descripteurs d'entrée pour chaque modèle grâce à une analyse de sensibilité, révélant leur importance individuelle et collective. Enfin, une comparaison entre nos modèles et des modèles d'études similaires dans la littérature a été réalisée afin d'évaluer l'impact et l'importance de nos travaux dans ce domaine de recherche spécifique.

Nous concluons cette étude en résumant les résultats clés et en fournissant des suggestions pour orienter les futurs travaux de recherche dans ce domaine spécifique. Cette synthèse servira de référence pour les prochaines étapes de développe

---

*Chapitre I*

*Généralités sur les hydrocarbures  
et les fractions pétrolières*

---

## **I.1. Introduction**

Le pétrole, également connu comme l'or noir, est hautement convoité à l'échelle mondiale en raison de son rôle crucial dans le progrès économique, énergétique et technologique des nations productrices (Tugcu, et all. 2012 ; AbdelAal 2013 ; Belaid 2017) C'est la principale matière première des carburants qui alimentent les transports (voitures, camions, avions).

C'est aussi une matière première utilisée par l'industrie de la pétrochimie pour un grand nombre de produits de la vie quotidienne : matières plastiques, peintures, colorants, cosmétiques, etc. le pétrole sert aussi comme combustible dans le chauffage domestique et comme source de chaleur dans l'industrie

L'ingénieur en raffinage et en pétrochimie a toujours eu pour préoccupation la compréhension des propriétés physico-chimiques du pétrole et de ses composants. Pour pouvoir efficacement séparer et transformer les produits pétroliers à chaque étape, de l'extraction dans les puits au stockage des produits finis, il est essentiel de bien comprendre les caractéristiques du pétrole et de ses dérivés (Rebas 2016)

L'objectif de ce chapitre est d'aborder divers concepts et termes de base relatifs à l'industrie pétrolière, qui seront explorés tout au long de cette thèse à travers plusieurs sections décrivant et expliquant les principes fondamentaux des hydrocarbures et du pétrole. Cela inclut l'historique et l'origine du pétrole, la composition et la caractérisation du pétrole brut et des différentes fractions pétrolières, les critères de classification des types de pétrole, les propriétés physico-chimiques et thermodynamiques des hydrocarbures et des fractions pétrolières, ainsi que les diverses méthodes de modélisation pour déterminer ces paramètres.

## **I.2. Généralités sur le pétrole bruts et ses différentes coupes :**

Le pétrole brut est un pétrole tel qu'il est lorsqu'il est extrait avant d'être raffiné, transformé, C'est un mélange de pentanes et d'hydrocarbures plus lourds, généralement extraits de réservoirs géologiques. Lorsque les pentanes et les hydrocarbures lourds se trouvent dans des réservoirs de gaz naturel, ils sont appelés "condensats", comparables, dans la pratique, au brut. Les réservoirs de pétrole peuvent produire des hydrocarbures liquides plus légers tels que le propane et le butane, qui entrent dans la catégorie des gaz naturels.

Il existe plusieurs types de coupes de pétrole brut, qui sont déterminées en fonction de leur poids moléculaire et de leur point d'ébullition. Les principaux types de coupes de pétrole brut incluent le naphta, le kérosène, le gazole et le fioul.

Le naphta est une fraction légère du pétrole brut, généralement utilisée comme matière première dans l'industrie pétrochimique pour la production de plastiques et de carburants synthétiques.

Le kérosène est une fraction intermédiaire plus lourde, souvent utilisée comme carburant dans les moteurs d'avion et comme combustible de chauffage.

Le gazole est une fraction plus lourde que le kérosène, utilisée comme carburant pour les véhicules diesel.

Enfin, le fioul est la fraction la plus lourde du pétrole brut, utilisée comme source d'énergie pour le chauffage et la production d'électricité.

### **I.3.Composition chimique des pétroles brut et des fractions pétrolières :**

Le pétrole brut est composé principalement d'hydrocarbures, avec des traces de soufre, d'azote, d'oxygène, et de composés métalliques (**Rebas 2016**)

Les principales fractions du pétrole brut sont les suivantes :

- Gaz naturel : composé principalement de méthane
- Essence : composée d'hydrocarbures légers tels que l'hexane et l'heptane
- Naphta : composée d'hydrocarbures plus lourds que l'essence
- Kérosène : composé d'hydrocarbures plus lourds que la naphta
- Diesel : composé d'hydrocarbures plus lourds que le kérosène
- Mazout : composé d'hydrocarbures très lourds

Chaque fraction a des propriétés et des utilisations spécifiques en fonction de sa composition chimique.

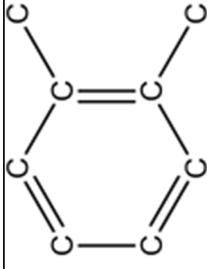
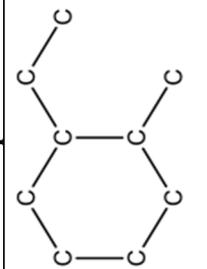
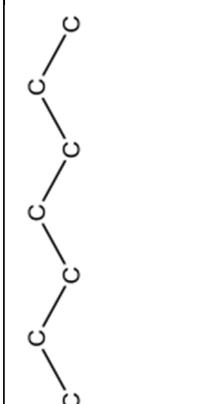
#### **I.3.1. Les hydrocarbures :**

Les hydrocarbures constituent la majeure partie des pétroles bruts et des fractions pétrolières. Stricto sensu, les hydrocarbures sont des molécules organiques composées

uniquement d'atomes de carbone et d'hydrogène. Elles se divisent en plusieurs familles chimiques selon leurs structures, qui reposent principalement sur la tétravalence du carbone. Les enchaînements moléculaires carbone-carbone peuvent être : (1) liés par une simple liaison (suffixe ane) ; les hydrocarbures saturés, ou (2) par des liaisons multiples doubles (suffixe ène) ou triples (suffixe yne) ; les hydrocarbures insaturés. Ces enchaînements moléculaires peuvent former soit des chaînes linéaires, soit des cycles, ces derniers pouvant être de structures simples ou ramifiées (**Rebas 2016 ; Vidal 1997 ; Wauquier 1998**)

Dans le secteur pétrolier, plusieurs méthodes de classification hiérarchique des hydrocarbures existent. Cependant, ils sont souvent divisés en deux grandes catégories : (1) les hydrocarbures saturés et (2) les hydrocarbures insaturés ou non saturés. En fonction de leurs structures moléculaires, ces derniers peuvent être classés en quatre familles chimiques : les paraffines, les naphènes, les aromatiques et les oléfines. Le tableau I.1 résume la classification de chaque famille et présente certaines de leurs principales caractéristiques.

**Tableau I.1 Classification des hydrocarbures en fonction de leurs familles (Dupuy 2015).**

Principales familles d'Hydrocarbures	Hydrocarbures saturés (Sans doubles liaisons)		Hydrocarbures insaturés (Avec doubles liaisons)			
	Atomes de carbone en ---Chaines---		Molécules contenant des ---Cycles---	Hydrocarbures comprenant des ---Cycles benzéniques---	Atomes de carbone en ---Chaines contenant des doubles liaisons---	
	Paraffines		Naphtènes	Aromatiques	Oléfines (une seul double liaison) Dioléfines (deux doubles liaisons)	
	En chaînes droites (Paraffines normale ou bien n-Paraffines)	En chaînes ramifiées (Iso-Paraffines)	(Cyclo-Paraffines)			
Hydrocarbures présents dans le pétrole brut						
Exemples						
Stabilité	Bonne	Bonne	Bonne		Oléfine $C-C-C-C=C-C$ Dioléfine $C=C-C-C=C-C$	Faible
Densité liquide	Faible	Faible	Moyenne			Faible
Teneur en H. Rapport H/C	Élevée	Élevée	Moyenne			Moyenne

### **I.3.1.1. Les paraffines :**

Le terme "paraffine" désigne toutes les molécules connues en chimie organique sous le nom d'hydrocarbures aliphatiques saturés, ou alcanes. Les paraffines sont des chaînes non cycliques d'atomes de carbone liés exclusivement par des liaisons simples. Chaque atome de carbone est toujours lié à quatre autres atomes, soit de carbone, soit d'hydrogène, avec une formule chimique générale de  $C_nH_{2n+2}$ . Les chaînes de paraffines peuvent être linéaires, appelées paraffines normales ou n-paraffines, ou ramifiées, dans ce cas appelées paraffines ramifiées ou iso-paraffines (**Rebas 2016 ; Dupuy 2015**).

Le Tableau I.1 présente un exemple de deux types de molécules que l'on peut trouver dans le pétrole brut : l'heptane pour les paraffines linéaires et le 3-éthylpentane pour les paraffines ramifiées.

### **I.3.1.2. Les naphènes :**

Les naphènes sont communément appelés cyclo alcanes en chimie organique. Ils font partie des hydrocarbures saturés qui comportent au moins une structure cyclique. Leur formule chimique générale est  $(n-c+1)$ , où  $c$  représente le nombre de cycles dans la molécule. Les structures cycliques les plus couramment rencontrées dans le pétrole brut sont celles à cinq atomes, comme le cyclopentane, et à six atomes, comme le cyclohexane. Les molécules naphéniques dérivent principalement de ces deux structures, soit par addition de chaînes paraffinées latérales, comme dans l'éthyle-cyclopentane, soit par fusion de plusieurs cycles saturés, comme dans le cas de la décline (**Rebas 2016 ; Wauquier 1998**).

Le tableau I.1 présente l'exemple d'une molécule de 1-éthyl-2-méthylcyclohexane, une des molécules naphéniques que l'on peut trouver dans le pétrole brut.

### **I.3.1.3. Les oléfines :**

Traditionnellement appelées alcènes, les oléfines sont des hydrocarbures aliphatiques insaturés possédant une ou plusieurs doubles liaisons carbone-carbone. Elles peuvent être structurées en chaînes linéaires, appelées oléfines normales, ou en chaînes ramifiées, appelées iso-oléfinés. Les oléfines ne se trouvent pas, ou très rarement, naturellement dans les pétroles bruts en raison de la réactivité élevée de leurs doubles liaisons. Cependant, lorsque le pétrole est soumis aux procédés de raffinage, en particulier ceux en absence d'hydrogène, ces

molécules peuvent se former. La formule chimique générale des oléfines est  $(n-d+1)$ , où  $d$  représente le nombre de doubles liaisons dans la molécule (Rebas 2016 ; Wauquier 1998).

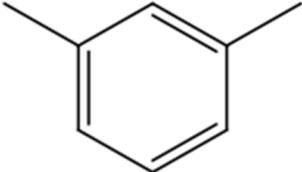
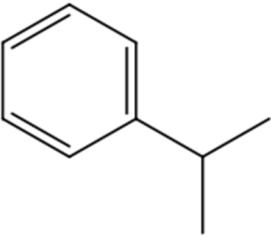
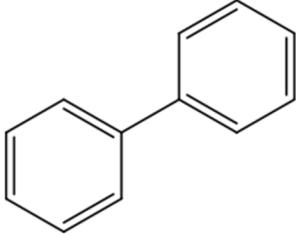
Le tableau I.1 donne deux exemples de molécules oléfiniques qui peuvent être rencontrées dans les coupes pétrolières, la (2E) -2-Pentène qui représente une oléfine simple et le 1,3-Butadiène qui représente une dioléfine simple.

#### **I.3.1.4. Les aromatiques :**

Les hydrocarbures aromatiques sont des composés polycycliques insaturés qui contiennent au moins une structure benzénique, composée de six atomes de carbone reliés par trois doubles liaisons conjuguées, alternant avec des liaisons simples. Cette conjugaison dans un cycle à six atomes confère une grande stabilité à la structure benzénique. Les autres structures aromatiques peuvent être dérivées de la structure benzénique de plusieurs façons : (1) par addition de chaînes paraffiniques latérales, comme dans le cas de l'éthylbenzène, (2) par jonction entre deux cycles benzéniques, comme dans la formation de diphenyle, (3) par fusion de cycles benzéniques, comme dans le naphthalène, et (4) par fusion de cycles saturés, comme dans la tétraline (Rebas 2016 ; Wauquier 1998).

La formule générale des hydrocarbures aromatiques varie selon leur dérivation par rapport à la structure benzénique de base. Pour les alkylbenzènes, la formule chimique générale est  $(n-6)$ . Les autres mono-aromatiques contenant des cycles naphténiques ont pour formule brute  $(n-c) -6$ , où  $c$  désigne le nombre de cycles naphténiques dans la molécule. Les d'aromatiques de type diphenyle alkylé ont pour formule brute  $C_nH_{2n-14}$ , tandis que la formule chimique générale est  $C_nH_{2n-12}$  pour les d'aromatiques alkyles avec deux cycles benzéniques fusionnés. Le Tableau I.2. Montre quelques exemples de molécules aromatiques présentes dans les fractions pétrolières (Rebas 2016 ; Belghit 2019)

**Tableau I.2. Structures de quelques hydrocarbures aromatiques ( Rebas 2016 ).**

Nom	Formule chimique	Structure moléculaire
m-xylène	$C_8H_{10}$	
i-propylbenzène	$C_9H_{12}$	
Diphényle	$C_{12}H_{10}$	

#### **I.3.1.5. Autres hydrocarbures :**

En plus des paraffines, naphènes, oléfines et aromatiques, il est possible de trouver à l'état de traces dans les mélanges pétroliers d'autres hydrocarbures insaturés, en particulier des alcynes et des cyclo-oléfines. Les alcynes sont des hydrocarbures aliphatiques contenant des triples liaisons carbone-carbone, tandis que les cyclo-oléfines possèdent une ou plusieurs doubles liaisons carbone-carbone sous forme cyclique. Les hydrocarbures alcyoniques sont très réactifs, c'est la raison qu'ils ne sont pas détectés dans les pétroles bruts et leurs teneurs dans les coupes pétrolières sont très faibles. En outre, il existe plusieurs isomères de position à l'état de traces dans les coupes pétrolières qui sont issues des effluents de procédés de conversion tels que les isomères cis, trans, para, méta et ortho (**Rebas 2016 ; Wauquier 1998 ; Belghit 2019**).

Le Tableau I.3 montre quelques exemples de ces molécules insaturés rencontrés dans les coupes pétrolières.

**Tableau I.3** Exemples de quelques structures d'autres hydrocarbures insaturés (Rebas 2016 ; Belghit 2019).

Nom	Formule chimique	Structure moléculaire
Éthylène	$C_2H_4$	
2-Butyne	$C_4H_6$	
Cyclopentane	$C_5H_8$	

#### **I.4. Propriétés usuelles des hydrocarbures et des fractions pétrolières :**

Les caractéristiques physiques et chimiques des pétroles bruts varient considérablement d'un champ de production à un autre, et même au sein d'un même gisement. Les propriétés physicochimiques, thermodynamiques, ainsi que les facteurs et indices spécifiques des hydrocarbures, du pétrole brut et de ses fractions jouent un rôle essentiel dans la caractérisation, l'identification et l'analyse de leurs qualités. Ces propriétés sont considérées comme des outils analytiques permettant de déterminer d'autres caractéristiques des pétroles bruts et des mélanges pétroliers. Ces outils doivent être non seulement simples à mettre en œuvre, mais également parfaitement reproductibles. De plus, ils doivent fournir des informations précieuses aux raffineurs, telles que la densité, l'indice de réfraction, la viscosité, et les indices d'octane ou de cétane (Rebas 2016).

Dans cette partie, je vais juste définir les propriétés les plus répandues et les organiserai de manière innovante pour faciliter la compréhension de chaque propriété.

#### **I.4.1. Propriétés liées aux fluides**

##### **I.4.1.1. Masse volumique**

La masse volumique d'une substance, également appelée densité volumique de masse, est une grandeur physique qui caractérise la masse de cette substance par unité de volume, généralement notée par la lettre grecque  $\rho$ . Cette grandeur dépend de la température et de la

pression : elle décroît avec l'augmentation de la température et augmente avec l'augmentation de la pression. La masse volumique s'exprime en kg/m<sup>3</sup>, g/cm<sup>3</sup> ou lb/ft<sup>3</sup> (**Daubert 1997**)

#### **I.4.1.2. Densité relative**

La densité relative d'un liquide, notée  $d^4_T$  et connue sous le terme anglo-saxon "Specific Gravity" ou SG, est un paramètre physique sans dimension spécifique à chaque produit (**Daubert 1997**).

Leur définition mathématique est indiquée par l'Eq :

$$d = \frac{\text{masse volumique du corps à}}{\text{masse du volume d'eau à } 4^{\circ}\text{C}}$$

#### **I.4.1.3. Gravité de l'API**

La densité API (en anglais : API Gravity), abréviation de "American Petroleum Institute Gravity", est une échelle utilisée pour mesurer la densité des produits pétroliers liquides. La densité API s'exprime en degrés API, avec des valeurs allant de 0, correspondant à une densité de 1,076, à 100, correspondant à une densité de 0,6112. Sa définition mathématique est indiquée dans l'équation suivante.

$$\text{API} = \frac{141,5}{\text{SG}(60^{\circ}\text{F}/60^{\circ}\text{F})} - \mathbf{131,5}$$

#### **I.4.1.4. Viscosité**

La viscosité est l'une des propriétés de transport utilisées pour caractériser les hydrocarbures. La viscosité dynamique, usuellement notée " $\mu$ ", est une grandeur physique qui mesure la résistance à l'écoulement laminaire d'un fluide incompressible. Elle s'exprime en pascal-seconde (Pa.s) dans le système SI et est déterminée en mesurant le temps d'écoulement d'un fluide dans un tube capillaire de longueur donnée à une température déterminée. La viscosité cinématique, notée " $\nu$ ", est le quotient de la viscosité dynamique par la masse volumique du fluide. Elle représente la capacité de rétention des particules du fluide et quantifie sa capacité à s'épancher, s'exprimant en mètres carrés par seconde (m<sup>2</sup>/s) (**Wauquier 1998**).

## **I.4.2. Autres propriétés physicochimiques**

### **I.4.2.1. Poids moléculaire**

Le poids moléculaire, noté  $M$ , est la somme des poids atomiques des éléments présents dans un composé. Il est utilisé pour convertir des quantités molaires en quantités massiques à des fins pratiques. Cette propriété est essentielle pour la caractérisation des hydrocarbures purs et des fractions pétrolières, et est généralement exprimée en kg/kmol, en g/mol ou en lb/mol.

### **I.4.2.2. Température d'ébullition**

Le point d'ébullition d'une fraction pétrolière, noté  $T_b$ , représente la température à laquelle cette fraction commence à bouillir et à passer de l'état liquide à l'état de vapeur sous pression atmosphérique (**Wauquier 1998**).

En termes simples, le point d'ébullition d'une fraction pétrolière est la température à laquelle sa pression de vapeur égale la pression atmosphérique. Il est utilisé pour déterminer d'autres paramètres ou propriétés physiques, tels que la volatilité, le poids moléculaire, le facteur acentrique ( $\omega$ ), les propriétés critiques, et l'indice de réfraction ( $\eta$ ).

### **I.4.2.3. Pression de vapeur**

À l'équilibre thermodynamique, la pression de vapeur ( $P_{vap}$ ) d'une substance donnée représente la force exercée par unité de surface des vapeurs, du liquide ou du solide présents dans un système (**Daubert 1997**).

La pression de vapeur est un indicateur de la volatilité d'un liquide, souvent utilisé pour calculer ses propriétés physiques, les pertes par évaporation et la plage d'inflammabilité. Plus la pression de vapeur d'un liquide est élevée, plus il est volatil. En général, les liquides volatils ont une pression critique plus élevée (**Riazi 2001**).

### **I.4.2.4. Point d'aniline**

Le point d'aniline est la température minimale à laquelle des volumes égaux d'aniline et d'un produit à analyser, comme une fraction pétrolière, se mélangent parfaitement. La perte de miscibilité se traduit par l'apparition d'un trouble net (**Rebas 2016**).

Le point d'aniline d'un hydrocarbure est influencé par son poids moléculaire et le nombre de carbones. Cette information est utile pour estimer la proportion d'aromatiques dans les fractions pétrolières et classer les hydrocarbures en différentes familles. Un point d'aniline

élevé indique une nature paraffinique, tandis qu'un point d'aniline bas indique une nature aromatique. Les hydrocarbures naphténiques se situent intermédiairement (**Rebas 2016 ; Belghit 2019 ; Daubert 1997**).

### **I.4.3. Propriétés thermodynamiques**

#### **I.4.3.1. Enthalpie standard de formation**

L'enthalpie standard de formation, ou chaleur standard de formation ( $\Delta H_{\text{of}}$ ), est une propriété thermodynamique fondamentale et importante des composés. Elle représente le changement d'enthalpie accompagnant la formation d'une mole de composé dans son état standard à partir de ses éléments constitutifs dans leurs états standards. Ce processus se déroule à la température spécifiée, généralement 298 K ou 25°C, et à une pression de 1 atm. La variation de  $\Delta H_{\text{of}}$  est principalement utilisée en thermochimie pour calculer les changements d'enthalpie des réactions chimiques ( $\Delta H_{\text{reaction}}$ ). Elle est mesurée expérimentalement en soustrayant la somme des enthalpies standards de formation des réactifs de la somme des enthalpies standards de formation des produits. De nombreuses méthodes de calcul de  $\Delta H_{\text{of}}$  existent dans la littérature, notamment les modèles basés sur les méthodes de contribution des groupes, qui sont largement utilisées (**Vatani et al. 2007**).

#### **I.4.3.2. Enthalpie de vapeur**

Les interactions entre dipôles permanents jouent un rôle majeur dans l'évaporation d'un liquide ou la condensation d'un gaz, ces changements d'état étant accompagnés d'une variation d'énergie. L'enthalpie de vaporisation ( $\Delta H^{\circ}\text{vap}$ ) est la quantité de chaleur nécessaire pour vaporiser une mole de substance à sa température d'ébullition et à pression constante.

### **I.4.4. Facteurs et indices spécifiques**

#### **I.4.4.1. Facteur de caractérisation de Watson**

Les interactions entre dipôles permanents jouent un rôle majeur dans l'évaporation d'un liquide ou la condensation d'un gaz, ces changements d'état étant accompagnés d'une variation d'énergie. L'enthalpie de vaporisation ( $\Delta H^{\circ}\text{vap}$ ) est la quantité de chaleur nécessaire pour vaporiser une mole de substance à sa température d'ébullition et à pression constante (**Riazi 2001**).

$$Kw = \left( \frac{1,8 \times Tb}{SG} \right)^{1/3}$$

#### **I.4.4.2. Facteur acentrique**

Le facteur acentrique ( $\omega$ ), développé par Pitzer, est un paramètre de caractérisation utilisé pour affiner la précision des équations d'état correspondantes appliquées aux hydrocarbures lourds. Il évalue la déviation des propriétés thermodynamiques d'un fluide par rapport aux valeurs calculées à partir du principe des états correspondants **(Riazi 2001)**.

#### **I.4.4.3. Facteur de compressibilité**

En thermodynamique, le facteur de compressibilité ( $Z$ ) est une grandeur intensive adimensionnelle qui exprime le rapport du volume molaire d'un mélange ou d'un corps pur réel, liquide ou gazeux, au volume molaire du gaz parfait correspondant aux mêmes pressions, températures et compositions. Pour un gaz parfait, le facteur de compressibilité est toujours égal à 1, quelle que soit la nature du gaz ou les conditions de pression, température et composition. En revanche, pour un fluide réel, le facteur de compressibilité varie généralement entre 0,2 et 1,2, et dépend de la nature du fluide ainsi que des conditions opératoires **(Sutton 1985 ; Elsharkawy 2000)**.

Le facteur de compressibilité est une variable essentielle dans les équations d'état telles que les équations d'état cubiques ou l'équation de Viriel. Il est crucial pour de nombreux calculs en ingénierie pétrolière, notamment pour estimer la chute de pression lors de l'écoulement de gaz à travers une conduite **(Elsharkawy et al.2001)**.

#### **I.4.4.4. Indice de réfraction**

L'indice de réfraction, parfois noté ( $I$ ) ou ( $\eta$ ), représente le rapport entre la vitesse de la lumière dans l'air, à une température et une pression donnée, et sa vitesse dans la substance ou le fluide testé, pour lequel l'indice de réfraction est mesuré à une longueur d'onde spécifiée. **(Riazi 2001)**.

L'indice de réfraction peut être défini comme le rapport entre le sinus de l'angle d'incidence et le sinus de l'angle de réfraction lorsque la lumière passe de l'air dans la substance testée, ce qui constitue l'indice de réfraction relatif. Pour obtenir l'indice de réfraction absolu par rapport au vide, il faut multiplier cet indice de réfraction relatif par 1,000 293, qui est l'indice de réfraction absolu de l'air. La valeur numérique de l'indice de réfraction des liquides varie inversement avec la longueur d'onde de la lumière et la température **(Riazi 2001)**.

#### **I.4.4.5. Indice d'octane**

La qualité des carburants pétroliers est déterminée par leur composition et les types d'hydrocarbures présents dans le mélange. L'indice d'octane, ou nombre d'octane (NO), est une caractéristique essentielle des carburants pour moteurs à allumage par étincelle, tels que l'essence et le kérosène. Il mesure la résistance à la détonation du carburant et est fortement influencé par le type d'hydrocarbure (**Albahri et al.2002 ; Albahri 2003**).

Le nombre d'octane d'un produit pétrolier est déterminé par la structure chimique de ses composants. Les hydrocarbures ramifiés présentent une meilleure résistance à la détonation que ceux à structure linéaire. Une augmentation de l'indice d'octane réduit les risques de cliquetis moteur. Les n-paraffines ont des indices d'octane bas, tandis que les iso-paraffines, les oléfines et les aromatiques ont des indices élevés (**Rebas 2016 ; Abdul Jameel et al. 2018**)

Les naphthènes présentent des valeurs intermédiaires. Ainsi, une essence issue d'un reforming catalytique ou d'un cracking catalytique a généralement un indice d'octane plus élevé qu'une essence de distillation directe. Le principal indice utilisé dans les stations-service est le RON (Research Octane Number), qui représente le comportement du carburant dans des conditions d'utilisation normales. En revanche, le MON (Motor Octane Number) reflète le comportement du carburant dans des conditions plus exigeantes, telles que des régimes élevés ou une conduite sur autoroute. La figure I.7 offre un aperçu des caractéristiques de chaque famille d'hydrocarbures en ce qui concerne le RON. Elle fournit également des informations sur les relations générales entre la structure chimique des hydrocarbures et leur indice d'octane (**Rebas 2016**).

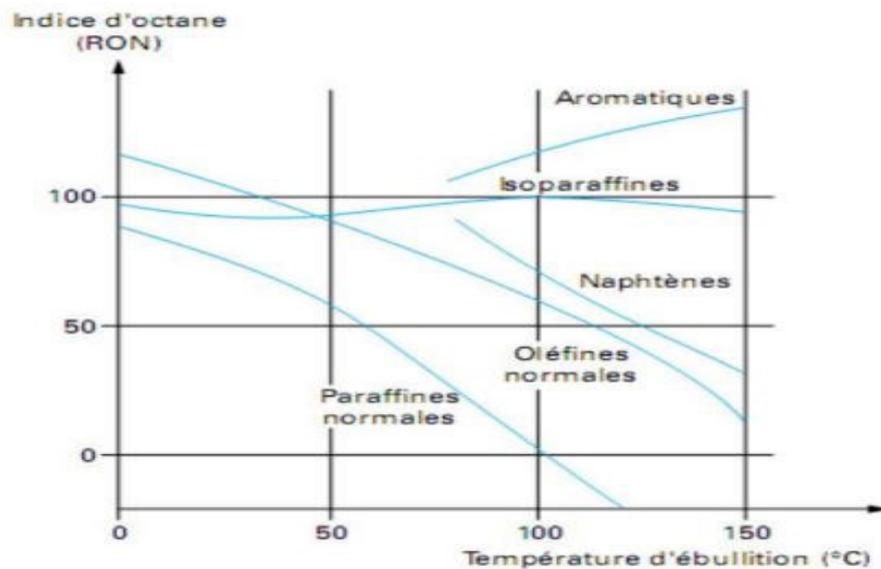


Figure I.1. Indices d'octane des différentes familles chimiques d'hydrocarbures.

## I.4.5. Propriétés liées à la sécurité et le risque d'explosion

### I.4.5.1. Point éclair

Le point éclair des hydrocarbures et de leurs mélanges est une donnée de sécurité cruciale pour leur traitement et leur manipulation. Il est spécifié dans la fiche de données de sécurité typique. Ce point représente la température minimale à laquelle la pression de vapeur d'un hydrocarbure est suffisante pour générer une vapeur inflammable en présence d'une source externe d'ignition, comme une étincelle ou une flamme (Atlas et al. 1992 ; Pan, Y et al. 2007 ; Alqaheem et al. 2017)

Le point d'éclair est un paramètre crucial pour les considérations de sécurité, notamment lors de la production, du traitement, du stockage et du transport des produits pétroliers volatils et des combustibles, comme le GPL, le naphta, l'essence et le diesel. Il est également important pour les carburants non-pétroliers tels que les carburants dérivés du charbon et les biocarburants, surtout dans des environnements à haute température (Atlas 1992 ; Alqaheem 2017)

Le point d'éclair est un indicateur du risque d'incendie et d'explosion d'un carburant. Un point d'éclair bas signifie un risque accru d'incendie. Il ne doit pas être confondu avec d'autres propriétés liées à la volatilité, telles que la limite d'inflammabilité, la température d'auto-inflammation, le point d'éclair et la pression de vapeur (Alqaheem 2017)

#### **I.4.5.2. Température d'auto-inflammation**

La température d'auto-inflammation, également connue sous le nom de température d'auto-inflammation (AIT) en anglais, est un paramètre crucial de l'inflammabilité qui requiert une attention particulière dans les industries et les processus manipulant des matériaux dangereux et inflammables (**Borhani et al. 2016**)

La température d'auto-inflammation est l'un des paramètres les plus essentiels pour classer les produits chimiques en termes d'inflammabilité et évaluer leur potentiel à former une atmosphère explosive (**Bagheri, et al. 2012**)

La température d'auto-inflammation (AIT) représente la température la plus basse à laquelle une substance s'enflamme dans l'air à la pression atmosphérique (**Borhani et al. 2016; Bagheri et al. 2012**)

### **I.5. Propriétés critiques :**

#### **I.5.1. Définition :**

Les propriétés critiques sont des caractéristiques physiques des corps purs. La température et la pression critiques définissent le point limite supérieur de la courbe de tension de vapeur, au-delà duquel le changement de phase n'est plus observé. En ce point, la densité et la composition des deux phases liquide et vapeur sont identiques ; la séparation d'un tel mélange et donc impossible en ce point.

La connaissance des propriétés critiques est indispensable dans le traitement des mélanges d'hydrocarbures où la coexistence d'une phase gazeuse et liquide est fréquente. Ils permettent de prévoir le comportement des fluides dans un large intervalle de températures et de pressions et définir ainsi les conditions d'existence de leurs différentes phases (**Rebas 2016**) Les propriétés critiques sont des caractéristiques physiques des corps purs. La température et la pression critiques définissent le point au-delà duquel le changement de phase n'est plus observable, et où les densités et compositions des phases liquide et vapeur deviennent identiques, rendant leur séparation impossible. Connaître le point critique permet de prévoir le comportement des substances sur une large gamme de températures et de pressions, en utilisant la loi des états correspondants ou les équations d'état (**Rebas 2016**)

La figure I.5 montre un diagramme pression-température où le point de convergence des courbes d'équilibres de sublimation, de fusion et de vaporisation d'un corps pur est appelé

"point triple". Ce point délimite les domaines des états liquide, solide et gazeux. Comme le montre la figure I.6, pour un mélange, la courbe d'équilibre de vaporisation se divise en deux courbes : la courbe de bulle et la courbe de rosée (Rebas 2016 ; Belghit 2019)

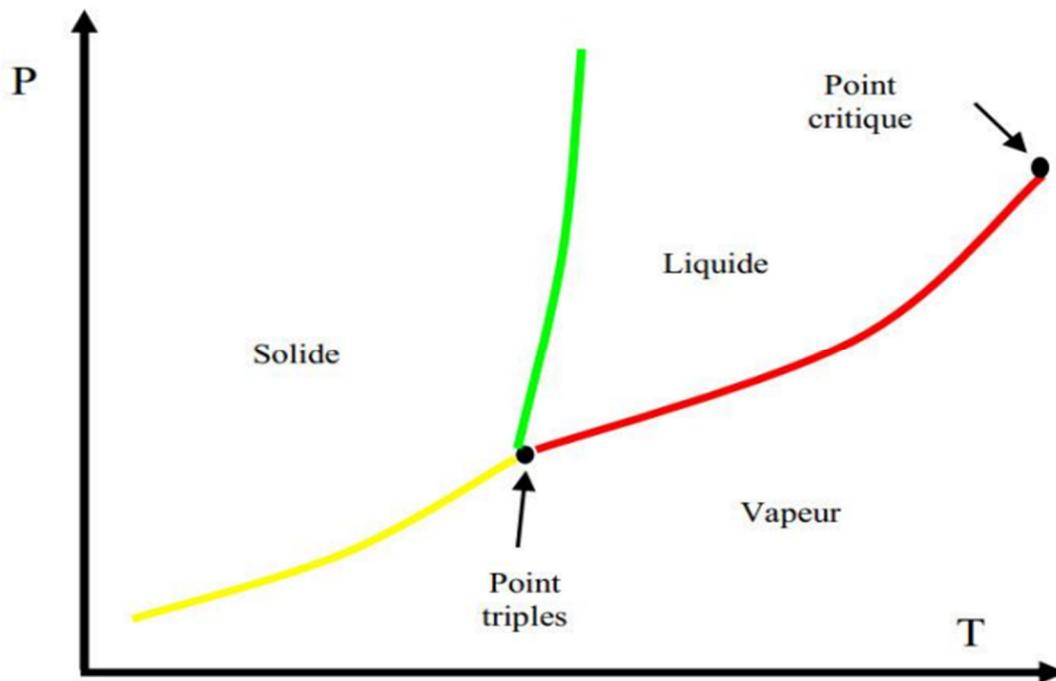
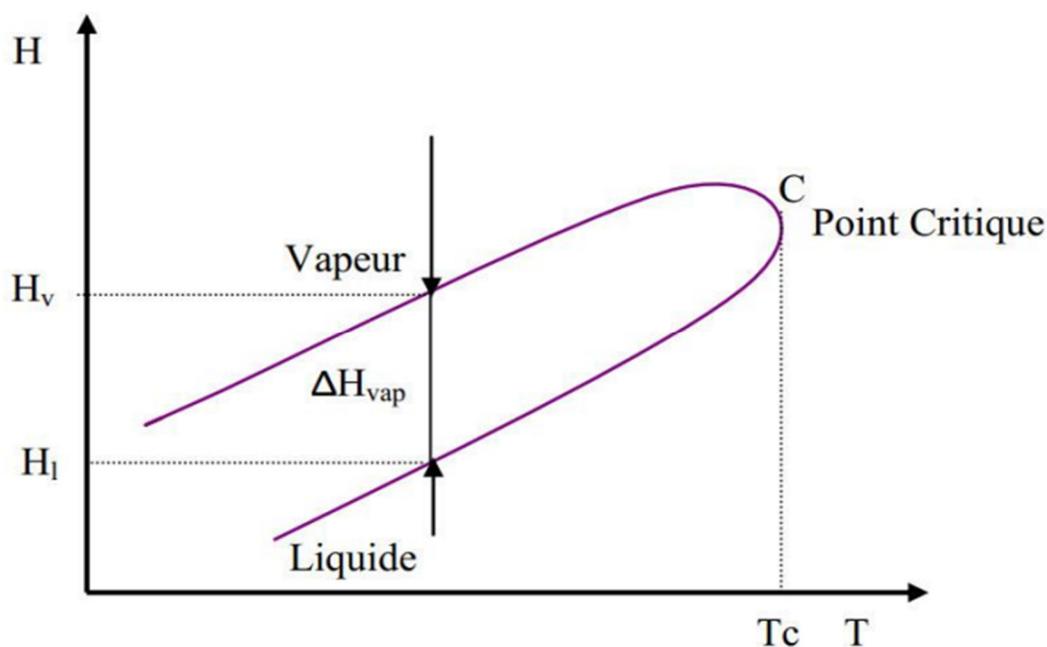


Figure I.2. Diagramme de coordonnées (P-T) des trois équilibres d'un corps pur.



**Figure I.3.** Représentation de la chaleur de vaporisation pour un mélange de corps purs.

Au-delà d'un point "C", appelé point critique, l'équilibre entre les phases liquide et gazeuse devient impossible pour les corps purs et leurs mélanges, comme illustré dans le diagramme P-T de la figure I.5. Ce point est défini par la température critique ( $T_c$ ), la pression critique ( $P_c$ ), et le volume critique ( $V_c$ ), où les propriétés du liquide et de la vapeur, telles que la densité et l'indice de réfraction, deviennent identiques, et la chaleur de vaporisation devient nulle. Les gaz au-dessus de leur température critique ne peuvent pas être liquéfiés par simple changement de pression, et les liquides au-dessus de leur pression critique ne peuvent pas passer en phase gazeuse. Le fluide au-dessus de son point critique est appelé fluide supercritique. Le volume critique ( $V_c$ ) est le volume molaire à  $T_c$  et  $P_c$  (Rebas 2016 ; Riazi, 2001)

Les propriétés critiques des hydrocarbures et des fractions pétrolières sont cruciales pour déterminer les conditions de phase, les plages de fonctionnement des réacteurs, et les équipements de transfert de matière comme les colonnes de distillation, les absorbeurs et les colonnes d'extraction (Daubert 1997). Ces propriétés sont également utilisées pour calculer certaines propriétés physiques, thermodynamiques et de transport des hydrocarbures et des fractions pétrolières.

Deux techniques couramment utilisées pour mesurer les propriétés critiques sont la méthode d'écoulement et la méthode d'ampoule. Ces deux méthodes reposent sur le principe de l'opalescence critique, où la température, la pression et le volume critiques d'une substance

Sont évalués lorsque le ménisque séparant les phases liquide et vapeur disparaît, laissant place à une bande d'un brouillard opalescent. Une technique vidéo permet de visualiser l'image de la cellule, tandis qu'un ordinateur enregistre la propriété critique mesurée. Pour les substances thermiquement instables, il est nécessaire de limiter le temps de séjour dans la cellule pour minimiser les réactions de décomposition **(Rebas 2016)**.

Étant donné que la détermination expérimentale de ces propriétés requiert un équipement coûteux et complexe, ainsi qu'une manipulation minutieuse et chronophage, il est souvent nécessaire d'utiliser des techniques de modélisation pour les estimer.

## **I.6. Mesures expérimentales :**

Deux techniques sont généralement utilisées pour la mesure des propriétés critiques, à savoir : la méthode d'écoulement et la méthode d'ampoule. Ces deux méthodes sont basées sur le principe de l'opalescence critique, qui consiste à évaluer la température, la pression et le volume critiques d'une substance lorsque le ménisque séparant les phases liquide et vapeur disparaît, laissant place à une bande d'un brouillard opalescent. Une technique de vidéo a permis de visualiser l'image de la cellule et un ordinateur enregistrerait la propriété critique mesurée. Pour les substances instables thermiquement, il faut effectuer les mesures de température et de pression lorsque le temps de séjour dans la cellule est très faible et ceci afin de limiter les réactions de décomposition **(Rebas 2016)** Comme la détermination expérimentale de ces propriétés nécessite un appareillage coûteux et sophistiqué ainsi qu'une manipulation longue et délicate, il est souvent nécessaire de faire appel à des techniques de modélisation pour les estimer.

## **I.6.Raffinage du pétrole :**

La capacité de l'industrie de raffinage est aujourd'hui suffisante pour assurer la transformation du pétrole en produit pétrolier. Cependant cette capacité n'est pas également répartie selon les régions du monde, tant en terme de nombres de raffineries qu'en terme de technologie utilisée par ces raffineries. La plus importante concentration de raffineries se trouve en Extrême Orient et plus particulièrement en Chine. Les pays développés comme l'Europe de l'Ouest et l'Amérique du Nord comptent à eux seuls pour environ 40 % du nombre total de raffineries dans le monde. Il est à noter que ces pays consomment environ 40% de la production mondiale de pétrole **(Bagheri et all.2012)**

## **I.7. Conclusion**

Dans ce chapitre, nous avons abordé les concepts fondamentaux ainsi que les généralités sur les hydrocarbures purs, le pétrole brut et ses différentes fractions. Nous avons inclus différentes définitions essentielles, leur classification et les diverses compositions. De plus, nous avons introduit des notions clés pour comprendre les caractéristiques physico-chimiques des hydrocarbures purs, des pétroles bruts et de leurs fractions pétrolières.

En outre, le processus de séparation des pétroles bruts, c'est-à-dire le procédé de raffinage, a été succinctement exposé dans ce chapitre.

Comme il est largement reconnu, déterminer les propriétés des hydrocarbures pose des défis en raison du temps et des ressources consommés par les méthodes expérimentales. Toutefois, une vue d'ensemble des méthodes de détermination et de prédiction des propriétés des hydrocarbures, incluant leurs principales techniques, a été présentée. En conclusion, il convient de noter que le domaine des hydrocarbures est vaste et complexe, ne pouvant être pleinement exploré en un simple chapitre. Heureusement, il existe des thèmes et des ouvrages de référence approfondis qui explorent et expliquent ce domaine de manière plus détaillée.

---

***Chapitre II***  
***Techniques de modélisation***  
***neuronal et méthodes de***  
***RQSP/RQSA***

---

## **II.1. Introduction**

Les caractéristiques physicochimiques essentielles et les propriétés structurelles d'une molécule influencent son comportement lors des processus physiques, chimiques, biologiques ou environnementaux, ce qui les rend intéressantes pour comprendre et représenter l'action des molécules composés.

Une relation de structure-propriété quantitative (RQSP), ainsi que la relation de structure-activité quantitative (RQSA), est le processus par lequel une structure chimique est liée à un effet spécifique, tel que l'activité biologique ou la réactivité chimique. La RQSP cherche à représenter la relation entre une structure chimique et différentes caractéristiques physiques ou chimiques. De la même façon, la RQSA cherche à établir une corrélation quantitative entre la structure moléculaire d'un composé et son influence biologique. via des descripteurs moléculaires (**Chitita (2017) ; Engel et all (2018)**).

Les descripteurs moléculaires peuvent être obtenus en appliquant les principes de diverses théories, comme la chimie quantique, l'informatique, la chimie organique et la théorie des graphes (**Engel et all (2018); Todeschini et all (2009)**)

Ils servent à modéliser un large éventail de caractéristiques dans tous les domaines de la chimie et de la biologie, tels que les sciences environnementales, la toxicologie, la chimie analytique, la chimie physique, la chimie médicinale et les sciences de la vie. En plus des descripteurs théoriques conventionnels (calculés), il existe également des descripteurs obtenus de façon expérimentale (comme le logP, le moment dipolaire) parfois employés dans les études RQSA (**Todeschini et all (2009)**). Dans la suite, on se concentre exclusivement sur les descripteurs moléculaires théoriques.

Ce chapitre donne un aperçu sur l'approche RQSP/RQSA. Les différentes méthodologies de développement, de validation et l'application de ces méthodes sont aussi présentées.

## **II.2. Approche relations quantitatives structures propriétés/activités RQSP/RQSA .**

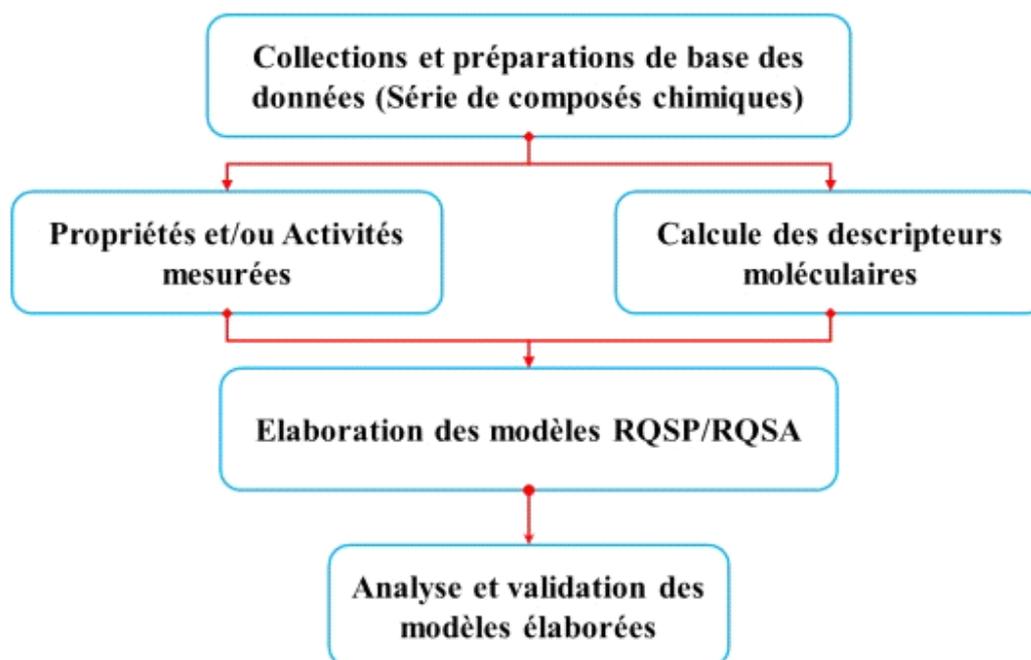
### **II.2.1. Domaines d'application et utilité des méthodes RQSP/RQSA**

Les méthodes RQSP/RQSA ont été utilisées au cours des dernières décennies dans toutes les disciplines où des produits chimiques sont employés ou étudiés, comme le domaine pharmaceutique, la pétrochimie, les sciences de l'environnement et d'autres. Les modèles déjà établis sont capables de prédire les caractéristiques ou les actions de nouveaux composés (**Hamadache (2014)**) En outre, les outils de description moléculaire et les méthodes de modélisation adéquates peuvent aider à comprendre et à rationaliser les liens entre la structure moléculaire et une propriété ou une activité dans une série de molécules (**Dearden et al (2009) ; Engel et al (2018)**). De plus, l'emploi des méthodes RQSP/RQSA présente de nombreuses applications ; nous présentons ci-dessous, à titre d'exemple, quelques exemples courants d'utilisation des modèles RQSP/RQSA dans différents secteurs (**Dearden et al(2009) ; Engel et al (2018) )**

- Prédiction des différentes propriétés physico-chimiques et activités biologiques à partir de leurs structures chimiques.
- Identification des composés dangereux dès les premiers stades de développement de tous types des produits.
- Comprendre et rationaliser les mécanismes d'action dans une série de produits chimiques.
- Réduire le coût de développement des différents produits industriels tels que les produits pharmaceutiques, les hydrocarbures, et les pesticides, ...
- Optimisation des propriétés pharmacocinétiques (ADME) des médicaments.
- Diminuer ou même remplacer les tests longs et coûteux sur les animaux.
- Conception et prédiction de la toxicité et des effets secondaires des nouveaux composés pour l'homme et les espèces environnementales.
- Prédire les effets ou la toxicité non ciblés de nouveaux composés.
- Prévision du devenir des composés rejetés dans l'environnement.

## II.2.2. Principe et méthodologie de base des études RQSP/RQSA

Les méthodes RQSP/RQSA sont basées sur l'hypothèse que la propriété ou l'activité d'un composé chimique est liée à sa structure. Par définition, Une RQSP/RQSA est un modèle mathématique qui associe un ou plusieurs paramètres quantitatifs dérivés de la structure chimique, à une mesure quantitative d'une propriété chimique ou d'une activité biologique (**Roubehie Fissa et all (2019); Chtita (2017)**).



**Figure II.1.** Principe et méthodologie de base des modèles RQSP/RQSA (**ROUBEHIE FISSA (2021)**)

Le principe de base adopté dans toutes études RQSP/RQSA (**Figure II.1**), consiste à trouver une relation mathématique reliant de manière quantitative un effet bien déterminé, comme une propriété chimique, ou une activité biologique, mesurée pour une série de composés similaires dans les mêmes conditions expérimentales, avec des descripteurs moléculaires à l'aide des méthodes statistiques (**Chtita, (2017)**).

L'objectif de ces études est d'analyser les données structurales afin de détecter les facteurs déterminants pour la propriété ou l'activité étudiée. Pour ce faire, différents types de méthodes statistiques peuvent être employés, et sont étudiés dans

la section II.6 de ce chapitre (voir plus loin la partie II.6. Méthodes d'apprentissage statistiques).

L'expression mathématique obtenue peut alors être utilisée comme moyen prédictif de la propriété ou l'activité étudiée pour de nouvelles molécules ou des molécules pour lesquelles les données expérimentales ne sont pas disponibles. Ceci peut être traduit par l'équation suivante :

$$\text{Propriété/Activité} = (\text{Descripteurs Moléculaires})$$

### II.3. Codage et représentation des structures chimiques

Les molécules sont représentées par différentes structures chimiques, ces structures sont figurées par des écritures de forme développées et semi-développées. Pour quantifier ces structures chimiques lors de la modélisation, il est nécessaire de les coder préalablement pour les rendre compréhensibles par les outils informatiques. Par la suite les informations structurelles seront extraites pour être utilisées lors de la construction du modèle de type RQSP/RQSA (**Roubehie Fissa et all (2019)**). Il existe de nombreuses manières ou langages de codage, telles que la théorie des graphes qui est considérée comme une source d'inspiration pour la création de plusieurs types de descripteurs moléculaires, notamment les descripteurs topologiques (**Balaban (1985) ; Bonchev (1991) ; Randić (1991) ; Lahiouel (1995a)**).

#### II.3.1. Utilisation de la théorie des graphes pour le codage des structures chimiques :

La théorie des graphes fait partie des mathématiques discrètes, en relation avec la topologie et la combinatoire. Elle aborde la manière dont les objets sont reliés et toutes les répercussions de leur connexion. La connectivité est donc une caractéristique essentielle d'un système. Théorie graphique (**Balaban (1985) ; Hansen et all (1988) ; Mihalić et all (1992)**)

La théorie chimique des graphes est une discipline de la chimie mathématique, et donc de la chimie théorique. Le but est d'utiliser des graphes chimiques, c'est-à-dire des graphes qui illustrent des systèmes chimiques. La théorie des graphes chimiques se concentre donc sur l'étude de toutes les répercussions de la connectivité dans un

système chimique. Autrement dit, la théorie chimique des graphes concerne l'ensemble des domaines d'application de la théorie des graphes **Randić (1991)** ; **Mihalić et all (1992)**).

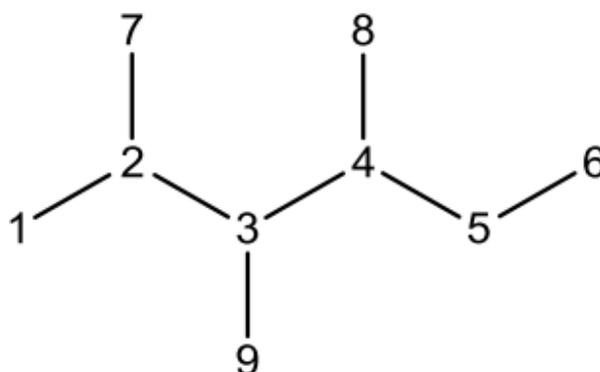
**Tableau. II.1.** Notions correspondant à la théorie des graphes et la théorie des graphes chimiques **Hansen et all (1988)**

<b>Théorie des graphes</b>	<b>Théorie des graphes chimiques</b>
point	Atome
Ligne	Liaison chimique
chemin	Sous-structure chimique
Degré	Valence de l'atome
Graphique cyclique	Composé cyclique
Arbre	Structure acyclique

### **II.3.1.1. Représentations d'une structure par quelques éléments basés sur la théorie des graphes chimiques**

Plusieurs représentations sont possibles pour les graphes. Un graphe chimique noté (**G**) peut être associé à plusieurs vecteurs ou matrices théoriques très importantes, telles que la matrice d'adjacence des sommets (en anglais : Vertex-Adjacency Matrix), la matrice des distances (en anglais : Distance Matrix) et le vecteur des valences (en anglais : Valences Vector) (**Mihalić et all (1992)** ; **Goulon et all (2008)**).

Le **Tableau II.2** montre un exemple de calcul de ces éléments pour un graphe moléculaire ( $G$ ) correspondant à la structure du 2.3.4-triméthylhexane présenté dans la **Figure II.2**, avec les sommets correspondant aux atomes et les lignes correspondant aux liaisons chimiques.



**Figure II.2.** Graphe moléculaire ( $G$ ) correspondant à la structure du 2.3.4-triméthylhexane.

**Tableau. II.2.** Exemple de calcul des matrices et vecteur basé sur la théorie des graphes chimique correspond à la structure du 2.3.4-triméthylhexane (**Mihalić et all (1992)**).

Matrice d'adjacence des Sommets $A(G)$	Matrice de distance $D(G)$	Vecteur des valences $V(G)$
$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 2 & 4 & 3 \\ 1 & 0 & 1 & 2 & 3 & 4 & 1 & 3 & 2 \\ 2 & 1 & 0 & 1 & 2 & 3 & 2 & 2 & 1 \\ 3 & 2 & 1 & 0 & 1 & 2 & 3 & 1 & 2 \\ 4 & 3 & 2 & 1 & 0 & 1 & 4 & 2 & 3 \\ 5 & 4 & 3 & 2 & 1 & 0 & 5 & 3 & 4 \\ 2 & 1 & 2 & 3 & 4 & 5 & 0 & 4 & 3 \\ 4 & 3 & 2 & 1 & 2 & 3 & 4 & 0 & 3 \\ 3 & 2 & 1 & 2 & 3 & 4 & 3 & 3 & 0 \end{pmatrix}$	$(1 \ 3 \ 3 \ 3 \ 2 \ 1 \ 1 \ 1 \ 1)$

### II.3.2. Notification SMILES et codage des structures chimiques

Dans la formalisation de la chimie, la première étape est de désigner un composé chimique. Il est essentiel d'avoir une notation précise et reproductible, que ce soit pour le plus simple atome ou pour la structure la plus complexe.

Un système de notation chimique appelé SMILES (Simplified Molecular Input Line Entry System) a été développé pour le traitement de l'information chimique contemporaine. La notation SMILES, basée sur les principes de la théorie des graphes moléculaires, offre une spécification de structure précise en utilisant une grammaire simple et extrêmement pratiquée **Weininger, D. (1988)**. Grâce à sa souplesse, à sa convivialité et à sa compatibilité avec les outils informatiques chimiques contemporains, le système de notation SMILES est particulièrement adapté au traitement des bases de données dans de nombreuses études de modélisation.

## **II.4. Notions de base sur les descripteurs**

### **II.4.1. Définition**

Un descripteur moléculaire est défini comme le résultat final d'une procédure logique et mathématique ou d'une expérience normalisée, qui transforme l'information chimique codée par une représentation symbolique d'une molécule en un nombre utile afin de pouvoir quantifier la molécule (**Roubehie Fissa et all (2019)** ; **Todeschini et all (2008)** ; **Mauri et all (2017)**).

### **II.4.2. Classification**

Les descripteurs peuvent être de différents types en fonction de leur méthode de calcul ou de détermination (**Roy et all (2015)**) Historiquement, deux grands schémas pour la classification des descripteurs moléculaires ont été établis : l'un en fonction de leur origine (constitutionnel, topologique, géométrique, quantique, thermodynamique...), et l'autre basée sur leurs dimensionnalités (0D, 1D, 2D, 3D ou 4D) **Karelson et all (2000)** ; **Dudek et all (2006)**). Dans ce qui suit-on donnera un aperçu sur ces catégories de descripteurs afin de bien clarifier les notions liées à chaque type

#### **II.4.2.1. Descripteurs 0-D**

Les descripteurs moléculaires de classe 0-D ne requièrent aucune donnée structurelle concernant la connectivité des atomes, ils ne renferment que quelques informations basiques telles que le nombre d'atomes et de liaisons, ainsi que la somme ou la moyenne des atomes

Éléments atomiques. Les descripteurs de cette classe sont toujours calculables sans l'utilisation des outils informatiques modernes, ne requièrent pas d'optimisation

des structures et sont facilement interprétables. Cependant, leur contenu informationnel est limité, mais ils peuvent tout de même jouer un rôle important dans la modélisation de diverses propriétés physico-chimiques ou participer à des modèles plus complexes (Chtita (2017); Bonachera (2011).)

#### **II.4.2.2. Descripteurs 1-D**

Les descripteurs moléculaires en 1-D se basent sur des caractéristiques de la structure moléculaire en une dimension, notamment le comptage des groupes fonctionnels et des fragments de sous-structure. Bien que similaires aux descripteurs en 0-D en termes de simplicité de calcul, les descripteurs en 1-D peuvent nécessiter l'utilisation d'outils informatiques pour accélérer le processus. Leur contenu informationnel est jugé acceptable, mais peut varier par rapport aux descripteurs en 0-D. Ces descripteurs présentent généralement une dégénérescence moyenne ou élevée et sont couramment utilisés dans la modélisation des propriétés physico-chimiques et biologiques (Tauler et al (2009) ; Nieto-Draghi et al (2015))

#### **I.4.2.3. Descripteurs 2-D**

A l'opposé des deux classes précédentes, les descripteurs 2-D nécessitent des informations structurelles sur les molécules notamment ceux qui sont liés à la connectivité des atomes et certains fragments moléculaires. De ce fait, toutes les molécules sont représentées sous formes des graphes chimiques, ensuite les informations nécessaires pour calculer les descripteurs de ce type sont extraites. C'est à partir de ce niveau que l'on peut espérer capturer des informations pertinentes pour la prédiction de la majorité des propriétés moléculaires (Chtita (2017); Karelson (2000)). Parmi les descripteurs les plus importants appartenant à cette classe, on trouve les indices topologiques qui ont été développés sur la base de la théorie des graphes chimiques, comme l'exemple de l'indice de Wiener "*W*" fondé en 1947 (Wiener (1947)), qui compte le nombre total des liaisons dans les chemins les plus courts entre les paires d'atomes (Goulon-Sigwalt-Abram (2008) ; Wiener (1947)), l'exemple de l'indice de Randić "*χ*" fondé en 1975 (Randić (1975)), qui se base sur la somme des bords pondérés dans un graphe moléculaire (Mihalić et al (1992) ; Randić (1975)) et l'exemple de l'indice de Balaban "*J*" fondé en 1982[53] , qui

compte la somme des distances pondérées dans une molécule (**Mihalić et all (1992)** ; **Balaban (1982)**)

#### **II.4.2.4. Descripteurs 3-D**

Ce type de descripteurs dépend essentiellement de la conformation tridimensionnelle des molécules. Leurs évaluations nécessitent de connaître les positions relatives de leurs atomes dans l'espace et de décrire des caractéristiques plus complexes. Une bonne connaissance de la modélisation "empirique" ou "ab-initio" ainsi que de la géométrie 3-D des molécules concernées, sont souvent les clés du calcul de ce type de descripteurs. Malgré le développement remarquable des outils informatiques, la plupart de ces descripteurs s'avèrent relativement coûteux en temps de calcul, mais ils fournissent plus d'informations par rapport aux autres classes précédentes, ils sont nécessaires à la modélisation de propriétés ou d'activités qui dépendent principalement de la structure 3-D (**Chtita (2017)**; **Kraim (2009)**) Il existe plusieurs familles de descripteurs appartenant à cette classe 3-D, le plus souvent on trouve :

- Les descripteurs géométriques qui peuvent être obtenus expérimentalement ou par Modélisation, empirique ou ab-initio. Ils sont basés sur l'arrangement spatial des atomes constituant la molécule et sont définis par les coordonnées des noyaux atomiques et la taille de la molécule représentée (**Chtita(2017)** ; **Kraim (2009)**) .Parmi ceux qui sont les plus importants, on trouve le volume moléculaire, la surface et le volume de Van Der Waals (**Higo et all (1989)** ; **Labute (2000)** ; **Saunders et all (2004)**).
- Les descripteurs quantiques électroniques qui caractérisent la distribution des charges moléculaires c'est-à-dire la polarité des molécules, mais aussi les paramètres de la chimie quantique qui nécessitent souvent des calculs plus sophistiqués. Les approches de la chimie quantique donnent accès à des informations supplémentaires telles que les données structurales, énergétiques, électroniques et spectroscopiques des molécules (**Chtita (2017)** ; **Kraim (2009)**). Parmi les plus importants on trouve l'énergie totale, le moment dipolaire, les énergies des orbitales frontières qui sont des paramètres très populaires dans la chimie quantique et dans les études RQSP/RQSA (**Kikuchi (1987)**), comme l'exemple des énergies "HOMO" et "LUMO" qui jouent un

rôle majeur dans de nombreuses réactions chimiques et dans les mécanismes. (Fukui (1970) ; Zhou (1990) ; Lewis, D. F. V., Ioannides et all (1994)).

- Les descripteurs physico-chimiques et thermodynamiques qui sont à l'origine des propriétés sont utilisés comme descripteurs introduits dans un modèle afin de prédire d'autres propriétés ou activités des systèmes étudiées (Chtita (2017) ; Kraim (2009)) La plupart de ces descripteurs sont obtenus expérimentalement, le plus souvent on trouve des paramètres comme le coefficient de partage octanol/eau, le parachor, la tension superficielle, la constante d'Henry, l'énergie libre de Gibbs et la chaleur de formation ou bien l'enthalpie de formation (Sugden (1924) ; Viswanadhan et all (1989) ; Bosque (2003)).

#### **II.4.2.5. Descripteurs 4-D**

Ils correspondent à la mesure des propriétés 3-D comme le potentiel électrostatique d'une molécule en tout point de l'espace. Ils permettent d'avoir des informations sur la structure cible souvent très complexes comme les protéines. Les descripteurs 4-D sont obtenus par le calcul des champs d'interactions moléculaires (CoMFA, CoMSIA) entre deux molécules (Chtita (2017) ; Ravi, M., Hopfinger et all (2001) ; Martins et all (2009) ; Andrade et all (2010)). On peut ainsi distinguer les descripteurs 4-D qui nécessitent un alignement de la molécule guidé par l'étude des complexes ligand-cibles avant d'être calculés ou, au moins, par des contraintes visant à optimiser le recouvrement spatial des champs électriques et stériques des ligands qui compensent le manque d'information sur le vrai mode de fixation dans la cible (Chtita (2017) ; Cramer et all (1988) ; Navajas et all(1996))

#### **II.4.3. Généralisation des descripteurs**

L'augmentation significative du nombre de descripteurs moléculaires au fil des décennies est largement attribuable aux progrès considérables réalisés dans le domaine mathématique et informatique. Ces avancées ont donné lieu à de nombreux logiciels spécialisés dans le calcul de divers types de descripteurs moléculaires (Chtita (2017) ; Cherkasov et all(2014)) parmi lesquels le logiciel "Dragon" occupe une place prépondérante. Avec sa dernière version, "Dragon 7.0", offrant jusqu'à 5270 descripteurs, cet outil est largement utilisé dans le domaine (Mauri et all (2006)) De même, la version en ligne gratuite "E-Dragon 1.0", utilisée dans des travaux de

recherche récents, propose environ 1666 descripteurs répartis dans différentes catégories (**Tetko et al (2005)**) Par ailleurs, d'autres logiciels et plateformes en ligne accessibles gratuitement, comme "ChemDes", offrent une multitude de fonctionnalités pour calculer des descripteurs moléculaires et des empreintes digitales. Ces outils comprennent également des logiciels de généralisation de descripteurs tels que "PaDEL Descriptors" (**Dong et al (2015)**) Malgré l'avantage indéniable que représente le calcul des descripteurs moléculaires à l'aide de ces divers outils, un défi majeur persiste : celui de sélectionner les meilleurs descripteurs adaptés à chaque étude.

### **II.5. Méthodes de sélection des descripteurs**

L'élaboration des modèles RQSP/RQSA nécessite parfois le calcul d'un grand nombre de descripteurs moléculaires, car les facteurs du processus étudié ne sont pas connus à priori (**Chitita (2017)**). Les descripteurs calculés sont différents les uns des autres par leur principe de calcul et leurs valeurs numériques pour une même molécule donnée. Ils n'ont donc pas tous une influence significative sur la grandeur à modéliser, et les variables ne sont pas toujours indépendantes. De plus, le nombre de descripteurs, c'est-à-dire la dimension de la base de données d'entrée, détermine la dimension du vecteur des paramètres à ajuster. Si cette dimension est trop importante par rapport au nombre des observations (molécules) de la base d'apprentissage, le modèle risque d'être sur-ajusté. Dans ce cas, elle est incapable de prédire la grandeur modélisée sur de nouvelles molécules et peut contenir des informations redondantes. Ce qui conduit à faire une réduction du nombre important de descripteurs et sélectionner les plus pertinents d'entre eux. Il ne s'agit pas d'utiliser alors le minimum de descripteurs pour expliquer la propriété ou l'activité ciblée, et il ne faut pas perdre des informations utiles. Comme Einstein a dit : "Tout devrait être fait aussi simple que possible, mais pas plus simple.". Dans une modélisation RQSP/RQSA, ce principe d'Einstein signifie que le modèle doit avoir le moins de paramètres possibles tout en traduisant au mieux l'information contenue dans la propriété ou l'activité visée (**Chitita (2017)**) Il existe plusieurs stratégies et outils pour la sélection des descripteurs moléculaires dans un procédé de construction des modèles RQSP/RQSA. Cependant, on peut classifier les stratégies utilisées en deux principales méthodes de sélections

(Chtita (2017) ; Kraim (2009)). Méthodes de sélection objective et Méthodes de sélection subjective.

### **II.5.1. Sélection objective**

Lors d'une analyse objective, il est essentiel de choisir soigneusement les variables pour réduire le nombre de caractéristiques sans considérer la variable dépendante (telle que l'activité ou la propriété ciblée). Dans ce processus de sélection, il est crucial d'éliminer les caractéristiques qui n'ont pas d'influence significative et d'éviter toute redondance. Par exemple, les caractéristiques présentant un pourcentage élevé de valeurs identiques pour l'ensemble des composés doivent être supprimées dès le départ, car elles ne fourniront aucune information distinctive utile. De même, si deux caractéristiques sont fortement corrélées, avec un coefficient de corrélation statistiquement significatif de plus de 0,95, il est préférable de ne pas les utiliser simultanément dans l'analyse des modèles pour éviter la redondance d'informations. Ce processus de sélection vise à réduire le nombre de caractéristiques sans perte d'informations essentielles.

### **II.5.2. Sélection subjective**

La sélection subjective de descripteurs est souvent nécessaire pour affiner les modèles RQSP/RQSA, même lorsque la sélection objective ne suffit pas. Cette approche consiste à choisir soigneusement les variables finales à partir d'un sous-ensemble déjà réduit de descripteurs afin de contribuer à la construction des modèles, en prenant en compte la variable dépendante. Différentes techniques d'optimisation sont utilisées pour déterminer de manière efficace le nombre optimal de descripteurs, sans évaluer toutes les combinaisons possibles. Les méthodes statistiques, telles que la sélection ascendante pas à pas, sont également employées pour éliminer les descripteurs qui n'ont pas d'influence significative sur les modèles RQSP/RQSA. Enfin, afin d'éviter que les modèles ne manquent de signification statistique ou ne contiennent des erreurs, le rapport entre les composés et les descripteurs doit être supérieur à 5, (Hamadache (2014) ; Topliss et al (1979) ; Pratim Roy et al (2009)).

## II.6. Méthodes d'apprentissage statistiques

Au cours de la dernière décennie, les méthodes d'apprentissage statistiques ont gagné en popularité tant dans le milieu académique que dans divers secteurs industriels (**Hamadache (2014); Ammi (2016)**). Elles ont été largement utilisées avec succès pour la modélisation prédictive, notamment dans la prédiction de diverses propriétés et activités ciblées ainsi que dans la sélection des descripteurs pertinents. Ces méthodes de sélection incluent l'approche pas à pas ascendante, basée sur des critères tels que la "Valeur-P" **Roubehie Fissa et all(2019) ; Roy et all (2015)**, ainsi que d'autres techniques comme les algorithmes génétiques et l'analyse factorielle. En ce qui concerne l'utilisation de ces méthodes pour la modélisation prédictive, des approches linéaires et non-linéaires telles que la régression linéaire multiple, les réseaux de neurones artificiels et les machines à vecteurs de support sont couramment utilisées **Hamadache (2014) ; Ammi (2016) ; Bitam (2018)**). Cette étude se concentrera particulièrement sur la régression linéaire multiple et les réseaux de neurones artificiels parmi ces méthodes.

### II.6.1. Régression linéaire multiple

En raison de sa simplicité la méthode de Régression Linéaire Multiple ou RLM (en anglais : Multiple Linear Regression ; MLR), est considérée parmi les méthodes les plus couramment utilisées dans le développement des modèles prédictifs (**Ammi (2016) ; Bitam (2018) ; Sarkhosh et all (2012)**). Son principe de base repose sur l'existence d'une relation linéaire quantitative entre une variable dépendante désignée parfois par  $Y$  ; représentant la propriété ou bien la réponse à étudier, et un nombre de  $N$  variables indépendantes désignées par  $X_i$  ; parfois ce sont les descripteurs moléculaires dans le cas des modèles de type RQSP/RQSA. La définition mathématique de cette méthode est donnée par l'Eq

$$Y = A_0 + \sum_{i=1}^N A_i * X_i \quad (II.01)$$

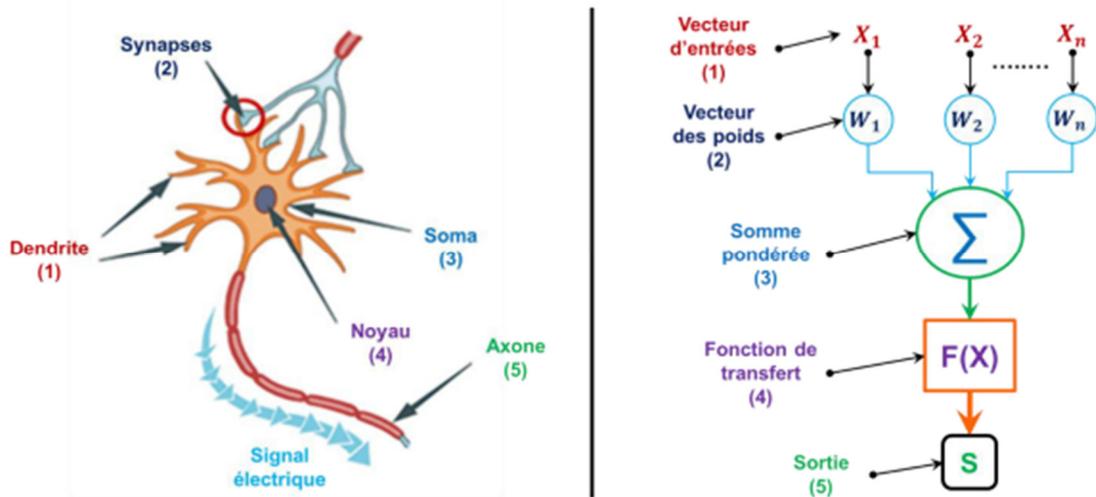
Avec  $Y$  et  $X_i$  : la sortie et les entrées du modèle, respectivement,  $A_i$  représente les coefficients ou bien les effets correspondant à chaque variable  $X_i$ ,  $A_0$  est la constante moyenne du modèle, appelée aussi l'intercepter, et  $N$  est le nombre des entrées du modèle.

## **II.6.2. Réseaux de neurones artificiels**

Les réseaux de neurones artificiels (RNAs) se démarquent comme l'une des technologies les plus avancées des deux dernières décennies, largement utilisées dans une multitude d'applications à travers divers domaines (**Gurney (1997) ; Suzuki et al (2011)**). Un RNA, souvent appelé simplement réseau de neurones, est un modèle mathématique computationnel inspiré par la structure et le fonctionnement des réseaux neuronaux biologiques du cerveau (**Suzuki et al (2011) ; Krose et al (1996)**). Ce système est composé de neurones artificiels interconnectés par des poids synaptiques, des unités de traitement non linéaires. L'apprentissage d'un réseau de neurones artificiels débute par l'ajustement de ces poids pour accomplir des tâches spécifiques (**Suzuki (2011) ; Hagan et al (1996)**).

### **II.6.2.1. Du neurone biologique au neurone artificiel formel**

Le neurone artificiel est un élément de base de tous les réseaux de neurones artificiels. Sa conception et ses fonctionnalités sont dérivées de l'observation d'un neurone biologique qui est la pierre angulaire des systèmes neuronaux biologiques qui comprennent le cerveau, la moelle épinière et les ganglions périphériques (**ROUBEHIE FISSA (2021) ; Gurney (1997) ; Krose et al (1996)**). On peut observer des similitudes dans la conception et les caractéristiques entre les deux côtés de la figure. Du côté gauche, on voit une représentation d'un neurone biologique comprenant son soma, ses dendrites et son axone. À droite, on a un schéma d'un neurone artificiel formel constitué de ses entrées, ses poids, sa fonction de transfert, son biais et ses sorties (**ROUBEHIE FISSA (2021) ; Suzuki(2011) ; Khaouane (2013)**).



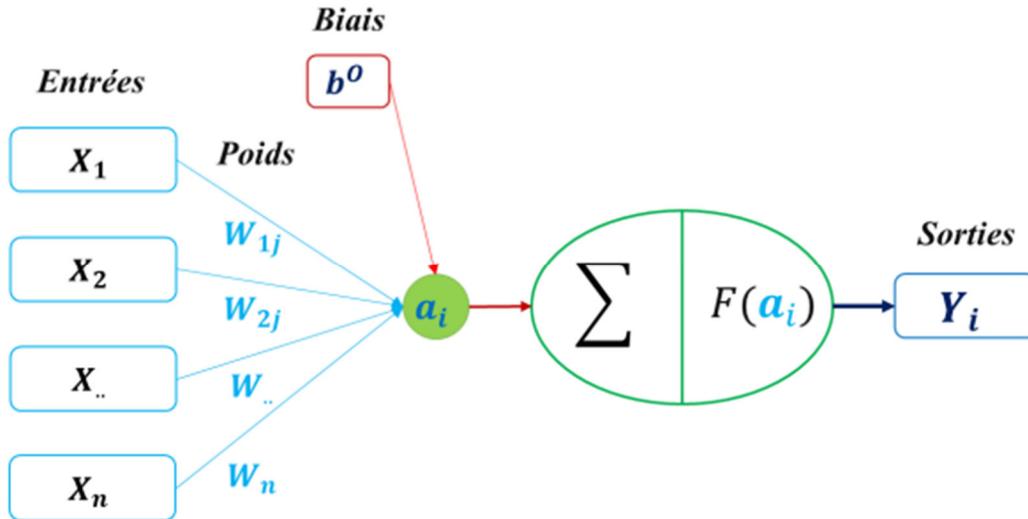
**Figure II.3.** Conception générale de neurones biologiques et artificiels.

Chaque neurone artificiel est comparable à un processeur élémentaire lorsqu'on le compare à un neurone biologique. Il reçoit un nombre variable d'entrées provenant d'autres neurones, et chaque entrée est associée à un poids (weight) qui définit la force de la connexion. Une fois traitées, ces entrées sont agrégées pour produire une sortie unique qui est ensuite transmise à d'autres neurones en aval, formant ainsi un réseau de communication (ROUBEHIE FISSA (2021) ; Hagan et all (1996) ; Khaouane (2013) ; Haykin (1998)).

La représentation mathématique de ce neurone artificiel formel est aussi bien expliquée dans l'Eq [34-77] (ROUBEHIE FISSA (2021) ; Bitam (2018)):

$$Y_i = F(a_i) = F(b^0 + \sum_{i=1}^N w_i * X_i) \quad (II.02)$$

Avec  $b^0$  est un poids effectué à une entrée constante appelée "biais",  $w_i$  est le poids de connexion de la  $i$  ème entrée,  $X_i$  et  $Y_i$  sont les valeurs de  $i$  ème entrée et sortie, respectivement. Une fonction  $F$ , appelée fonction d'activation ou bien fonction de transfert, appliquée sur l'unité  $a_i$  appelé le potentiel du neurone, cette unité calcule la somme pondérée de toutes les entrées (ROUBEHIE FISSA (2021))



**Figure II.4.** Schématisation distinctive d'un neurone artificiel formel ROUBEHIE FISSA (2021).

Comme on le voit à partir du modèle de neurone artificiel et de son Eq (II.2), la principale variable inconnue de notre modèle est sa fonction de transfert. La fonction de transfert est un élément constitutif important pour les réseaux de neurone artificiels et pour n'importe quel modèle mathématique, car elle détermine ses essentielles propriétés. Nous l'avons choisi sur la base du problème neurone artificiel (réseau de neurones artificiels) qui doit résoudre dans la plupart des cas (Suzuki (2011)). Contrairement à des neurones biologiques qui sont dans un état binaire, la plupart des fonctions de transfert sont continues, proposant une infinité de valeurs possibles comprises dans l'intervalle de  $[0, +1]$  ou  $[-1, +1]$  (Si-Moussa (2010)). Il existe de nombreuses formes possibles pour la fonction de transfert. Les Tableaux II.3 et II.4, donnent quelques fonctions de transfert pour les réseaux de neurone artificiels, utilisées respectivement dans le logiciel Statistica (Cybenko (1989); Bishop (1995); Statsoft (2020a); Statsoft. (2020b)) et dans le programme de calcul techniques Matlab (ROUBEHIE FISSA (2021); Suzuki (2011); Si-Moussa (2010)).

**Tableau.II.3.** Fonctions de transfert (d'activation) usuelles dans logiciel Statistica ROUBEHIEFISSA (2021); Bishop (1995); Statsoft (2020a); Statsoft (2020b)

## Chapitre II. Techniques de modélisation neuronal et méthodes de RQSP/RQSA

Nom du fonction	Définition mathématique	Intervalle	Description	Désignation correspondante	dans
Linéaire	$\alpha = n$	$(-\infty, +\infty)$	L'activation du neurone est transmise directement en tant que sortie	Statistica	
Logarithmique Sigmoidale (Logistique)	$\alpha = \frac{1}{1 + e^{-n}}$	$(0, +1)$	Une courbe en forme de S	Identity	
Tangente hyperbolique	$\alpha = \frac{e^{+n} - e^{-n}}{e^{+n} + e^{-n}}$	$(-1, +1)$	Une courbe sigmoïde (tangente hyperbolique) similaire à la fonction logistique. Fonctionne souvent mieux que la fonction logistique en raison de sa symétrie. Idéal pour les perceptrons multicouches, en particulier les couches cachées	Tanh	
Exponentielle	$\alpha = e^{-n}$	$(0, +\infty)$	La fonction exponentielle négative	Exponential	
Sinus	$\alpha = \sin(n)$	$[0, +1]$	Peut-être utile si la reconnaissance des données distribuées radialement	Sine	
Exponentielle Normalisée	$\alpha = \frac{e^{n_i}}{\sum e^{n_i}}$	$[0, +1]$	Principalement utilisé pour (mais sans s'y limiter) les tâches de classification. Utile pour construire des réseaux de neurones avec des sorties multiples normalisées, ce qui le rend particulièrement adapté à la création de classificateurs de réseaux de neurones avec des sorties probabilistes	Softmax	
Gaussienne	$\alpha = \frac{1}{\sqrt{2\pi\sigma}} \exp \left[ -\frac{(x - \mu)^2}{2\sigma^2} \right]$	/	Ce type de fonction d'activation gaussienne isotrope est uniquement utilisé par les unités cachées d'un réseau neuronal RBF, également appelées fonctions de base radiales. L'emplacement (également connu sous le nom de vecteurs prototypes) et les paramètres de propagation sont équivalents aux poids de couche cachés en entrée d'un réseau neuronal MLP	Gaussian	

Avec :  $\mathbf{a} = (n)$ ,  $(n)$  C'est l'entrée nette d'un neurone (Somme pondérée des entrées du neurone pour les MLP). Pour la fonction gaussienne :  $(\mathbf{x})$  Représente le vecteur des entrées de neurones,  $\mu$  ( $\boldsymbol{\mu}$ ) Vecteur des poids d'entrée (centre du RBF),  $\sigma$  ( $\boldsymbol{\sigma}$ ) Propagation (largeur) du RBF

**Tableau.II.4.** Fonctions de transfert usuelles dans le logiciel Matlab [34-85-90].

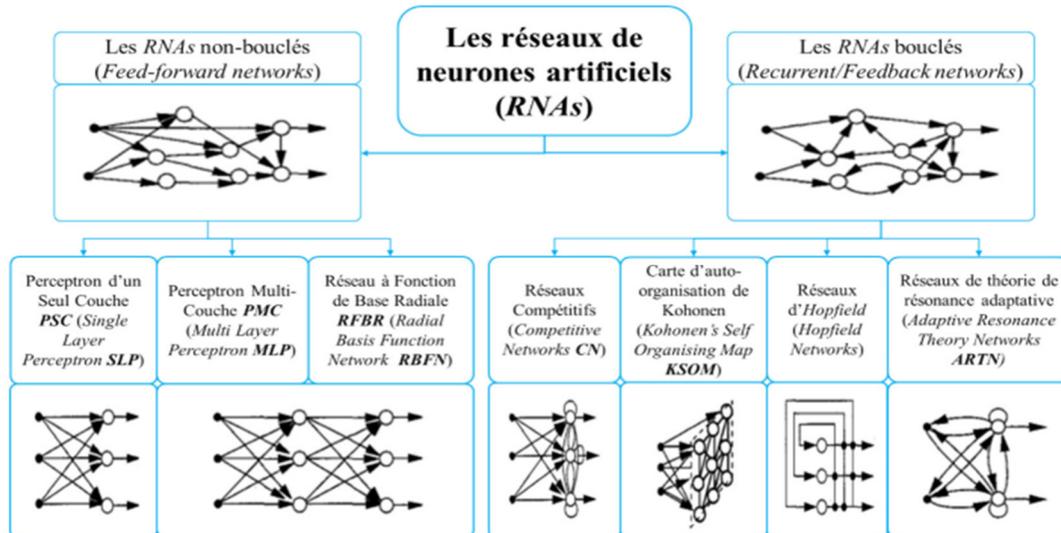
Nom du fonction	Relation entrée/sortie	Graphe de la fonction	Désignation correspondante dans Matlab
Seuil	$a = 0$ Si $n < 0$ $a = 1$ Si $n \geq 0$		<b>hardlim</b>
Seuil symétrique	$a = -1$ Si $n < 0$ $a = 1$ Si $n \geq 0$		<b>hardlims</b>
Linéaire	$a = n$		<b>purelin</b>
Linéaire saturée	$a = 0$ Si $n < 0$ $a = n$ Si $0 \leq n \leq 1$ $a = 1$ Si $n > 1$		<b>satlin</b>
Linéaire saturée symétrique	$a = -1$ Si $n < -1$ $a = n$ Si $-1 \leq n \leq 1$ $a = 1$ Si $n > 1$		<b>satlins</b>
Linéaire positive	$a = 0$ Si $n < 0$ $a = n$ Si $n \geq 0$		<b>poslin</b>
Logarithmique sigmoïde (Logistique)	$a = \frac{1}{1 + e^{-n}}$		<b>logsig</b>
Tangente hyperbolique	$a = \frac{e^{+n} - e^{-n}}{e^{+n} + e^{-n}}$		<b>tansig</b>
Triangulaire de base	$a = 1 -  n , -1 \leq n \leq +1$ $a = 0$ , autrement		<b>tribas</b>
Avec : $a = (n)$ , (n) C'est l'entrée nette d'un neurone (Somme pondérée des entrées du neurone pour les MLP)			

### II.6.2.2. Classification des réseaux de neurones selon leurs architectures

L'interconnexion des neurones artificiels individuels est appelée topologie dans un réseau de neurones artificiels. L'architecture d'un RNA englobe toutes les caractéristiques de base telles que le nombre d'entrées et de sorties, la direction de transmission de l'information, la position des connexions (poids et biais), leur comportement et le mode d'interconnexion entre les neurones. Il existe de nombreuses façons d'interconnecter les neurones, avec de nombreuses topologies possibles. Ces topologies peuvent être regroupées en deux grandes familles de base, comme illustré dans la figure ci-dessous (**ROUBEHIE FISSA (2021) ; Suzuki (2011) ; Dreyfus et all (2002) ; Dreyfus (2005) ; Jain et all (1996)**)

.Le côté droit de la figure représente les réseaux de neurones bouclés), qui ont une topologie récurrente simple (graphe semicyclique), où une partie des informations circulent non seulement dans la direction de l'entrée vers la sortie, mais également dans le sens opposé (**ROUBEHIE FISSA (2021) ; Suzuki (2011) ; Jain et all (1996)**).

Le côté gauche de la figure représente les réseaux de neurones non-bouclés ,qui ont une topologie à action directe simple (graphe acyclique), dans lequel les flux d'informations circulant à partir des entrées vers les sorties dans une seule direction sans retour en arrière (**ROUBEHIE FISSA (2021) ; Dreyfus et all (2002) ; Dreyfus (2005) ; Basheer et all (2000)**)



**Figure II.5.** Classification des réseaux de neurones artificiels selon les architectures D'interconnexions entre neurones (ROUBEHIE FISSA (2021) ; Jain et all (1996) ; Basheer et all (2000))

Nous avons adopté un type de réseau de neurones non-bouclé, un Perceptron Multi Couches (PMC), qui sera expliqué ultérieurement dans la suite

### 1.6.2.3. Classification des réseaux de neurones selon leurs modes d'apprentissage

L'apprentissage est une caractéristique fondamentale des réseaux de neurones, impliquant le calcul des poids des neurones pour rapprocher au mieux les sorties du réseau des sorties souhaitées. La règle d'apprentissage dicte comment les poids du réseau doivent être ajustés précisément entre les cycles d'entraînement successifs, appelés époques. Pour les réseaux de neurones (ROUBEHIE FISSA (2021) ; Basheer et all (2000)), l'apprentissage repose sur l'utilisation d'algorithmes spécifiques. Ces algorithmes modernes modifient les poids des connexions de manière à aligner la réponse du réseau avec les données disponibles, tirées de l'expérience acquise (Chtita (2017) ; ROUBEHIE FISSA (2021) ; Ammar (2007)).

Il existe trois grands paradigmes d'apprentissage ; apprentissage supervisé, apprentissage non-supervisé et apprentissage imposé ou forcé. Habituellement, ils peuvent être employés par n'importe quel type d'apprentissage de réseau de neurone artificiel. Chaque paradigme possède de nombreux algorithmes d'apprentissage (ROUBEHIE FISSA (2021) ; Suzuki (2011)).

L'apprentissage supervisé est une méthode d'apprentissage automatique où un modèle, comme un réseau de neurones artificiels, est entraîné à partir de données d'entrée et de sortie connues. Les données d'entraînement sont des exemples de paires d'entrées et de sorties souhaitées, représentées souvent sous forme de vecteurs de données. Le modèle ajuste ses paramètres en fonction de ces exemples pour pouvoir prédire correctement la sortie correspondante pour de nouvelles entrées (Dearden et all (2009) ; ROUBEHIE FISSA (2021) ; Suzuki (2011) ; Basheer et all (2000) ; Berkane (2010) ; Laidi (2013)).

L'apprentissage renforcé est une technique similaire à l'apprentissage supervisé, mais au lieu de fournir des résultats désirés au réseau, on lui accorde un score mesurant sa performance après plusieurs itérations. Ainsi, le réseau reçoit une évaluation de son exactitude plutôt que la réponse correcte elle-même. Ce type d'apprentissage consiste à donner simplement un retour binaire sur la qualité de la sortie, ce qui entraîne une adaptation sélective des poids des connexions. Les algorithmes utilisant l'apprentissage renforcé sont principalement utilisés dans le domaine des systèmes de contrôle (**ROUBEHIE FISSA (2021) ; Ammar (2007) ; White (1992)**).

L'apprentissage non-supervisé ou bien compétitif est un type d'apprentissage qui se fait seulement sur la base d'informations locales existant à l'entrée (**Samarasinghe (2006) ; Laidi (2009)**). Il représente la capacité des réseaux de neurones à apprendre sans connaître la sortie désirée, les informations présentées à l'entrée provoquent une autoadaptation du réseau afin de produire des valeurs de sortie qui soient proches des réponses à des valeurs d'entrée similaires (**Hamadache (2014) ; ROUBEHIE FISSA (2021) ; Hassoun (1995)**).

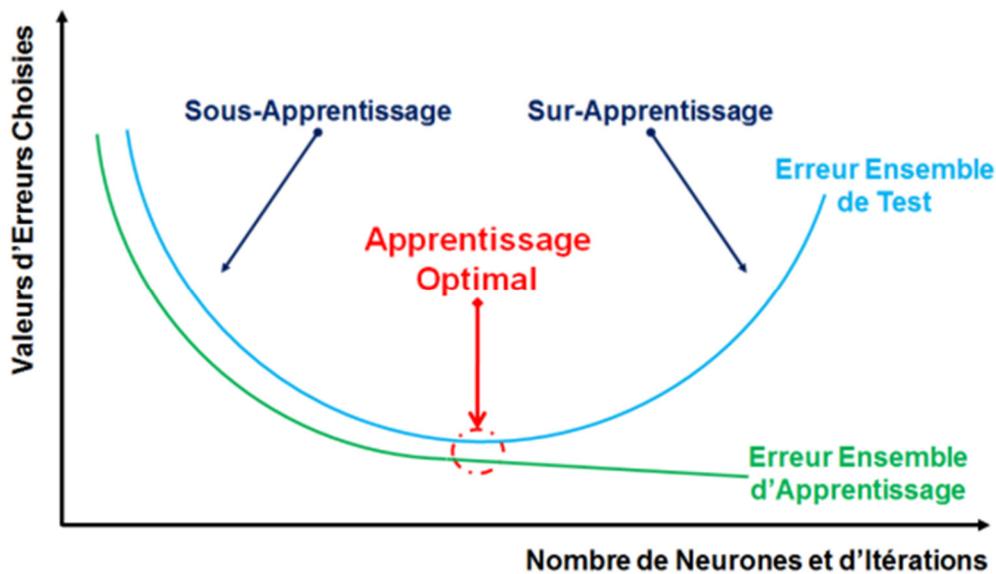
#### **II.6.2.4. Phénomènes de sous-apprentissage et sur-apprentissage**

Le point crucial du développement d'un réseau de neurones est son apprentissage. Il s'agit d'une procédure adaptative par laquelle les connexions des neurones sont ajustées face à une source d'information **Ammar (2007) ; Rumelhart et all (1985)**). Malheureusement, la tâche d'apprentissage n'est pas toujours réussie en raison des obstacles que nous pouvons rencontrer souvent. Les plus célèbres d'entre eux sont deux phénomènes ; le sous-apprentissage et le surapprentissage (**Bishop 1995 ; Haykin 2010 ; Roubéihi Fissa et all 2021**).

. Le phénomène de sous-apprentissage (en anglais : under-learning où parfois under-fitting) est définie comme l'insuffisance d'apprentissage dans un RNA (peu d'apprentissage) à cause de plusieurs raisons et pouvant être dû au très faible nombre d'itérations ou aussi le trop faible nombre de neurones de connexion des couches cachées (Si-Moussa 2010). Ce qui affecte sur la performance du réseau obtenu, de sorte qu'on trouve une erreur très élevée dans les deux sous base d'apprentissage et de test (**ROUBEHIE FISSA (2021) ; Bishop(1995) ; Haykin (2010)**).

Par contre le phénomène de sur-apprentissage (en anglais : over-learning où des fois overfitting) est définie comme un apprentissage exagéré d'un RNA (apprentissage par cœur) (Baughman et all (1995) ; Hawkins (2004)). Cela se produit généralement lorsqu'il y a un trop grand nombre de neurones cachés, qui conduisent à un déséquilibre entre les deux sous-base d'apprentissage et de test en termes de leurs valeurs d'erreurs trouvées, alors qu'on espère les trouver très faibles dans la base d'apprentissage (résultat souhaitable) et trop élevés dans la base de test (résultat non-souhaitable) (Hamadache (2014); ROUBEHIE FISSA (2021) ; Si-Moussa (2010)).

La Figure montre bien les différences entre les trois phénomènes d'apprentissage ; le sous-apprentissage, le sur-apprentissage et enfin l'apprentissage optimale ou bien la généralisation optimale. Ce dernier est l'état désiré pour n'importe modèle de réseau de neurone artificiel pendant leur processus d'apprentissage.



**Figure II.6.** Phénomènes d'apprentissage optimale, sous-apprentissage et sur-apprentissage.

En résumé, plus on augmente le nombre de neurones dans la couche cachée, meilleur sera l'apprentissage. Cependant, baser le choix du modèle uniquement sur les performances de l'ensemble d'apprentissage peut conduire à un risque de sur-apprentissage. Pour éviter ce problème, il est essentiel de contrôler continuellement les performances des deux sous-ensembles (d'apprentissage et de validation) simultanément, de manière à ce qu'elles soient comparables et minimisées (Hamadache (2014); ROUBEHIE FISSA (2021))

Dans la plupart des situations, il n'y a pas une façon pour déterminer le nombre adéquat de neurones cachés sans faire l'apprentissage de plusieurs réseaux et d'évaluer ensuite l'erreur de généralisation de chaque réseau (**ROUBEHIE FISSA, (2021) ; Si-Moussa (2010)**).

La sous-utilisation de neurones cachés peut entraîner des erreurs d'apprentissage et de généralisation élevées en raison du sous-apprentissage. En revanche, une surutilisation de neurones cachés peut conduire à une erreur d'apprentissage très faible mais une erreur de généralisation élevée, car cela peut prolonger excessivement la phase d'apprentissage, introduisant ainsi un grand nombre de paramètres ajustables qui peuvent affecter négativement les performances globales du réseau neuronal. **ROUBEHIE FISSA (2021) ; Si-Moussa (2010) ; Ammar (2007)**).

Afin d'éviter tous les problèmes de sous-apprentissage et de sur-apprentissage, on a adopté une stratégie basée sur l'équilibre entre le nombre de neurones de connexion dans les trois couches ; entrée, cachée et sortie, et puis la taille de la base des données utilisées, qui sera expliquée dans le chapitre III (Matériels et méthodes).

#### **II.6.2.5. Algorithme d'apprentissage :**

L'apprentissage des réseaux de neurones artificiels implique l'ajustement des connexions neuronales pour améliorer les performances du réseau en fonction des informations reçues. Les algorithmes d'apprentissage, tels que ceux utilisant la rétro propagation, modifient les poids des connexions pour que les sorties du réseau correspondent aux exemples expérimentaux. L'algorithme BFGS, une méthode d'optimisation Quasi-Newtonienne (**Ammar (2007) ; Dreyfus et all (2008) ; Urbani, (1995) ; Oussar (1998) ; Lucea (2006) ; SIGMA. (2020)**), est utilisé dans cette étude pour ajuster progressivement les poids du réseau et réduire l'écart entre les prédictions et les données réelles.

Généralement, l'approximation par la méthode "Quasi-Newtonienne" n'est efficace que si elle est appliquée au voisinage d'un minimum. Il est donc recommandé d'utiliser une autre méthode comme la méthode du gradient simple ou du gradient stochastique, au début de l'apprentissage, vu qu'ils sont efficaces lorsqu'on est loin du minimum et sa convergence ralentit considérablement lorsqu'on s'approche du

minimum ; de mettre, ensuite en œuvre la méthode de BFGS lorsqu'on estime être suffisamment proche d'un minimum (Si-Moussa (2010) ; Ammar (2007) ; Dreyfus et al (2008)).

#### **II.6.2.6. Réseaux de neurones perceptron multicouche :**

Le Perceptron Multi-Couches (PMC), ou Multi Layer Perceptron (MLP) en anglais, est un type de réseau de neurones non-bouclés appartenant à la famille des réseaux de neurones à propagation avant (**Feed-forward Networks**). Dans un PMC, les neurones sont organisés en couches consécutives où l'information circule d'une seule direction, de la couche d'entrée à la couche de sortie. Chaque neurone dans une couche est connecté à tous les neurones de la couche précédente et de la couche suivante, mais pas à ceux de la même couche.

Le réseau PMC est composé de trois types de couches distinctes :

1. La couche d'entrée : Cette couche contient les informations relatives à l'entrée du réseau de neurones, telles que des données structurales, des caractéristiques qualitatives ou quantitatives, ou des facteurs liés aux sorties souhaitées.
2. Les couches cachées : Ces couches peuvent être une ou plusieurs et contiennent des neurones connectés aux informations de la couche d'entrée en vue de prédire les réponses désirées de la couche de sortie.
3. La couche de sortie : Cette couche est composée d'un ou plusieurs neurones en fonction du nombre de réponses souhaitées. Ces neurones fournissent les valeurs de sortie en combinant linéairement les fonctions de transfert des couches cachées.

**La Figure II.7**, illustre l'architecture d'un réseau de neurone perceptron multicouche non-bouclé avec une couche d'entrée, une couche cachée et une seule couche de sortie.

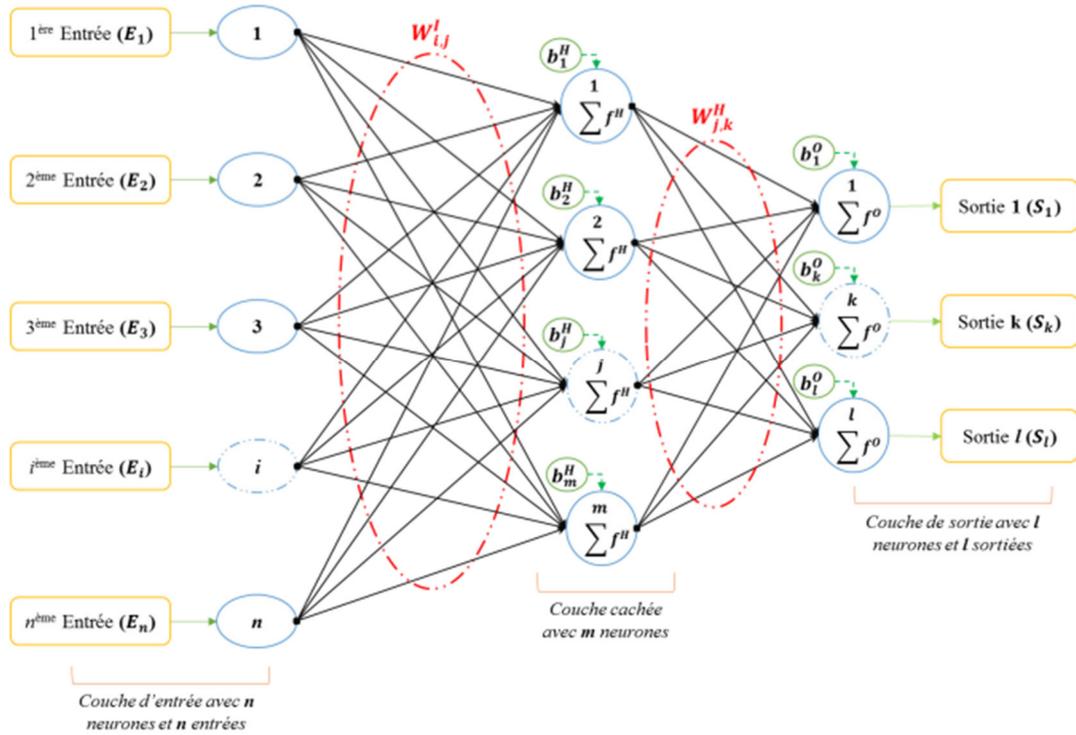


Figure II.7. Structure d'un perceptron multi couches non-bouclé (Roubehie Fissa et all (2023)).

Les neurones sont reliés entre eux par des connexions pondérées ; ce sont les poids de ces connexions qui dirigent le fonctionnement du réseau et programment une application de l'espace des entrées vers l'espace des sorties à l'aide de la transformation non linéaire (Hamadache (2014) ; Ammi (2016) ; Khaouane (2013)). Le modèle mathématique neuronal final de l'exemple précédent est donné par l'Eq (II.)

$$S_k = f^O(\sum_{j=1}^m w_{kj}^H * f^H(\sum_{i=1}^n w_{ij}^I E_i + b_j^H) + b_k^O) \quad (II.3)$$

Où  $S_k$  la sortie du  $k$  ème neurone ( $k = 1, 2, \dots, l$ ) ;  $l$  est le nombre de neurones dans la couche sortie, alors que  $E_i$  est l'entrée du  $i$  ème neurone ( $i = 1, 2, \dots, n$ ) ; avec  $n$  est le nombre de neurones dans la couche entrée,  $w_{kj}^H$  est le poids de la couche cachée associé au  $j$  ème et au  $k$  ème neurone,  $w_{ij}^I$  est le poids de la couche entrée associé au  $i$  ème et au  $j$  ème neurone ( $j = 1, 2, \dots, m$ ) ; avec  $m$  le nombre de neurones dans la couche cachée,  $b_j^H$  biais de la couche cachée connectée avec le  $j$  ème neurone de la même couche cachée,  $b_k^O$  biais de la couche sortie connecté avec le  $k$  ème neurone de la même couche sortie,  $f^H$  fonction d'activation (ou de

transfert) de la couche cachée,  $f$   $O$  fonction d'activation (ou de transfert) de la couche sortie.

L'architecture d'un modèle neuronal final dépend en effet des fonctions d'activation choisies et du nombre de neurones utilisé dans chacune des trois couches : d'entrée, cachées et de sortie. Dans les deux derniers chapitres, je vais expliquer en détail toutes les expressions mathématiques relatives aux cinq paramètres étudiés dans ce thème.

## **II.7. Analyses et validation des modèles**

Les meilleurs modèles de RQSP-RNA dans cette étude ont été évalués en termes de quatre types d'erreurs différents ainsi que de performances de régression. Les erreurs incluent la racine d'erreur quadratique moyenne (REQM), l'erreur standard de prédiction (ESP), l'erreur relative moyenne (ERM) et l'erreur relative absolue moyenne (ERAM). Les performances de régression ont été évaluées à l'aide des coefficients de régression ( $R$ ), de détermination ( $R^2$ ), de détermination ajustée ( $R_{adj}^2$ ), ainsi que le vecteur de régression comprenant les valeurs de la pente ( $\alpha$ ) et de l'interception ( $\beta$ ). Ces évaluations ont été réalisées sur les ensembles de données de test, d'apprentissage et sur l'ensemble global de la base de données.

## **II.8. Conclusion :**

Il semble que vous vous référeriez à un passage ou un chapitre d'un document académique ou d'un thème traitant des méthodes RQSP/RQSA, de la méthodologie de développement, de l'analyse et de la validation des modèles, ainsi que de la quantification des informations moléculaires à l'aide de différents descripteurs moléculaires. Le texte mentionne également l'utilisation de méthodes d'apprentissage statistique telles que les réseaux de neurones artificiels, en mettant particulièrement l'accent sur les perceptrons multicouches. Il est discuté en détail de l'architecture d'un réseau neuronal, de ses paramètres de base tels que les fonctions d'activation, le processus d'apprentissage et d'autres phénomènes connexes. Il semble s'agir d'une section clé du travail, visant à expliquer en profondeur les notions essentielles liées au développement de ces modèles.

---

## *Chapitre III*

# *Modélisation des propriétés critiques des hydrocarbures par les RQSP-RNA*

---

### **III.1. Introduction**

Généralement, le développement des modèles de type Relation Quantitatives Structures Propriétés (RQSP) suit trois étapes clés : au départ, le choix des propriétés à étudier telles que les propriétés critiques ou d'autres propriétés physicochimiques. Ensuite, l'extraction des informations structurales des molécules et la sélection des meilleurs descripteurs moléculaires pertinents. Enfin, l'utilisation des méthodes d'apprentissage mathématique ou numérique pour relie les deux premières étapes (**Roubehie Fissa et all 2023 ; Hamadache et all 2018**).

Les propriétés critiques : la température critique ( $T_c$ ), la pression critique ( $P_c$ ), et le volume critique ( $V_c$ ) ; assurent un rôle fondamental dans l'ingénierie chimique et d'autres domaines liés aux l'industrie pétrolière (**Riazi, M. R. 2005 ; Fahim et all 2009**). En outre, ces propriétés sont également considérées comme des paramètres clés dans la production, la séparation et le traitement des matières énergétiques, telles que les hydrocarbures et les fractions pétrolières (**Roubehie Fissa et all 2023**).

Le but de cette étude était de développer des modèles de type RQSP-RNA pour prédire trois paramètres physicochimiques qui sont les propriétés critiques :  $T_c$ ,  $P_c$  et  $V_c$  des hydrocarbures purs. Environ 400 molécules d'hydrocarbures purs ont été utilisées comme base de données. Des milliers de descripteurs ont été calculés initialement afin d'extraire les informations structurales de ces molécules. En suit une sélection a été réalisée en utilisant une méthodologie statistique pour conserver que les descripteurs moléculaires pertinents. Enfin, la relation entre les structures représentées par les descripteurs moléculaires et les propriétés a été lié par un RNA-PMC. Toutes les étapes citées sont appliquées pour développer les modèles de RQSP-RNA. Les meilleurs modèles ont été testés avec divers types d'erreur et paramètres statistiques afin de prédire :  $T_c$ ,  $P_c$  et  $V_c$  des hydrocarbures purs avec une bonne précision.

### **III.2. Matériels et méthodes**

#### **III.2.1. Outils informatiques utilisés**

Divers logiciels et outils informatiques ont été utilisés pendant la réalisation de ce travail, tels qu'ils sont illustrés dans la **Figure III.01**, et dans les parties ci-dessous :

- L'organisation des données des hydrocarbures et certaines étapes de prétraitement a été réalisé par le logiciel "Microsoft Office Excel 2010".

- La préparation des notifications SMILES ont été faite par le logiciel "ChemDraw Office version 20.1" (ChemOffice 2012).
- Concernant les calculs des descripteurs moléculaires et l'optimisation des molécules, on a utilisé le logiciel "Dragon v.7" intégré sur la platform "OCHEM" (Sushko 2011).
- Certaines étapes de réduction des descripteurs et principalement l'apprentissage des modèles neuronaux ont été exécutés par le logiciel "Statistica v. 14 (TIBCO Statistica 14.0.0 (2020)).
- À propos des erreurs des modèles neuronaux développées le logiciel "MATLAB R2020b" a été utilisé (The MathWorks Inc. USA 2020) pour calculer quelques erreurs et paramètres statistiques ainsi pour tracer certains graphes.

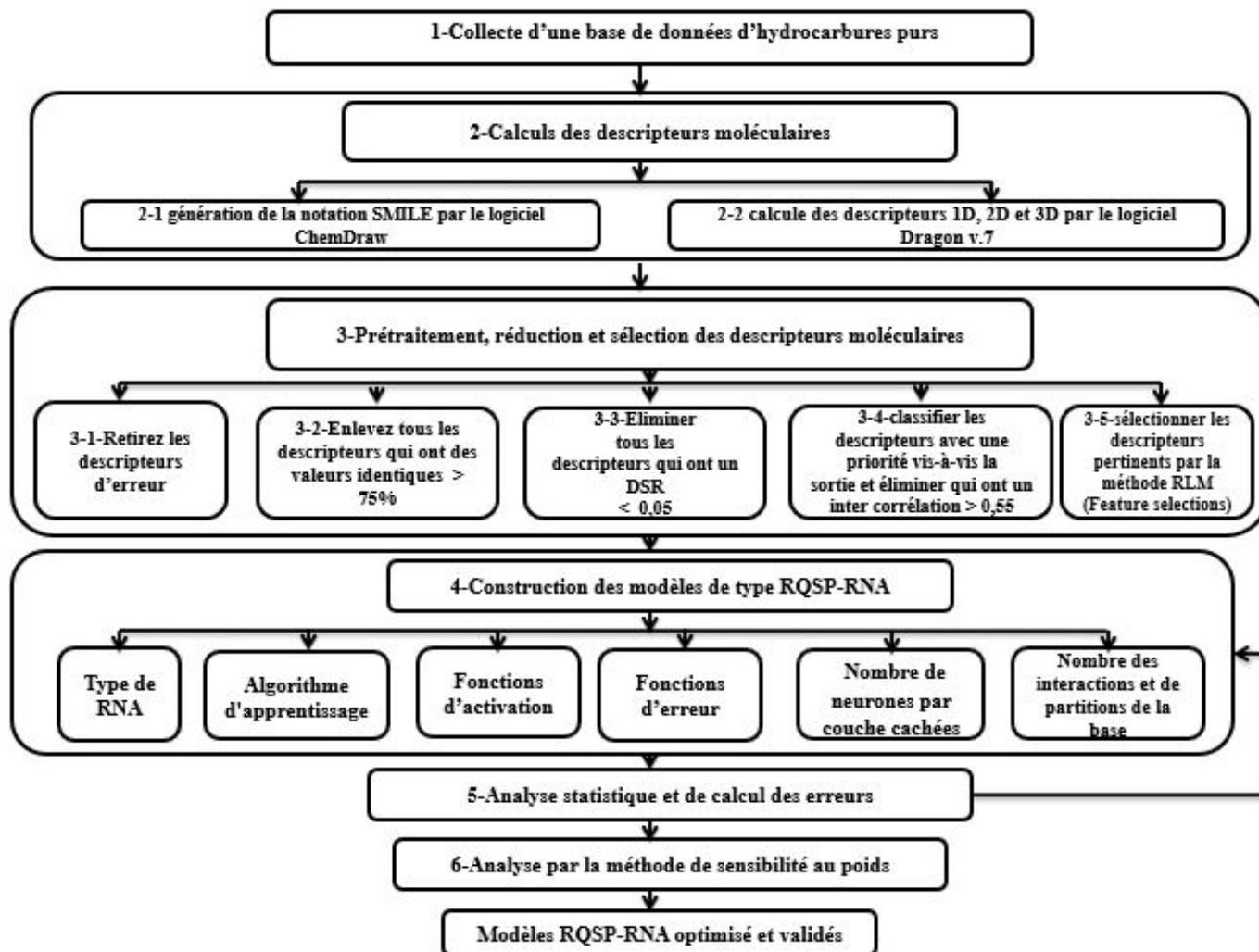
Il convient de noter que tous les calculs de ce travail ont été effectués à l'aide d'un ordinateur personnel Hp : doté d'un processeur Intel(R) Core(TM) i5-6300U CPU @ 2.40GHz 2.50 GHz de CPU et de 8,00 Go de mémoire vive (RAM), avec Windows 10 Professionnel comme système d'exploitation.



Figure III.01. Capture d'écran des logiciels utilisés dans cette étude : Microsoft Office Excel, ChemDraw Office, MATLAB, OCHEM Dragon V. 7 et, TIBCO Statistica

### III.2.2. Méthodologie générale de la modélisation

Plusieurs étapes ont été utilisées dans le développement des modèles de type RQSP-RNA. Cependant, la méthodologie adoptée dans ce travail se divise en six étapes, résumée dans l'organigramme de la **Figure III.02**



**Figure III.2.** Organigramme générale de la méthodologie de développement suivies dans cet travail.

#### III.2.2.1. Collection et préparation de la base des données

La première étape concerne la préparation des données ; qui sont concéderez comme la pierre angulaire pour le développement d'un modèle de qualité. De ce fait la base de données

expérimentale est collectée à partir de la littérature (Yaws et al 2003). Elle comprend 400 molécules d'hydrocarbures purs réparties en différentes classes : saturées et insaturées, qui représentent les principaux groupes d'importance industrielle tels que les n-paraffines, les iso-paraffines, les oléfines simples et aromatiques. L'objectif est de prédire trois paramètres critiques : la température critique ( $T_c$ ) exprimé en **K**, la pression critique ( $P_c$ ) en **bar** et le volume critique ( $V_c$ ) en **cm<sup>3</sup>/mole**. La plupart de ces catégories comprennent plusieurs isomères de position tels que : *cis*, *trans*, *méta*, *ortho* et *para*. Voici un exemple où une partie de ces données se trouve dans le tableau III.1.

**Tableau III.1.** Partie d'une 20 molécule de la base des données utilisées dans ce travail (Yaws et al 2003).

N°	Name	Formula	Tc (K)	Pc (bar)	Vc (cm <sup>3</sup> /mol)
1	acenaphthene	C <sub>12</sub> H <sub>10</sub>	803,15	31	553
2	acetylene	C <sub>2</sub> H <sub>2</sub>	308,32	61,39	113
3	allene	C <sub>3</sub> H <sub>4</sub>	393,15	54,7	162
4	alpha-methylstyrene	C <sub>9</sub> H <sub>10</sub>	654	33,6	423
5	alpha-phellandrene	C <sub>10</sub> H <sub>16</sub>	649	28,2	500
60	decylbenzene	C <sub>16</sub> H <sub>26</sub>	753	17,7	881
61	decylcyclohexane	C <sub>16</sub> H <sub>32</sub>	751,25	16,5	858
62	decylcyclopentane	C <sub>15</sub> H <sub>30</sub>	730,64	16,29	816,5
63	1-decylcyclopentene	C <sub>15</sub> H <sub>28</sub>	718,6	16,77	802,5
64	1-decylphthalene	C <sub>20</sub> H <sub>28</sub>	838,67	15,06	969,5
178	3-ethyl-3,4-dimethylhexane	C <sub>10</sub> H <sub>22</sub>	624,5	24,22	511
179	3-ethyl-3-hexene	C <sub>8</sub> H <sub>16</sub>	565,63	26,87	464,5
180	1-ethyl-3-isopropylbenzene	C <sub>11</sub> H <sub>16</sub>	665,63	26,3	537,5
181	4-ethyl-3-methylheptane	C <sub>10</sub> H <sub>22</sub>	614,3	22,8	530
182	3-ethyl-3-methylheptane	C <sub>10</sub> H <sub>22</sub>	620	23,1	532
396	2,2,4,6,6-pentamethylheptane	C <sub>12</sub> H <sub>26</sub>	629,3	18,52	679,5
397	2,2,3,4,5-pentamethylhexane	C <sub>11</sub> H <sub>24</sub>	630,42	20,09	622,5
398	2,2,3,3,5-pentamethylhexane	C <sub>11</sub> H <sub>24</sub>	629,72	20,15	623,5
399	propylene	C <sub>3</sub> H <sub>6</sub>	364,76	46,13	181
400	4-propylheptane	C <sub>10</sub> H <sub>22</sub>	601	21,78	545

Pour plus de transparence et clarté sur la base de données ; le **tableau III.2** présente les valeurs moyennes, ainsi que les bornes maximale et minimale des données pour les trois propriétés critiques étudiées.

**Tableau III.2.** Description quantitatives et qualitatives de la base des données.

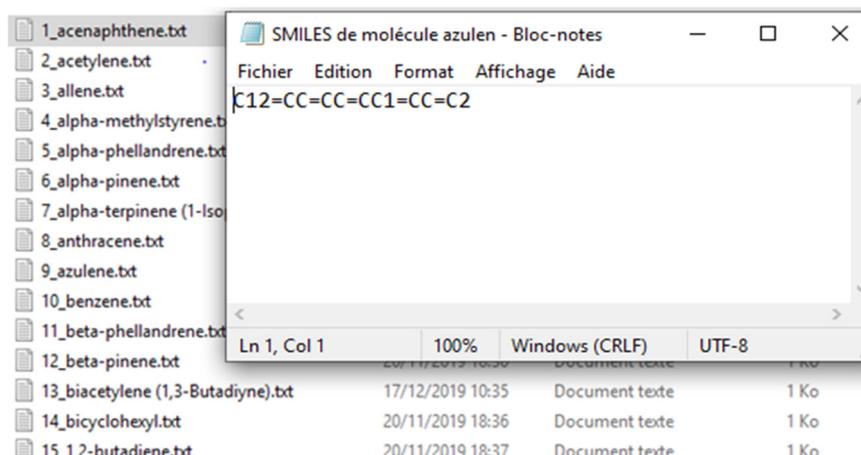
Paramètre	Valeurs minimale	Valeurs moyen	Valeurs maximale	Nombre des molécules
Tc (°K)	190,58	661,8356	3711,11	400
Pc (bar)	1,75	26,2081	61,39	
Vc (cm <sup>3</sup> /mol)	99,3	610,188	3731,5	

#### *III.2.2.2. Méthode de génération de descripteurs moléculaires*

Le calcul des descripteurs moléculaires est une étape très importante dans la création des modèles de type RQSP, car cela permet de convertir les structures chimiques des molécules étudiées en données numériques, facilitant ainsi leur représentation.

Dans cette partie, deux des étapes principales ont été suivies pour calculer les descripteurs.

1. La première est la génération des notations SMILES à partir du logiciel ChemDraw qui permet de traduire les structures chimiques en langage informatique. Chaque notation SMILES est ensuite enregistrée dans un fichier texte, avec son nom et son numéro d'identification dans la base de données, comme elle indique la **Figure III.03**.
2. La deuxième est les calculs des descripteurs moléculaires en utilisant le logiciel Dragon v.7. implanter dans la Platform OCHEM Sushko. Les nombres de descripteurs calculés et leurs classe correspondantes est illustrée dans le **tableau.III.3**,



**Figure III.3.** Capture d'écran du fichier texte de la molécule azulène.

**Tableau. III.3.** Nombres de descripteurs initialement calculés par chaque classe correspondante

N° de classe	Nom des classes		Nombre des descripteurs
	"En anglais"	"Traduisez en français"	
(1)	Constitutional descriptors	Descripteurs constitutionnels	(47)
(2)	Topological indices	Indices topologiques	(75)
(3)	Connectivity indices	Indices de connectivité	(37)
(4)	2D matrix-based descriptors	Descripteurs matriciels 2D	(607)
(5)	Burden eigenvalues	Valeurs propres du fardeau	(96)
(6)	ETA indices	Indices ETA	(23)
(7)	Geometrical descriptors (3D)	Descripteurs géométriques (3D)	(38)
(8)	3D autocorrelations (3D)	Autocorrélations 3D (3D)	(80)
(9)	3D-MORSE descriptors (3D)	Descripteurs 3D-MORSE (3D)	(224)
(10)	GETAWAY descriptors (3D)	Descripteur ESCAPADE (3D)	(273)
(11)	Functional group counts (3D)	Nombre de groupes fonctionnel (3D)	(154)
(12)	Atom-type E-state indices	Indices d'état E de type atome	(172)
(13)	2D Atom pairs	Paires d'atome 2D	(1596)
(14)	Charge descriptors (3D)	Descripteur de frais (3D)	(15)
(15)	Drug-like indices	Indices de type drogue	(28)
(16)	Ring descriptors	Descripteurs d'anneaux	(32)
(17)	Walk and path counts	La marche et la chemin comptent	(46)
(18)	Information indices	Indices d'information	(50)
(19)	2D auto correlation	Auto corrélations 2D	(213)
(20)	P_VSA-like descriptors	Descripteur de type P_VS A	(55)

(21)	Edge adjacency indices	Indices de contiguïté des bords	(324)
(22)	3D matrix-based descriptors (3D)	Descripteurs basés sur une matrice 3D	(99)
(23)	RDF descriptors (3D)	Descripteurs RDF 3(D)	(210)
(24)	WHIM descriptors (3D)	Descripteurs WHIM (3D)	(114)
(25)	Randic molecular profiles (3D)	Profils moléculaires Randic (3D)	(41)
(26)	Atom-centred fragments	Fragments centrés sur l'atome	(115)
(27)	CATS 2D	CHATS 2D	(150)
(28)	3D Atom pairs (3D)	Paires d'atomes 3D (3D)	(36)
(29)	Molecular properties	Propriétés moléculaires	(20)
(30)	2 CATS 3D (3D)	2 CHATS 3D (3D)	(300)
<b>Nombre total des descripteurs moléculaires</b>		<b>5270</b>	

Un nombre considérable de 5720 descripteurs ont été calculés, ce qui donne une quantité significative à prendre en compte lors de la création des modèles de RQSP-RNA. Il est largement admis que la plupart de ces descripteurs ne sont pas directement liés aux paramètres que l'on cherche à modéliser.

Il est essentiel de réduire ce nombre à un niveau acceptable en sélectionnant les descripteurs les plus pertinents pour chaque paramètre. Afin d'atteindre nos objectifs, nous avons mis en place un processus visant à réduire les descriptions les plus significatives pour les trois propriétés critiques à étudier.

### ***III.2.2.3. Procédure de prétraitement et sélection des descripteurs***

L'un des défis de cette étude réside dans le choix des éléments les plus appropriés pour chaque paramètre à analyser. Le processus de prétraitement des données vise à éliminer les caractéristiques qui ne sont pas pertinentes. Cela implique la réduction de la dimension des variables d'entrée du modèle en supprimant celles qui ont peu d'impact sur la variable de sortie à modéliser.

Nous avons employé une approche statistique pour réduire et sélectionner le nombre de descripteurs, en vue de les intégrer par la suite dans la création des modèles RQSP-RNA. Cette approche s'appuie sur les principaux éléments énoncés dans l'étude de **(Roubehie Fissa et al 2019)**.

- (1) Nettoyer tous les descripteurs avec une valeur d'erreur en laissant une case vide après avoir effectué le calcul à l'aide du logiciel en ligne "Dragon.V.7".

- (2) Supprimer les descripteurs d'une même valeur répétés plus de 75%, en comparant les trois premiers quartiles.
- (3) calcul la déviation standard relative (RSD) et éliminez ceux dont les valeurs sont inférieures à 0,05. Les étapes suivantes ont été effectuées en utilisant le logiciel "STATISTICA".
- (4) Pour sélectionner uniquement les descripteurs pertinents, nous avons supprimé ceux dont le coefficient d'inter corrélation (R) entre chaque paire de descripteurs était supérieur à 0,55. Les calculs des coefficients ont été effectués avec MATLAB et l'analyse comparative et la suppression ont été réalisées dans Excel 2010.
- (5) Enfin, nous avons utilisé une approche pas à pas de régression linéaire multiple, connue sous le nom de régression MLR (forward stepwise) ou par autre nom (Feautere selection ), pour éliminer les descripteurs avec une valeur-p supérieure à 0,005. Ce processus a été effectué par le logiciel "STATISTICA". A la fin seulement on a retenus les meilleures 20 descripteurs

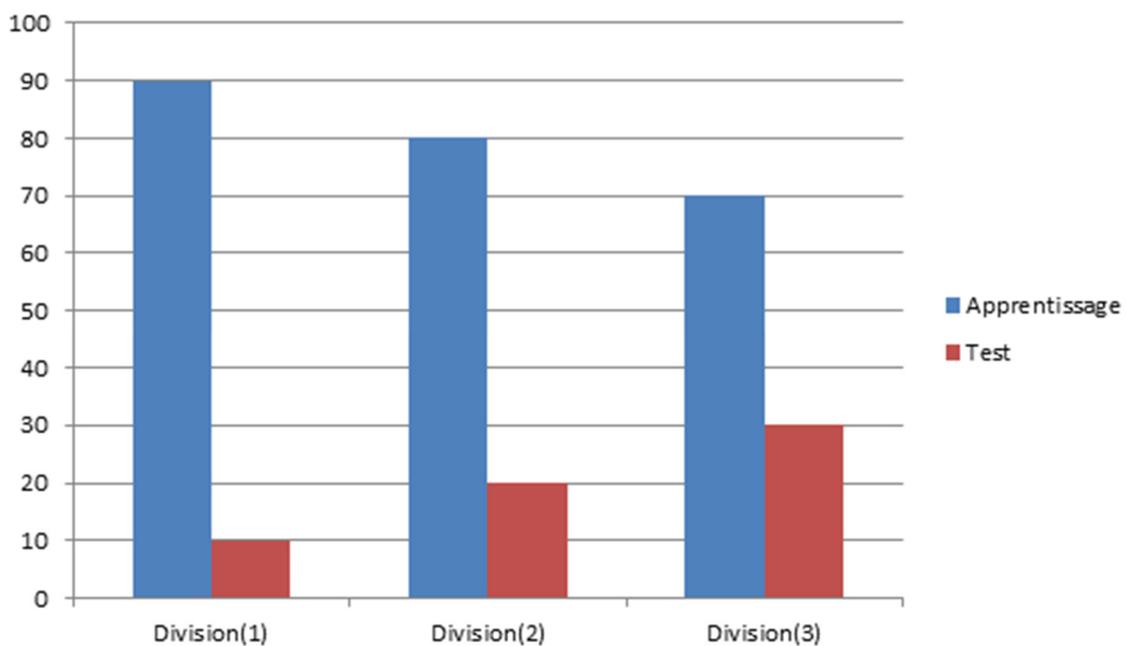
#### *III.2.2.4. Choix des paramètres pour le développent des modèles de type RQSP-RNA*

Après avoir effectué le prétraitement des données, j'ai réduit au minimum le nombre de descripteurs pour obtenir uniquement les plus pertinents pour chaque paramètre. Cela m'a permis de les utiliser de manière efficace lors de la construction des modèles de Réseau de Neurones Artificiels (RNA) pour la Régression Quantile sur les Prix (RQSP). Ce processus a constitué un défi supplémentaire en raison du grand nombre de choix de paramètres dans les RNA. Afin de relever ce défi, j'ai exploré différentes options pour obtenir les meilleurs modèles de RQSP-RNA pour chaque paramètre (**Roubehie Fissa et all 2019**).

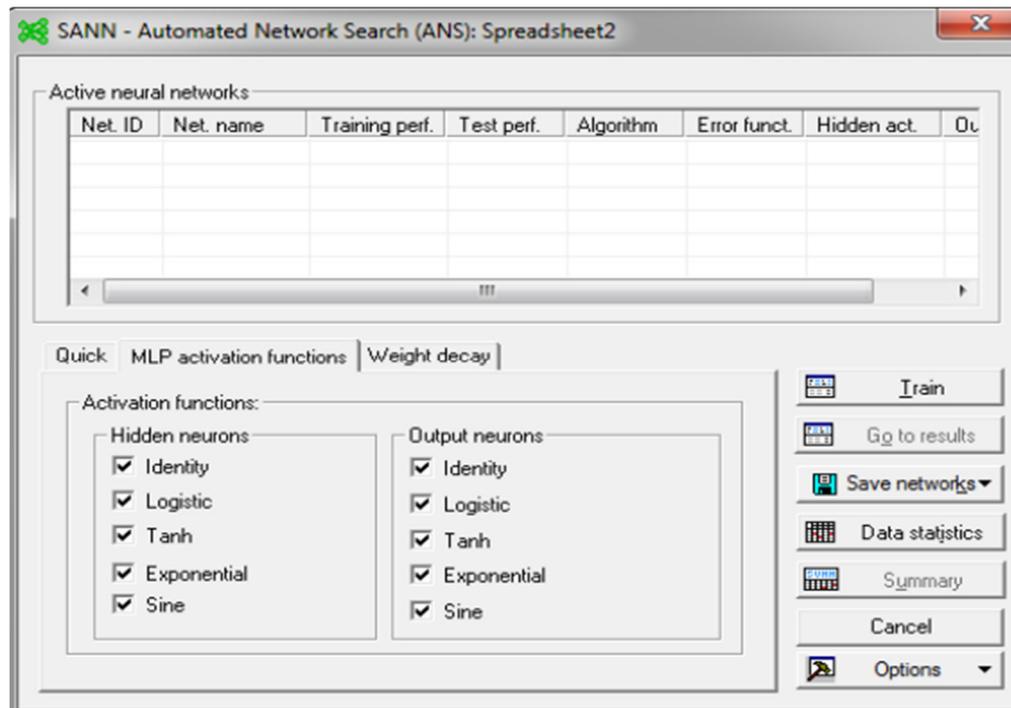
Les paramètres du RNA sont déterminés en se basant sur des études antérieures et en utilisant le logiciel "STATISTICA". Ainsi, nous allons suivre ces étapes pour choisir les meilleurs paramètres :

- L'ensemble de données a été divisé au hasard en deux sous-ensembles l'un pour l'apprentissage et l'autre pour le test ; avec trois partition différents pour chaque propriétés critique par la manière suivante : 70% et 30% pour la phase d'apprentissage et de test, de même pour la deuxième partition avec 80% et 20% pour chaque phase, enfin avec 90% et 10% pour la troisième partition (**Figure III.4**).

- Le perceptron multicouche (PMC) est choisi comme le type de réseau de neurones utilisé dans cette étude en raison de sa popularité pour résoudre des problèmes d'approximation et de prédiction.
  - L'algorithme Quasi Newton à propagation en arrière (Quasi Newton back propagation algorithm, plus précisément : Broyden-Fletcher-Goldfarb-Shanno algorithm *BFGS*) a été utilisé comme un algorithme d'apprentissage (**Wessel et al 1994**).
  - Cinq fonctions d'activation distinctes sont appliquées à chaque couche cachée et de sortie. Ainsi, à la fin du processus, le logiciel identifie la combinaison optimale des deux fonctions d'activation relatives à la couche en question (**Figure III.5**).
  - La somme des carrés (Sum of Squares **SOS**) a été choisie en tant que fonction d'erreur.
  - Dans ce travail, une seule couche cachée est employée, avec une variation de 2 à 30 neurones dans cette couche. Ceci est fait pour garantir que les caractéristiques des modèles de réseaux de neurones restent cohérentes avec la taille des données.
  - Trois nombre d'exécutions différents ont été effectués pour chaque propriété dans chaque partition; qui sont 2000, 4000 et 6000 itérations. Pour obtenir les meilleurs modèles de type RQSP-RNA dans chaque paramètre de sortie.



**Figure III.4.** Représentation graphique de division de la base des données.



**Figure III.5.** Capture d'écran pour le choix des fonctions d'activation sur STATISTICA

### *III.2.2.5. Analyses statistiques et calculs des erreurs des meilleurs modèles développés*

Dans cette étude, les performances des régressions et la précision des modèles de type RQSP-RNA ont été évaluées à l'aide des quatre types d'erreur, et deux coefficients statistiques en plus de deux autre paramètre. Le **Tableau III.4.** Est montres ensembles de ces grandeurs de performances utilisées :

**Tableau III.4.** Désignation de grandeurs de performances.

Désignations	Nom de grandeurs de performances
REQM	Racine d'Erreur Quadratique Moyenne
ESP	Erreur Standard de Prédiction
ERM	Erreur Relative Moyenne
ERAM	Erreur Relative Absolue Moyenne
R	Coefficient de Régression
R <sup>2</sup>	Coefficient de Détermination
$\alpha$	La pente
$\beta$	Interception

Les définitions mathématiques de l'ensemble des paramètres de performance sont données par les **Equations (III.01–III.06)**, respectivement.

$$\mathbf{REQM} = \sqrt{\sum_{i=1}^N (y_i^{\text{exp}} - y_i^{\text{cal}})^2 / N} \quad (\text{III.01})$$

$$\mathbf{ESP} = (\mathbf{REQM} / \hat{y}^{\text{exp}}) * 100 \quad (\text{III.02})$$

$$\mathbf{ERM} = (100/N) * \sum_{i=1}^N (|(y_i^{\text{exp}} - y_i^{\text{cal}}) / y_i^{\text{cal}}|) \quad (\text{III.03})$$

$$\mathbf{ERAM} = (100 / N) * \sum_{i=1}^N (|(y_i^{\text{exp}} - y_i^{\text{cal}}) / y_i^{\text{exp}}|) \quad (\text{III.04})$$

$$\mathbf{R} = \frac{\sum_{i=1}^N [(y_i^{\text{exp}} - \hat{y}^{\text{exp}})(y_i^{\text{cal}} - \hat{y}^{\text{cal}})]}{\sqrt{\sum_{i=1}^N (y_i^{\text{exp}} - \hat{y}^{\text{exp}})^2} \sqrt{\sum_{i=1}^N (y_i^{\text{cal}} - \hat{y}^{\text{cal}})^2}} \quad (\text{III.05})$$

$$\mathbf{R}^2 = 1 - \sum_{i=1}^N (y_i^{\text{exp}} - y_i^{\text{cal}})^2 / \sum_{i=1}^N (y_i^{\text{exp}} - \hat{y}^{\text{exp}})^2 \quad (\text{III.06})$$

$$\hat{y}^{\text{cal}} = \sum_{i=1}^N y_i^{\text{cal}} / N \quad (\text{III.07})$$

$$\hat{y}^{\text{exp}} = \sum_{i=1}^N y_i^{\text{exp}} / N \quad (\text{III.08})$$

Où  $N$  est le nombre des composés dans l'ensemble des données pour chaque phase (apprentissage, test, ou l'ensemble entier des données),  $k$  étant le nombre de variables du modèle (dans ce cas c'est le nombre de descripteurs sélectionnés),  $y^{\text{cal}}$  est la valeur calculée (prédite) de  $y$  ; pour l'ensemble d'apprentissage ou du test,  $y^{\text{exp}}$  est la valeur expérimentale (observée) de  $y$ ,  $\hat{y}^{\text{cal}}$  la moyenne des valeurs calculées de  $y$  ; (**Eq (III.7)**), et  $\hat{y}^{\text{exp}}$  la moyenne des valeurs expérimentales de  $y$  ; (**Eq (III.8)**).

**III.2.2.6. Mesure de la contribution des descripteurs des modèles RQSP-RNA développées**

- L'analyse de sensibilité est une méthode essentielle dans l'évaluation des modèles mathématiques, permettant de valider, d'étudier le comportement des variables et de confirmer le choix des variables en évaluant leur contribution. C'est particulièrement crucial dans le cas des modèles RQSP-RNA, où cette analyse joue un rôle clé dans la validation et l'optimisation des résultats **Wessel et all (1994)**.

Dans cette étude, une analyse de sensibilité basée sur la méthode des "Poids" a été employée pour évaluer l'influence de chaque variable d'entrée, représentée par les descripteurs moléculaires pertinents dans notre cas, sur les paramètres critiques ( $T_c$ ,  $P_c$  et  $V_c$ ) de chaque modèle RQSP-RNA. Cette évaluation est cruciale pour valider notre sélection des descripteurs employés dans la création des modèles RQSP-RNA, démontrant ainsi l'efficacité de notre approche de prétraitement.

- La procédure de cette étude, connue sous le nom de méthode de Poids, a été soigneusement planifiée et mise en œuvre à l'aide de "Microsoft Excel". Les étapes pratiques décrites par **Goh Gevrey et all (2003)**, dans la réalisation de cette analyse. Comme l'indique le nom de la méthode "Poids", on introduit les poids de connexion de partitionnement en : **(1)** Poids de la couche entrée ( $w_{ij}^I$ ) (connexions entre les neurones d'entrée et cachés), **(2)** Poids de la couche cachée ( $w_{1j}^H$ ) (connexions entre les neurones cachés et de sortie). Pour trouver, à la fin une signification quantifiée de chaque descripteur d'entrée sur chaque sortie du modèle, ce qui se traduit par l'importance relative (**IR**) de chaque descripteur (**Khaouane, L. (2013) ; Wessel et all (1994)**).

Pour mieux expliquer cette méthode j'ai repris l'exemple d'application utilisé dans l'étude de **Goh Goh, A. T. (1995)**, et l'étude de Gevrey et ses co-auteurs **Goh Gevrey et all (2003)**. L'exemple est donné pour un réseau de neurones d'architecture composée de : trois **(3)** neurones dans la couche entrée, quatre **(4)** neurones dans la couche cachée et un **(1)** neurone dans la couche sortie, le processus de calcul est basé sur les quatre étapes suivantes

- 1<sup>ère</sup> étape : Pour chaque neurone caché ( $i$ ) ; multiplier la valeur absolue des poids de connexions cachée–sortie  $|W_{1j}^H|$  par la valeur absolue des poids de connexions cachée–entrée  $|W_{1j}^i|$ . Faire ceci pour chaque variable d’entrée ( $j$ ).
- 2<sup>ème</sup> étape : Pour chaque neurone caché ( $i$ ) ; diviser chaque ( $P_{ij}$ ) par la somme de toutes les variables d’entrée pour obtenir ( $Q_{ij}$ ). Par exemple pour le neurone caché 01,  
$$Q_{11} = \frac{P_{11}}{P_{11}+P_{12}+P_{13}} = 0,26445.$$
- 3<sup>ème</sup> étape : Pour chaque neurone d’entrée ( $j$ ), additionner les résultats des produits ( $Q_{ij}$ ). obtenus par l’étape précédente, afin d’obtenir les sommes ( $S_j$ ). Par exemple :  $S_1 = Q_{11} + Q_{21} + Q_{31} + Q_{41}$ .
- 4<sup>ème</sup> étape : Deviser ( $S_j$ ) par la somme de toutes les variables d’entrée, exprimé en pourcentage, cela donne l’importance relative ou bien la contribution des variables d’entrée données. Par exemple : pour le neurone d’entrée 01, l’importance relative ( $IR$ ) égale à :  $(S_1 * 100) / (S_1 + S_2 + S_3) = 43,7\%$ .

### III.3. Résultats et discussions

#### III.3.1. Évolution du nombre de descripteurs au cours de la procédure de sélection

Selon la section (III.2.2.3) de ce chapitre, la méthode de sélection utilisée dans cette étude repose sur une approche statistique comprenant plusieurs étapes distinctes, chacune présentant ses propres caractéristiques. Chaque étape influence le traitement préliminaire et affecte donc le nombre de descripteurs à choisir. Le **tableau III.5** illustre la variation du nombre de descripteurs au cours du prétraitement pour chaque paramètre analysé.

**Tableau III.5.** Évolution du nombre de descripteurs après chaque étape de prétraitement pour les trois paramètres étudiés.

<i>Nom des classes de descripteurs</i>	<i>(0)</i>	<i>(1)</i>	<i>(2)</i>	<i>(3)</i>	<i>(4)</i>			<i>(5)</i>		
					<i>Pour Tc</i>	<i>Pour Pc</i>	<i>Pour Vc</i>	<i>Pour Tc</i>	<i>Pour Pc</i>	<i>Pour Vc</i>
<i>Nombre total des descripteurs</i>	5270	5270	2188	2188	<u>833</u>	<u>827</u>	<u>833</u>	<u>420</u>	<u>565</u>	<u>506</u>

*NB : L'étape (0) indique le nombre des descripteurs initiaux calculés et les étapes (1–5) sont indiquées dans la partie (III.2.2.3) de ce chapitre.*

Après la dernière étape de prétraitement, les descripteurs pertinents pour chaque paramètre ont été obtenus sont : **420** descripteurs pour Tc, **565** descripteurs pour Pc et **506** descripteurs pour Vc, les valeurs P de chaque descripteur et leurs désignations sont indiquées dans le **tableau III.6**

**Tableau III.6.** Caractéristique des descripteurs retenus pour les trois paramètres étudiés.

<i>Paramètres</i>	<i>Pour Tc</i>		<i>Pour Pc</i>		<i>Pour Vc</i>	
	<i>Nom_Des</i>	<i>Valeur_P</i>	<i>Nom_Des</i>	<i>Valeur_P</i>	<i>Nom_Des</i>	<i>Valeur_P</i>
Entrée 01	VE2_L	6,022 10 <sup>-222</sup>	BIC0	2,41 10 <sup>-189</sup>	VE2_Dz(Z)	1,08 10 <sup>-188</sup>
Entrée 02	Mor05m	2,67 10 <sup>-125</sup>	SpMaxA_EA(bo)	8,47 10 <sup>-119</sup>	Vx	3,00 10 <sup>-168</sup>
Entrée 03	Mor23p	4,41 10 <sup>-101</sup>	SpMin4_Bh(p)	4,48 10 <sup>-95</sup>	SpMin4_Bh(i)	2,11 10 <sup>-106</sup>
Entrée 04	SpMin6_Bh(s)	4,62 10 <sup>-94</sup>	VE2_D	7,4 10 <sup>-93</sup>	SpMin4_Bh(p)	1,88 10 <sup>-104</sup>
Entrée 05	Mor30u	3,03 10 <sup>-86</sup>	SpMin6_Bh(v)	1,16 10 <sup>-88</sup>	SpMin5_Bh(p)	2,18 10 <sup>-104</sup>
Entrée 06	Mor09m	1,85 10 <sup>-85</sup>	SpMin3_Bh(p)	8,27 10 <sup>-80</sup>	SpMin3_Bh(p)	4,22 10 <sup>-96</sup>
Entrée 07	G2u	3,78 10 <sup>-85</sup>	SpMin3_Bh(i)	9,46 10 <sup>-79</sup>	SpMin7_Bh(m)	4,44 10 <sup>-84</sup>
Entrée 08	Mor30i	1,59 10 <sup>-81</sup>	SpMin5_Bh(p)	8,64 10 <sup>-78</sup>	R5s+	1,62 10 <sup>-82</sup>
Entrée 09	Mor24v	7,74 <sup>-63</sup>	JGI5	5,09 10 <sup>-72</sup>	JGI5	1,81 10 <sup>-82</sup>
Entrée 10	Wi_Dt	5,44 10 <sup>-51</sup>	SM6_D/Dt	8,263 10 <sup>-68</sup>	Wi_Dz(Z)	8,39 10 <sup>-82</sup>
Entrée 11	H8u	1,25 10 <sup>-45</sup>	SpMin5_Bh(m)	1,39 10 <sup>-66</sup>	JGI8	1,54 10 <sup>-79</sup>
Entrée 12	R6u+	8,40 10 <sup>-43</sup>	SpMin4_Bh(i)	9,65 10 <sup>-66</sup>	R2i+	1,60 10 <sup>-77</sup>
Entrée 13	HATS2p	6,43 10 <sup>-38</sup>	H4e	2,08 10 <sup>-60</sup>	SpMin3_Bh(i)	9,52 <sup>-71</sup>
Entrée 14	SM15_AEA (ri)	7,49 10 <sup>-37</sup>	JGI3	2,56 10 <sup>-60</sup>	Mor05m	1,70 <sup>-69</sup>
Entrée 15	SM14_AEA (ri)	1,63 10 <sup>-33</sup>	Wi_Dz(Z)	2,16 10 <sup>-56</sup>	SM14_AEA(ri)	1,73 <sup>-59</sup>
Entrée 16	Mor10m	8,94 10 <sup>-32</sup>	H5i	1,54 10 <sup>-55</sup>	HATS3v	4,07 <sup>-59</sup>
Entrée 17	SpMAD_G/D	3,29 10 <sup>-31</sup>	GGI8	1,07 10 <sup>-53</sup>	JGI3	2,70 <sup>-58</sup>
Entrée 18	SpMin3_Bh(e)	2,29 10 <sup>-29</sup>	HATS1i	3,65 10 <sup>-53</sup>	L1i	9,87 <sup>-57</sup>
Entrée 19	SpMin6_Bh(p)	1,38 10 <sup>-28</sup>	SpMin3_Bh(v)	2,54 10 <sup>-51</sup>	Tm	4,48 <sup>-55</sup>
Entrée 20	SM14_AEA (dm)	1,67 10 <sup>-28</sup>	ATS7m	1,16 10 <sup>-49</sup>	L1v	2,34 <sup>-53</sup>
<b>Totale</b>	<b>P (420_Descripteurs) &lt; 0,0000</b>		<b>P (565_Descripteurs) &lt; 0,0000</b>		<b>P (506_Descripteurs) &lt; 0,0000</b>	

D'après les résultats présentés dans le **tableau III.6**, on constate que tous les descripteurs choisis pour les trois propriétés critiques sont statistiquement significatifs.

### III.3.2. Architectures des meilleurs modèles RQSP-RNA obtenus

Après avoir passé par les étapes de réduction, construction et calcul, seuls les modèles les plus performants ont été retenus pour évaluer leurs performances en analysant différentes erreurs et paramètres statistiques. Les différentes architectures des meilleurs modèles pour chaque paramètre, accompagnées de certains paramètres statistiques essentiels, sont répertoriées de manière structurée dans le **tableau III.7**.

Les modèles RQSP-RNA démontrent des valeurs de SDC satisfaisantes pour chacun de leurs paramètres, tant dans les ensembles d'apprentissage, de test que de validation. De plus, les coefficients de régression R affichent des performances élevées pour les trois modèles étudiés.

**Tableau III.07.** Architectures des meilleurs modèles obtenus pour chaque paramètre.

Couches	Nombre de neurones	Fonctions d'activation	Valeurs de R et SDC	
<b>Pour le paramètre Tc</b>			$R^{App}$	0,9976
Couche d'entrée	20	<i>(Aucune)</i>	$R^{Tes}$	0,9986
Couche cachée	14	Tanh	$SDC^{App}$	36 ,6837
Couche de sortie	1	Identity	$SDC^{Tes}$	178 ,0269
Algorithme d'App	BFGS 203			
<b>Pour le paramètre Pc</b>			$R^{App}$	0,9895
Couche d'entrée	20	0,9767	$R^{Tes}$	0,9767
Couche cachée	28	<i>Exponential</i>	$SDC^{App}$	0,6101
Couche de sortie	1	Logistic	$SDC^{Tes}$	1 ,7819
Algorithme d'App	BFGS 116			
<b>Pour le paramètre Vc</b>			$R^{App}$	0,9990
Couche d'entrée	20	<i>(Aucune)</i>	$R^{Tes}$	0,9988
Couche cachée	16	Exponential	$SDC^{App}$	124,9625
Couche de sortie	1	Identity	$SDC^{Tes}$	209,5224
Algorithme d'App	BFGS 83			

NB : Apprentissage (*App*), Test (*Tes*), Validation (*Val*), Broyden-Fletcher-Goldfarb-Shanno (**BFGS**)

Comme indiqué dans le **tableau III.6**, les architectures des meilleurs modèles de type sont (20-14-1) pour le modèle de Tc, (20-28-1) pour le modèle de Pc, et (20-16-1) pour le modèle de Vc. Donc, les réseaux comportent de 20, 20 et 20 entrées pour Tc, Pc et Vc,

respectivement. Avec une sortie ( $S$ ) pour chaque modèle. Pour les couches cachées ; 14, 28 et 16 neurones sont apparus respectivement dans les mêmes modèles.

Les deux fonctions d'activations les plus fréquemment utilisées dans les trois modèles étudiés est l'exponentielle "Exponential", et identity qui a été mentionnée deux fois. Ensuite, les fonctions d'activation suivantes ont été retrouvées dans un seul modèle chacune : tangente hyperbolique "Tanh", et la sigmoïde logarithmique "Logistic". Les définitions mathématiques des fonctions d'activations apparues dans cette étude sont données dans le **tableau III.8**. (Voir les Eqs (1), (2), (3) et (4)) StatSoft 2007; Beale et al (2012)

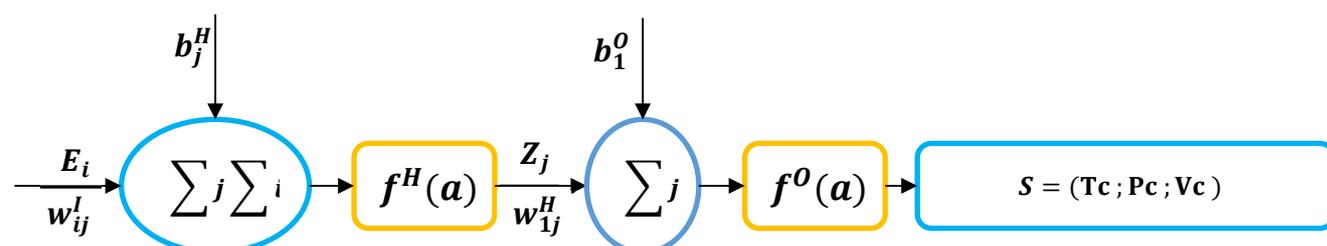
**Tableau III.8.** Fonctions d'activation apparues dans cette étude StatSoft 2007; Beale et al (2012)

	Fonctions d'activation			
	Exponentielle	Tangent Hyperbolique	Sigmoïde Logarithmique	Linéaire
<b>Notations dans STATISTICA</b>	<b>Exponential</b>	<b>Tanh</b>	<b>Logistic</b>	<b>Identity</b>
<b>Notations dans Matlab</b>	<b>exp</b>	<b>tanh</b>	<b>Logis</b>	<b>identity</b>
<b>Formules Mathématiques</b>	$a = e^{-n}$ (1)	$a = \frac{e^{+n} - e^{-n}}{e^{+n} + e^{-n}}$ (2)	$a = \frac{1}{1 + e^{-n}}$ (3)	$a = n$ (4)

*(e):* La fonction exponentielle; *(n):* Les variables des modèles de RNA (Voir les Eqs. (III.10) à (III.15)).

### III.3.3. Application des modèles obtenus dans la prédiction des propriétés critiques des hydrocarbures

Pour prédire les trois propriétés critiques des hydrocarbures à l'aide des modèles RQSP-RNA développés dans cette étude, il est essentiel de comprendre le contenu mathématique de chaque modèle RQSP-RNA obtenu.



**Figure III.6.** Représentation schématique d'un réseau du modèle de type RQSP-RNA.

Tous les modèles que nous allons étudier suivent la structure illustrée dans la **figure III.6**. La seule variation entre ces modèles réside dans le nombre de neurones présents dans la couche d'entrée. Pour chaque modèle, nous devons définir les descripteurs moléculaires utilisés en entrée ainsi que le nombre de neurones dans la couche cachée. Les équations (III.10 à III.15) décrivent mathématiquement les sorties de la couche cachée ( $Z_j$ ) et de la couche de sortie ( $S$ ) pour chaque paramètre.

$$Z_j = f^H \left[ \sum_{i=1}^{20} W_{ij}^I E_i + b_j^H \right] = \frac{e^{+\left[\sum_{i=1}^{20} W_{ij}^I E_i + b_j^H\right]} - e^{-\left[\sum_{i=1}^{20} W_{ij}^I E_i + b_j^H\right]}}{e^{+\left[\sum_{i=1}^{20} W_{ij}^I E_i + b_j^H\right]} + e^{-\left[\sum_{i=1}^{20} W_{ij}^I E_i + b_j^H\right]}} \quad (\text{III.10})$$

$$S = f^O \left[ \sum_{i=1}^{14} W_{1j}^H Z_j + b_1^O \right] = \sum_{i=1}^{14} W_{1j}^H Z_j + b_1^O \quad (\text{III.11})$$

$$Z_j = f^H \left[ \sum_{i=1}^{20} W_{ij}^I E_i + b_j^H \right] = e^{-\left[\sum_{i=1}^{20} W_{ij}^I E_i + b_j^H\right]} \quad (\text{III.12})$$

$$S = f^O \left[ \sum_{i=1}^{28} W_{1j}^H Z_j + b_1^O \right] = \frac{1}{1 + e^{-\left[\sum_{i=1}^{28} W_{1j}^H Z_j + b_1^O\right]}} \quad (\text{III.13})$$

$$Z_j = f^H \left[ \sum_{i=1}^{20} W_{ij}^I E_i + b_j^H \right] = e^{-\left[\sum_{i=1}^{20} W_{ij}^I E_i + b_j^H\right]} \quad (\text{III.14})$$

$$S = f^O \left[ \sum_{i=1}^{16} W_{1j}^H Z_j + b_1^O \right] = \sum_{i=1}^{16} W_{1j}^H Z_j + b_1^O \quad (\text{III.15})$$

Chaque neurone de la couche cachée traite les informations reçues via une fonction d'activation. Le résultat de ce calcul, noté ( $Z_j$ ), (avec j allant de 1 à 14-28 selon le modèle), est ensuite transmis au neurone de sortie. Ce dernier synthétise ces informations pour donner la réponse finale du réseau, notée ( $S$ ). Pour ce qui est du calcul précis du signal de sortie de chaque neurone caché ( $Z_j$ ), il diffère selon le modèle utilisé : les équations (III.10), (III.12) et (III.14) s'appliquent respectivement aux modèles Tc, Pc et Vc. De la même manière, la sortie finale du réseau ( $S$ ) est déterminée par l'équation (III.11) pour le modèle Tc, l'équation (III.13) pour le modèle Pc et l'équation (III.15) pour le modèle Vc.

Avec  $f^H$  la fonction d'activation de la couche cachée,  $f^O$  la fonction d'activation de la couche de sortie,  $S$  la sortie du réseau (paramètres de sortie : Tc, Pc et Vc),  $Z$  le signal de sortie

des neurones cachés,  $W_{ij}^I$  les poids de la couche d'entrée (connexions entre les neurones d'entrée et cachés),  $W_{ij}^H$  Les poids de la couche cachée (connexions entre les neurones cachés et de sortie),  $E_i$  les neurones d'entrée (dans le cas actuel, les descripteurs moléculaires pertinents),  $b_j^H$  sont les biais de la couche cachée (biais des neurones cachés),  $b_1^0$  est le biais de la couche de sortie (biais du neurone de sortie),  $i$  le nombre d'entrées pour chaque modèle ( $i = 1$  à  $20$  ; pour les modèles de Tc, Vc et Vc), et  $j$  est le nombre de neurones cachés ( $j = 1$  à  $14$  ; pour le modèle de Tc,  $j = 1$  à  $28$  ; pour le modèle de Pc, et  $j = 1$  à  $16$  ; pour le modèle de Vc).

Les valeurs des poids ( $w_{ij}^I, w_{ij}^H$ ) et des biais ( $b_j^H, b_1^0$ ) des trois meilleurs modèles RQSP-RNA pour chaque couche cachée et sortie, sont données respectivement dans : le tableau (III.9) pour le modèle Tc, dans le tableau (III.10) pour le modèle de Pc, dans le tableau (III.11) pour le modèle de Vc.

**Tableau III.9** Poids et biais du meilleur modèle RQSP-RNA pour Tc (20-14-1)

Connexion des couches : (Caché- Sortie)		Connexion des couches : (Entrée-Caché)																				Poids	Biais	Connexion des couches : (Caché- Sortie)
		$W_{j1}^1$	$W_{j2}^1$	$W_{j3}^1$	$W_{j4}^1$	$W_{j5}^1$	$W_{j6}^1$	$W_{j7}^1$	$W_{j8}^1$	$W_{j9}^1$	$W_{j10}^1$	$W_{j11}^1$	$W_{j12}^1$	$W_{j13}^1$	$W_{j14}^1$	$W_{j15}^1$	$W_{j16}^1$	$W_{j17}^1$	$W_{j18}^1$	$W_{j19}^1$	$W_{j20}^1$			
-0,47130	0,59370	-0,73993	0,21731	0,21731	0,23175	-1,36487	-0,94892	-0,36777	0,46677	-0,24519	-0,38790	0,65392	-0,20320	0,42936	1,21737	-1,32082	-0,48572	-0,52995	-0,59133	-0,70752	0,68529			
2,43132	0,38830	-0,08459	-0,31679	-0,17173	0,92288	-0,10012	-0,10664	0,12650	-9,84581	-0,25561	0,06335	0,71125	0,18211	0,00315	-0,48517	-0,11180	-0,10559	0,01435	0,14136	-1,59457	-0,62780			
-0,76193	0,53139	-0,33261	-0,25641	-0,38662	-0,21027	0,03488	0,26015	0,58253	0,01510	0,17764	-0,37299	-0,43945	0,48558	0,12394	0,06009	-0,93884	0,11223	0,54803	-0,46516	-0,55101	-0,59776			
-0,48859	1,26050	-0,16278	0,09160	0,17349	-0,05623	0,17079	0,96988	-0,20612	2,07153	0,01162	-0,27441	2,97478	0,49335	-0,08372	0,02266	-0,31249	0,19501	-0,19917	0,05648	-1,85017	0,30242			
0,73268	-0,65977	-0,00081	-0,27551	-0,13517	-0,26418	0,79818	-0,49393	-0,05458	-0,72536	0,11617	0,29094	-0,44224	-0,04585	-0,23286	-0,31903	0,01762	0,08922	-0,17332	0,30632	0,28557	0,04474			
0,90261	-1,05102	-0,58418	-0,03710	0,59750	-0,03714	-3,14517	0,25031	0,15333	-0,40371	0,31177	0,19399	-2,01541	-1,30044	-0,38916	0,29587	0,56820	0,40208	-1,04565	-1,04775	1,98199	-0,03835			
0,91942	-0,05151	-0,01601	0,31658	0,17134	0,11509	0,35788	-0,39049	-0,39692	0,35674	0,02781	-0,25958	-0,38121	-0,04601	0,25100	0,23221	0,45993	-0,24264	0,88899	-0,13934	-0,20235	0,34578			
-0,55015	-0,11389	0,14932	0,13087	0,08175	-0,59259	0,85987	-0,08991	-0,27485	0,38761	-0,33435	-0,32917	0,18284	-0,93733	1,05703	-0,48746	-0,43719	1,45403	-0,25724	-0,19255	-1,57166	0,43017			
1,92592	-1,94134	1,38542	0,50537	1,16649	1,18187	-1,23010	0,74793	-0,72569	2,88694	1,16362	1,74098	-2,50653	-2,03358	-0,64872	1,36344	-0,74496	-0,03444	-0,76683	0,09313	0,99204	1,53133			
0,09049	-0,86722	0,41951	1,10097	0,23826	0,93783	-0,80219	0,78606	0,09208	0,36929	0,59049	-0,79689	0,15103	-0,53196	-0,57085	-2,69291	-0,80817	1,10825	-0,14467	-1,32187	2,91030	-0,13173	0,34959		
-2,31533	0,64477	-0,11500	0,38136	0,47525	-0,00796	0,40153	0,27364	-0,28536	0,29489	0,06654	-0,43463	1,71227	0,09266	0,11207	0,25277	-0,02201	0,08796	0,16473	-0,33644	-0,24937	-0,26995			
-0,75890	0,00867	-0,03596	0,38271	0,00113	0,36481	0,08007	-0,22248	-0,69548	0,29434	-0,12068	0,37334	-0,05261	-0,52224	-0,24998	0,13661	0,67566	0,36967	-0,44320	0,55726	0,14551	-0,20106			

0,99942	-0,97848	-0,05943	0,64957	0,78761	-0,40700	-1,43132	0,39355	0,13376	-3,05669	-0,75554	-0,04893	0,59885	0,29797	-0,10039	-1,07030	0,20765	2,65428	-1,07624			
-1,66448	1,73445	0,06984	-0,15974	0,04092	-0,28157	0,92310	-0,17358	-0,59261	0,64080	-0,22566	-0,76709	2,55197	1,66548	0,26508	-0,21498	-0,24897	-0,52320	0,80227	0,28952	-0,44676	-0,13527

Tableau III.10 Poids et biais du meilleur modèle RQSP-RNA pour Pc (20-28-1).

Connexion des couches : (Entrée-Caché)																					
Connexion des couches : (Caché-Sortie)																					
Poids										Biais											
$W_{j1}^1$	$W_{j2}^1$	$W_{j3}^1$	$W_{j4}^1$	$W_{j5}^1$	$W_{j6}^1$	$W_{j7}^1$	$W_{j8}^1$	$W_{j9}^1$	$W_{j10}^1$	$W_{j11}^1$	$W_{j12}^1$	$W_{j13}^1$	$W_{j14}^1$	$W_{j15}^1$	$W_{j16}^1$	$W_{j17}^1$	$W_{j18}^1$	$W_{j19}^1$	$W_{j20}^1$	$b_j^H$	$b_1^0$
0,99509	0,40050	-0,60907	0,36600	-0,61692	-0,03305	-0,09941	-0,69380	0,07868	-0,41003	0,49163	0,01948	-1,05218	-0,09714	-0,13236	-0,24106	0,23155	0,06694	-0,03555	0,40051	-0,03322	0,40803
-0,80154	-3,16499	1,54143	-0,92765	0,06647	0,41026	2,86264	0,49033	0,62397	-2,87265	0,51927	0,65753	0,09662	0,47412	0,82197	-0,32848	-1,45396	-3,19910	0,29681	-0,85521	-0,93396	0,59472
-1,68972	-0,66627	0,27942	-0,98317	-0,53003	0,02156	-0,30593	-0,33384	-0,56568	-0,40118	0,18163	0,26844	-0,22837	0,01612	0,04947	-0,12782	0,12580	0,50575	-0,65425	0,68707	0,46551	-0,86107
1,70519	-0,11992	-0,48284	-0,45843	-0,73554	0,16921	-0,12247	-0,52560	-0,34205	0,20604	0,48660	-0,21179	-0,93669	-0,43963	0,02522	-0,37399	0,09633	-0,68505	0,09230	-0,10205	-0,24378	0,65155
-0,62680	0,01006	-0,07340	-0,38529	-0,26821	0,10295	-0,05035	0,04517	-0,15206	0,01877	-0,01660	-0,10414	-0,16316	-0,05539	-0,03078	-0,18161	0,07099	-0,15051	0,16669	-0,03445	-0,02560	-0,41607

Chapitre III. Modélisation des propriétés critiques des hydrocarbures par les RDSP-RNA

1,81944	-0,26297	-0,63852	-1,50151	0,18040	-0,72491	0,03603	0,33670	-0,35957	-0,04458	0,72667	-0,07993	-0,22061	0,72688	0,65474	-0,19293	-0,75306	-1,30060	0,35607	-0,52946	-0,17263	1,13201
-0,55674	-0,12191	0,02855	-0,27482	-0,28983	0,01351	-0,00669	0,01776	-0,21542	-0,11830	-0,00188	-0,07818	-0,19661	-0,07548	-0,06534	-0,14726	0,09987	-0,15287	0,10613	0,03236	0,01411	-0,36657
-0,80462	-0,06865	-0,09433	-0,14651	-0,40629	-0,02710	0,05689	0,18955	-0,34849	-0,09704	-0,20896	0,02429	-0,20944	-0,13484	-0,04692	-0,24183	0,10816	-0,22504	0,10788	0,03561	0,07986	-0,35516
0,35144	0,05238	-0,60200	0,22199	-0,51052	0,00468	-0,36152	-0,73795	-0,14610	-0,04647	-0,07480	-0,30721	-0,69189	-0,37820	-0,22129	-0,24295	0,34442	-0,09299	-0,13031	0,02399	0,07328	0,29109
-2,95883	0,17674	0,25830	-0,54224	-0,14135	-0,00229	-0,21142	0,59021	-0,21412	-0,51581	0,26982	-0,11697	-0,91865	-0,61961	0,14857	-0,47998	-0,02669	-0,50273	-0,21910	0,54848	0,26082	-0,95870
1,49917	-0,01801	-0,16085	0,11977	0,21099	-0,38654	-0,43470	-0,00892	0,03392	-0,10434	-0,46111	-0,82295	0,05110	-0,48868	-0,08395	-0,38834	-0,14209	0,07480	-0,75900	0,05414	-0,26084	0,43345
-0,21876	0,00938	-0,07851	-0,23230	-0,29306	0,03154	-0,10741	-0,05337	-0,11642	0,02063	-0,04068	-0,07828	-0,16769	0,01437	-0,04325	-0,18384	0,04356	-0,16673	0,17426	-0,16753	-0,10745	-0,18756
-0,61984	-0,30356	-0,02858	-0,21614	-0,22605	0,16290	-0,06152	-0,02915	0,04619	0,24640	-0,06171	-0,19170	0,00211	-0,00352	0,11657	-0,21742	-0,08251	-0,35501	0,31806	-0,20133	0,01747	-0,27230
0,65400	-0,06599	-0,11707	-0,22699	-0,15176	0,02449	-0,04532	-0,00457	0,00399	0,02069	-0,04232	-0,19093	-0,13452	-0,13211	0,01743	-0,03042	0,06008	-0,29152	0,18575	-0,19048	-0,28866	-0,18488
0,76007	0,00350	-0,33955	0,07059	-0,16947	-0,05120	-0,23778	-0,22244	0,04625	0,04377	-0,07201	-0,37638	-0,20863	-0,19134	-0,01213	-0,19207	0,09631	0,06041	-0,04725	-0,05070	-0,16998	-0,05237
-1,42609	0,14004	0,17480	-0,93505	-0,20209	0,43275	-0,49672	-0,12953	0,94029	0,52941	-0,27519	-0,63596	-0,12599	-0,29313	0,22634	-0,03070	-0,63931	-0,17008	-0,01342	-0,23351	-0,03975	-0,62400
2,61568	-1,70034	-0,00497	-1,73873	-0,45814	0,67455	-1,04311	0,02784	-0,19900	-0,34597	-0,13668	0,10009	0,19149	-1,57084	0,11142	0,20755	-0,05224	-0,91557	-0,82417	-0,07605	0,19402	1,46359
2,44463	1,71637	1,10779	-1,27974	-0,32660	0,23919	-0,69271	-0,41975	-2,01938	-0,10659	0,40916	0,27928	-0,79399	-2,34329	-1,03905	-0,04309	-1,33230	-2,68956	0,71605	0,50728	-0,52662	0,85004
0,75624	-0,07592	-0,19448	-0,18658	-0,14391	-0,01784	-0,13013	-0,09651	-0,00201	-0,02088	-0,02932	-0,29433	-0,18904	-0,13896	0,02669	-0,11421	0,07203	-0,24811	0,15503	-0,19958	-0,28828	-0,19075
-0,68600	-0,02668	0,41917	0,95707	-0,14584	0,54247	-0,06071	0,02536	-0,32836	-0,24684	-0,13717	-0,34155	-0,29433	-0,16773	0,06213	-0,41059	0,10180	0,86066	0,30693	0,01750	0,21526	0,23547
-2,76111	-0,45734	0,34540	0,44986	-0,07227	0,07146	0,22979	-0,26774	-0,52764	-0,36049	0,31016	-0,16804	-0,41726	-0,28293	0,37972	-0,52107	-0,09678	-0,05122	-0,08460	-0,62565	0,59265	-0,57935
1,24974	0,06754	1,14196	0,92642	-0,94142	1,10159	-0,05250	0,60502	-0,39462	-0,23338	0,13289	-0,28303	-0,54353	-0,19997	-0,06114	-0,35805	0,47826	2,45481	-0,65969	0,22125	-1,15665	0,07579
1,12910	-0,16764	0,00534	-0,28894	-0,03634	0,12831	0,08485	0,10928	-0,06105	-0,15612	-0,10333	-0,20818	-0,13717	-0,29632	-0,08176	0,09466	0,00775	-0,63744	0,42650	-0,10002	-0,40378	-0,40871
-0,89185	-0,03116	-0,05479	-0,11658	-0,41767	0,03834	-0,11618	-0,02466	-0,16715	-0,03734	-0,15484	-0,10446	-0,28145	-0,03691	-0,02321	-0,29157	0,04720	-0,20965	0,14745	-0,13278	-0,05217	-0,23208
-0,47147	0,04298	-0,02906	-0,32849	-0,25784	0,13871	-0,22314	-0,08201	-0,06826	0,05497	0,07508	-0,05189	-0,11562	0,02890	-0,02372	-0,11679	-0,01257	-0,12101	0,11766	-0,03921	-0,04365	-0,33891
-0,01831	-0,11202	-0,32700	0,00436	-0,29845	0,10288	-0,21144	-0,32509	0,05277	0,17625	-0,19410	-0,28339	-0,25551	-0,11969	-0,04167	-0,19711	0,25422	-0,06924	0,05686	-0,10612	0,04279	-0,00408
-0,92343	1,07141	1,36861	0,35297	-0,02851	1,48669	1,10172	-0,91251	0,10570	-2,06745	1,49055	-0,11684	-0,44174	-1,00230	-0,88463	-0,08902	-0,03442	-1,48324	-0,22520	-0,80106	-0,28917	-0,23650
-5,16832	0,06575	-0,46932	0,42525	0,28283	0,27751	-0,42867	-0,03229	-0,56101	0,56789	0,43162	-0,02439	0,17058	-1,63991	1,82186	0,25851	0,15019	-1,06657	0,09477	-0,17137	-0,45875	-1,24006

0,56208

Tableau III.11 Poids et biais du meilleur modèle RQSP-RNA pour  $V_c$  (20-16-1).

Connexion des couches : (Caché-Sortie)		Connexion des couches : (Entrée-Caché)																				Connexion des couches : (Caché-Sortie)		
		Poids																						
		Biais																						
$W_{1j}^1$	$b_j^H$	$W_{j1}^1$	$W_{j2}^1$	$W_{j3}^1$	$W_{j4}^1$	$W_{j5}^1$	$W_{j6}^1$	$W_{j7}^1$	$W_{j8}^1$	$W_{j9}^1$	$W_{j10}^1$	$W_{j11}^1$	$W_{j12}^1$	$W_{j13}^1$	$W_{j14}^1$	$W_{j15}^1$	$W_{j16}^1$	$W_{j17}^1$	$W_{j18}^1$	$W_{j19}^1$	$W_{j20}^1$	$b_j^H$	$W_{1j}^H$	$b_1^0$
0,156686	0,301449	0,553803	0,060596	0,039908	-0,031204	-0,124726	-0,077049	-0,088576	0,079910	-0,359135	-0,087049	0,100849	-0,017134	0,393625	0,133771	-0,002731	-0,064640	0,065522	0,065522	-0,240921	-0,240921	0,301449	0,301449	-0,160898
0,079313	0,168577	0,233261	-0,216027	0,038295	-0,216243	0,186293	-0,293381	0,103107	0,045444	-0,145684	-0,033138	0,034247	-0,088028	0,184836	0,078129	0,149481	0,156986	0,044784	0,044784	-0,024564	-0,024564	0,168577	0,168577	0,160898
-0,025129	0,292571	0,417833	-0,121971	-0,066597	-0,047517	0,078736	0,168021	0,096991	0,066236	0,013594	-0,044686	-0,129432	0,070932	-0,076306	0,050990	0,132066	0,115906	0,294853	0,294853	-0,103337	-0,103337	0,292571	0,292571	0,160898
0,145216	-0,206126	-0,429581	-0,028094	-0,071706	0,013472	-0,036868	-0,121677	-0,007357	0,209985	0,183550	0,085633	-0,102130	-0,192406	0,122883	-0,027614	-0,094989	0,062159	0,094389	0,094389	0,059070	0,059070	-0,206126	-0,206126	0,160898
0,028528	0,168677	0,238056	-0,135042	-0,034000	-0,100451	-0,022829	-0,281481	-0,000910	-0,028613	-0,173592	-0,106148	-0,032375	-0,134552	0,023853	0,000661	0,142462	0,153980	0,181782	0,181782	0,040347	0,040347	0,168677	0,168677	0,160898
-0,034752	0,089091	0,123577	-0,133463	-0,102186	-0,068848	-0,047132	-0,148948	-0,020668	0,006916	-0,083769	-0,079175	-0,016676	-0,031467	-0,017166	-0,002952	0,103987	0,085931	0,057072	0,057072	0,041471	0,041471	0,089091	0,089091	0,160898
0,427083	-0,265804	-0,341288	-0,221592	-0,019577	-0,197377	-0,059569	0,006433	0,148824	-0,144791	-0,677946	-0,174584	0,079011	0,091027	0,053847	0,336097	-0,039303	0,113618	0,166636	0,166636	0,255156	0,255156	-0,265804	-0,265804	0,160898
0,042357	0,066568	0,168705	-0,055305	-0,075404	-0,016097	-0,019002	0,000475	0,007856	-0,000294	-0,048155	-0,047033	0,037664	0,131615	0,121226	0,075207	-0,066822	-0,084075	-0,086102	-0,086102	-0,135753	-0,135753	0,066568	0,066568	0,160898
-0,068354	-0,323638	-0,635929	0,015241	-0,126023	0,050790	-0,006388	-0,133758	0,013934	0,019670	-0,330509	0,001181	0,067014	0,106665	0,186108	-0,004388	0,187929	0,034166	0,190944	0,190944	0,063147	0,063147	-0,323638	-0,323638	0,160898
-0,055102	-0,102826	-0,122556	-0,131729	0,115945	0,038486	-0,081356	-0,012156	0,029089	-0,024200	0,019468	-0,041922	0,025049	0,035139	0,075863	0,055990	0,189756	0,135506	-0,000881	-0,000881	0,036442	0,036442	-0,102826	-0,102826	0,160898
0,001692	0,251279	0,292241	-0,166977	-0,145492	0,074949	-0,107026	-0,321127	-0,004388	0,060859	0,010212	-0,089453	-0,181086	-0,053376	-0,022455	-0,001282	-0,012842	0,258522	0,109871	0,109871	0,020418	0,020418	0,251279	0,251279	0,160898

*Chapitre III. Modélisation des propriétés critiques des hydrocarbures par les RDSP-RNA*

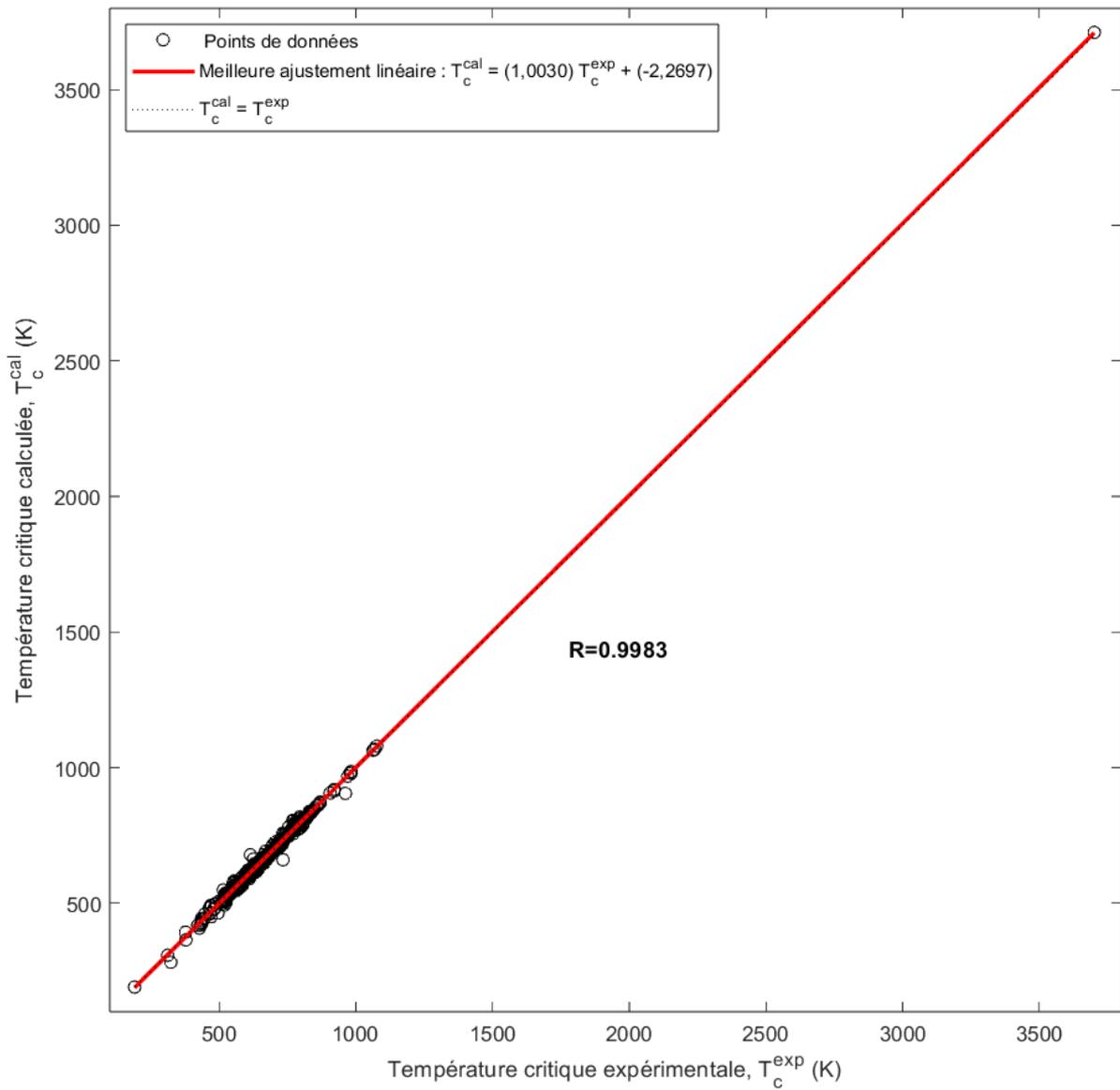
0,085646	0,522391	0,017071	-0,186688	0,026663	-0,085812	0,145196	-0,037125	-0,161692	-0,025079	0,184407	-0,202844	0,044120	0,112241	-0,008005	-0,059038	-0,022584	0,150305	0,115689	0,185190	-0,435155	0,192238
-0,450965	0,484211	-0,358641	-0,195224	-0,125849	0,037685	-0,113160	-0,266712	0,027552	0,038907	0,097137	-0,088791	-0,018712	0,152564	-0,334256	-0,055373	-0,244486	0,276919	0,381440	0,386487	0,050933	-0,165218
0,026721	0,100804	-0,005615	-0,024359	0,074510	-0,076114	0,102803	0,005506	0,142226	-0,080997	-0,005580	-0,067772	-0,063876	0,038235	0,193161	0,133690	0,089381	-0,156298	-0,141788	-0,134431	-0,120386	0,123190
0,043886	0,082721	-0,033898	-0,066842	-0,037033	-0,080998	-0,055427	0,013764	0,007631	-0,001487	0,029669	-0,083493	-0,036845	-0,013074	0,000711	-0,052502	-0,036881	0,061382	0,077747	0,046877	-0,010172	0,035728
-0,030238	-0,146211	0,076560	-0,066271	-0,014421	0,027269	-0,069908	-0,065700	0,100484	-0,046327	0,067514	0,004765	0,001126	0,009366	-0,123057	0,010775	-0,018096	0,020103	0,031037	0,085719	0,055987	-0,159706

### III.3.4. Estimation de la performance des résultats d'estimation issus des modèles RQSP-RNA développés

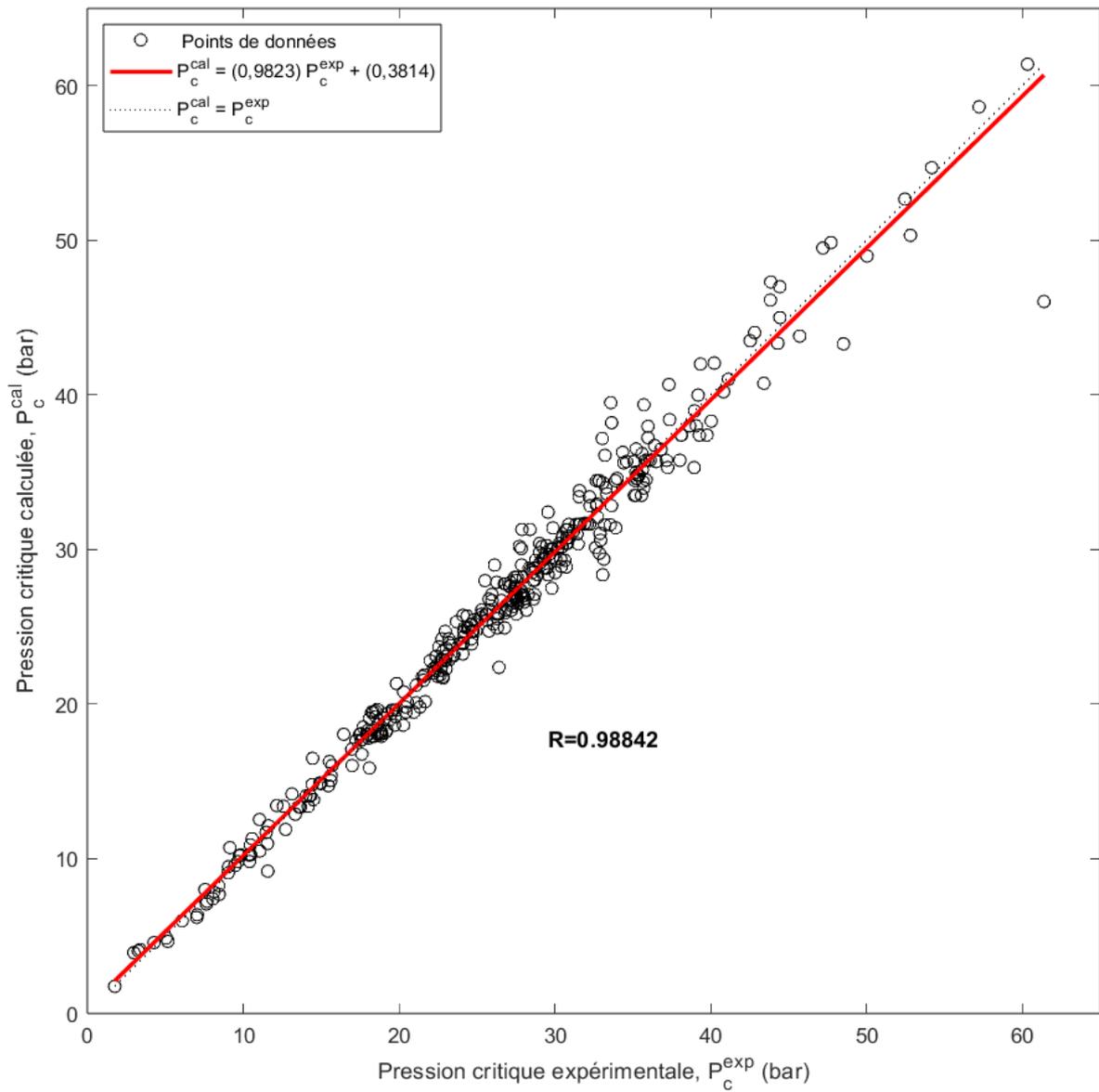
#### *III.3.4.1. Comparaison graphique entre les valeurs expérimentales et calculées à partir des modèles RQSP-RNA*

Pour prédire les trois propriétés critiques des hydrocarbures ( $T_c$ ,  $P_c$  et  $V_c$ ), nous employons des modèles qui intègrent des valeurs d'entrée ( $E_i$ ) représentant les descripteurs pertinents sélectionnés pour chaque paramètre. En outre, nous prenons en compte les poids ( $w_{ij}^I, w_{1j}^H$ ) et les biais ( $b_j^H, b_1^O$ ) associés à chaque connexion dans les modèles.

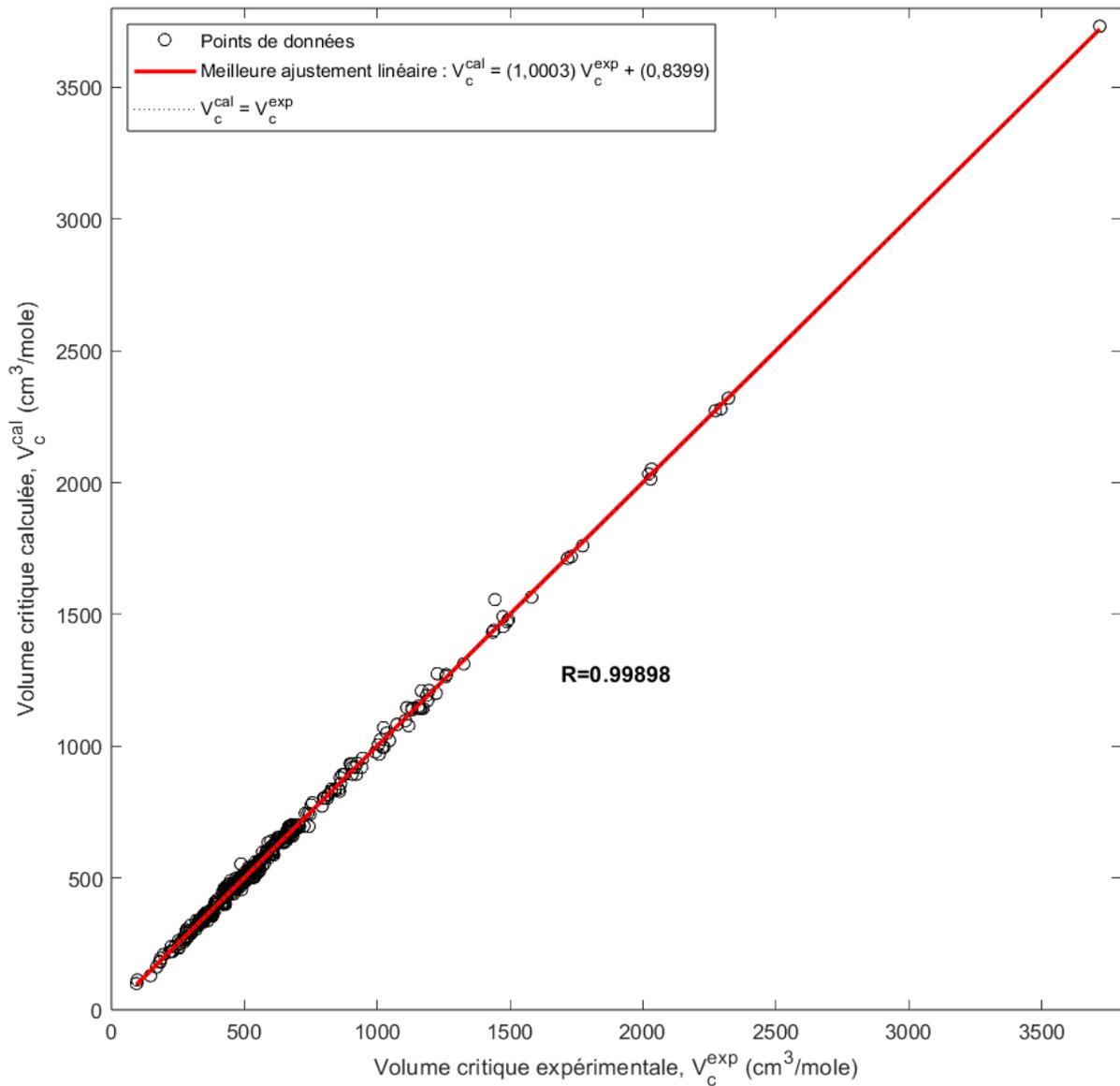
Les modèles de RQSP-RNA développés illustrent les comparaisons entre les valeurs expérimentales et calculées sur les figures. : **Figure III.7** pour le modèle de  $T_c$ , **figure III.8** pour le modèle de  $P_c$ , et sur la **figure III.9** pour le modèle de  $V_c$ . Pour l'ensemble de la base des données entier.



**Figure III.7.** Comparaison entre les valeurs expérimentales et calculées par le modèle RQSP-RNA de la température critique  $T_c$  pour la base des données entier.



**Figure III.8.** Comparaison entre les valeurs expérimentales et calculées par le modèle RQSP-RNA de la pression critique  $P_c$  pour la base des données entier.



**Figure III.9.** Comparaison entre les valeurs expérimentales et calculées par le modèle RQSP-RNA du volume critique  $V_c$  pour la base des données entier.

Une distribution étroite de données près de la ligne de pente unitaire suggère que les modèles peuvent prédire de manière précise les trois valeurs critiques. Les résultats obtenus à partir des corrélations vectorielles indiquent une adéquation idéale avec des valeurs proches de l'idéal (i.e.  $\alpha = 1$  (pente),  $\beta = 0$  (ordonnée à l'origine),  $R = 1$ ). Les équations de régression présentes dans les graphiques des valeurs critiques  $T_c$ ,  $P_c$  et  $V_c$  confirment cette précision pour l'ensemble des données. Les modèles démontrent ainsi de bons résultats avec des coefficients de détermination élevés, comme le montre également le **tableau III.12**.

**III.3.4.2 Calculs des paramètres de performance et analyse statistique des modèles de type RQSP-RNA obtenus**

Le **tableau III.12** présente un résumé des performances des meilleurs modèles. Il compare leur précision en utilisant la racine de l'erreur quadratique moyenne (REQM) et l'erreur standard de prédiction (ESP) pour chaque sous-ensemble de données : apprentissage, test, validation et la base de données globale. En plus de ces mesures, le tableau fournit également les valeurs de corrélation (R et R<sup>2</sup>), la pente ( $\alpha$ ), l'interception ( $\beta$ ) et le nombre de points de données (N) pour chaque phase.

**Tableau III.12** Paramètres statistiques des meilleurs modèles RQSP-RNA pour les trois propriétés.

Paramètres Statistiques	R (-)	R <sup>2</sup> (-)	REQM (K ; bar ; cm <sup>3</sup> /mol)	ESP (%)	ERM (%)	ERAM (%)	$\alpha$ (-)	$\beta$ (-)	N (-)
Phase d'apprentissage									
<i>T<sub>c</sub><sup>App</sup></i>	0,9976	0,9953	8,5655	1,3072	1,08116	1,0811	0,9995	0,3061	320
<i>P<sub>c</sub><sup>App</sup></i>	0,9895	0,9791	1,4128	5,4049	3,4921	3,4704	0,9851	0,3523	360
<i>V<sub>c</sub><sup>App</sup></i>	0,9991	0,9981	15,8090	2,5685	2,2004	2,1973	0,9999	0,0658	320
Phase de test									
<i>T<sub>c</sub><sup>Test</sup></i>	0,9986	0,9973	18,8694	2,7418	1,9917	2,0188	1,0050	-4,7526	80
<i>P<sub>c</sub><sup>Test</sup></i>	0,9767	0,9539	1,8878	7,0387	5,5718	5,7616	0,9554	0,7238	40
<i>V<sub>c</sub><sup>Test</sup></i>	0,9988	0,9976	20,4706	3,4758	3,1467	3,0908	1,0022	3,6688	80
Phase entière (apprentissage + test)									
<i>T<sub>c</sub><sup>Ent</sup></i>	0,9983	0,9966	11,3976	1,7221	1,2636	1,2686	1,0030	-2,2697	400
<i>P<sub>c</sub><sup>Ent</sup></i>	0,9884	0,9770	1,4673	5,5986	3,7001	3,6995	0,9823	0,3814	400
<i>V<sub>c</sub><sup>Ent</sup></i>	0,9990	0,9980	16,8448	2,7606	2,3897	2,3760	1,0003	0,8399	400

Le **tableau III.12** présente différents coefficients statistiques et types d'erreur qui illustrent la performance des modèles développés. Les résultats s'avèrent très prometteurs : en effet, la valeur du coefficient de détermination ( $R^2$ ) dépasse 0,9800 pour chacun des trois modèles, et ce, sur l'ensemble des sous-bases de données. Ces résultats témoignent de la robustesse et de la précision des modèles développés.

Les calculs de valeurs **REQM** confirment la haute précision de nos modèles. En effet, les résultats obtenus, pour les quatre phases des sous-ensembles possibles, se situent dans des plages très cohérentes : entre 8,5655 et 18,8694 **K** pour le paramètre  $T_c$ , entre 1,4128 et 1,8878 **bar** pour le paramètre  $P_c$ , et entre 15,8090 et 20,4706 **cm<sup>3</sup>/mol** pour le paramètre  $V_c$ .

L'erreur standard de prédiction (**ESP**) et la **REQM** sont deux mesures similaires, mais l'**ESP**, exprimée en pourcentage, offre une meilleure lisibilité pour comparer les modèles. Dans le cas présent, les modèles TC, PC et VC affichent tous des ESP inférieurs à 6% pour les quatre sous-ensembles étudiés. Plus précisément, l'**ESP** varie de 1,3072% à 2,7418% pour  $T_c$ , de 5,4049% à 7,0387% pour  $P_c$  et de 2,5685% à 3,4758% pour  $V_c$ , ce qui indique une bonne performance de ces modèles.

C'est vrai, la similarité frappante des valeurs d'erreur entre les différents sous-ensembles (apprentissage, test, validation et ensemble entier) pour chaque paramètre est un indicateur fort de l'équilibre atteint par nos modèles RQSP-RNA. Cet équilibre délicat entre sous-apprentissage et sur-apprentissage témoigne de la stabilité et de la robustesse de nos modèles, confirmant ainsi leur fiabilité **Bishop, C. M. (1995) ; Pigram et all (2001) ; Hawkins, D. M. (2004)**.

#### *III.3.4.3 Analyse de sensibilité des descripteurs dans chaque modèle RQSP-RNA par la méthode de poids*

Pour évaluer l'importance de chaque descripteur d'entrée sur nos modèles, et comme évoqué dans le paragraphe III.2.2.6, nous allons utiliser la méthode d'analyse de sensibilité au poids. Cette méthode nous permettra de déterminer l'influence de chaque descripteur d'entrée sur les résultats de chaque modèle développé.

Les contributions des descripteurs choisis sur les modèles de RQSP-RNA ont été établies et illustrées sur les **figure III.10**, **figure III.11** et la **figure III.12**, respectivement pour les modèles de  $T_c$ ,  $P_c$  et  $V_c$ .

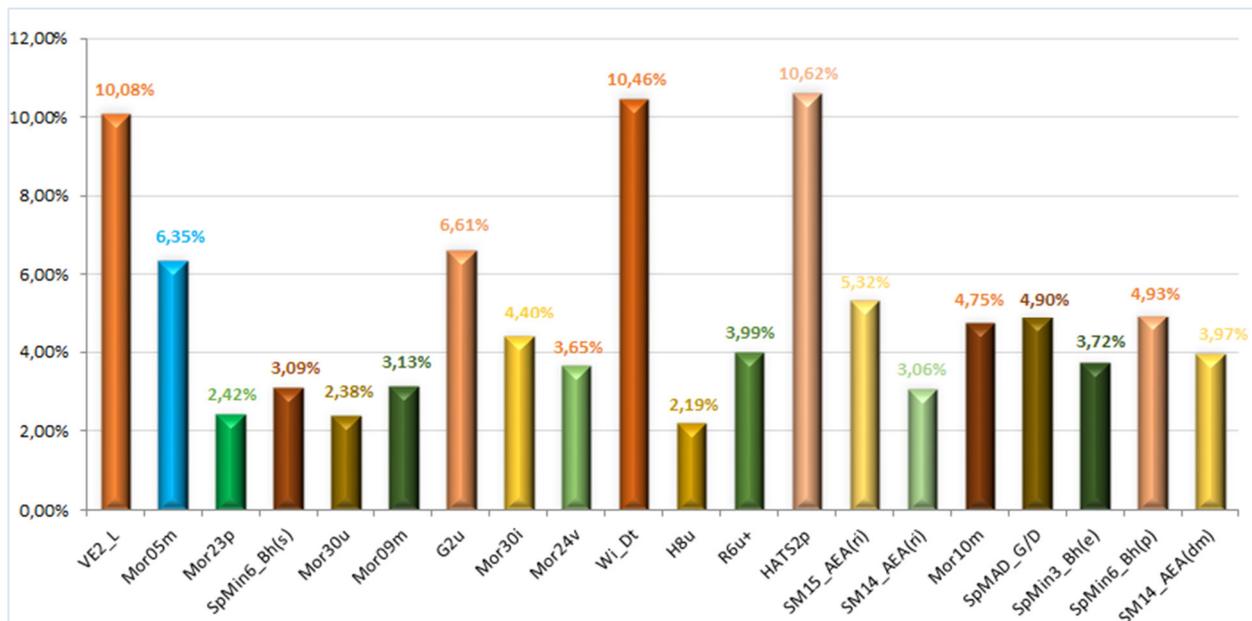


Figure III.10. Contributions relatives des descripteurs pertinents d'entrée pour le modèle de RQSP-RNA de la température critique Tc.

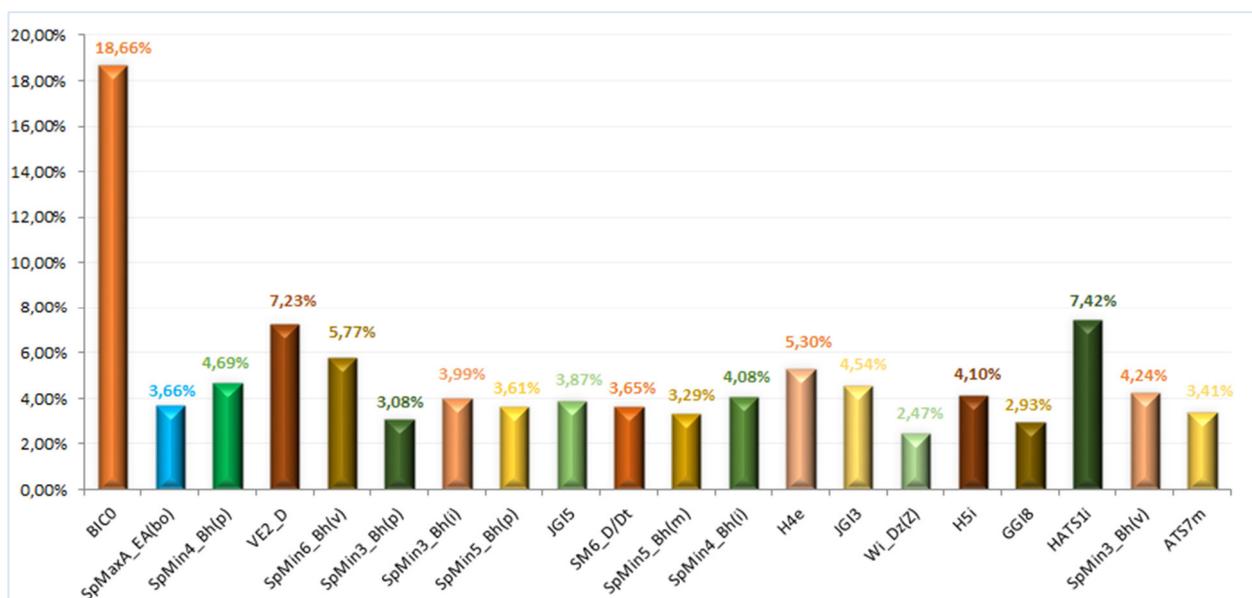


Figure III.11. Contributions relatives des descripteurs pertinents d'entrée pour le modèle de RQSP-RNA de la pression critique Pc.

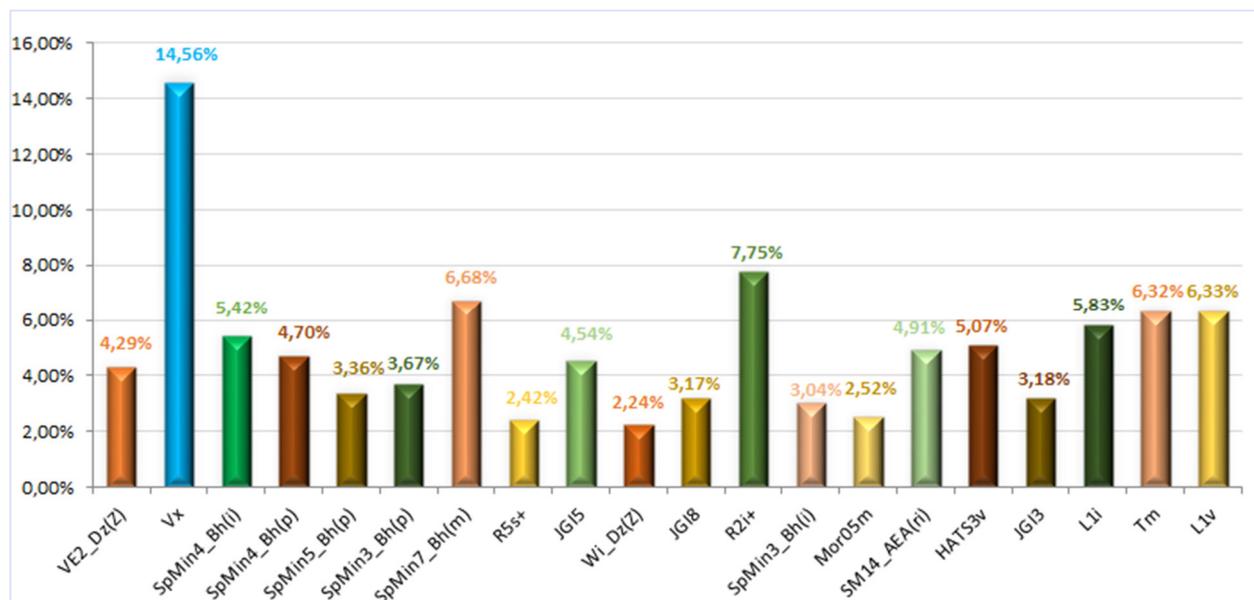


Figure III.12. Contributions relatives des descripteurs pertinents d'entrée pour le modèle de RQSP-RNA du volume critique  $V_c$ .

L'analyse de sensibilité, via l'importance relative ( $IR$ ) et les valeurs de contributions moyennes ( $VCM$ ), révèle que la plupart des descripteurs ont un impact significatif dans les trois modèles (TC, PC et VC). En effet, malgré quelques descripteurs à faible contribution, la majorité se trouvent autour de la valeur de ( $VCM$ ) qui égale à 5,00% pour les trois propriétés critiques  $T_c$ ,  $P_c$  et  $V_c$ .

Il est intéressant de constater que la différence entre les descripteurs, notamment ceux ayant une forte ou moyenne contribution aux ( $VCM$ ) de chaque modèle, n'est pas significative. Cela suggère que tous les descripteurs sélectionnés ont joué un rôle, à des degrés divers, dans le développement des modèles de RQSP-ARN.

### III.4. Conclusion

Ce chapitre a présenté l'ensemble des travaux de modélisation des trois paramètres critiques des hydrocarbures purs : la température critique ( $T_c$ ), la pression critique ( $P_c$ ) et le volume critique ( $V_c$ ). Pour plus de clarté, le chapitre est organisé en deux parties principales.

La première partie de notre étude a porté sur la présentation des outils et méthodes mis en œuvre pour développer des modèles de type RQSP-RNA. Nous avons détaillé chaque étape du processus de développement, en précisant les logiciels et outils informatiques utilisés. Un

organigramme a permis de résumer ces étapes. Ensuite, nous avons exposé en détail chaque méthode de développement des modèles, en commençant par les étapes de prétraitement. Cela incluait la collecte et le traitement de la base de données, l'optimisation des structures chimiques et le codage SMILES. Enfin, nous avons abordé le choix des paramètres des modèles avant leur exécution.

La deuxième partie de ce chapitre a présenté les résultats des modèles RQSP-RNA développés pour prédire trois propriétés critiques des hydrocarbures purs : Tc, Pc et Vc. L'analyse a porté sur plusieurs aspects, notamment : la sélection des descripteurs pertinents lors du prétraitement, l'architecture et les performances des meilleurs modèles, l'estimation des erreurs et le calcul des coefficients statistiques (REQM, ESP), l'analyse de sensibilité des descripteurs pour chaque paramètre et la validation finale des modèles RQSP-RNA.

L'analyse de la performance des modèles a révélé que les meilleurs d'entre eux estiment les paramètres critiques des hydrocarbures purs avec une grande précision. Cette conclusion est supportée par les valeurs élevées des coefficients de détermination ( $R^2$ ) et les faibles valeurs des erreurs REQM et ESP.

---

***Conclusion générale et  
perspectives***

---

Les hydrocarbures, constituant principal du pétrole brut, incluent des composants saturés et insaturés essentiels pour divers mélanges pétroliers. Leurs propriétés physicochimiques et thermodynamiques sont cruciales pour des industries comme la pétrochimie, car elles aident à comprendre les phénomènes et opérations unitaires liées au traitement, à la production, au transport et au stockage. Cependant, la détermination des propriétés des mélanges pétroliers repose sur celles de leurs composants purs, dont certaines, comme les propriétés critiques, sont difficiles à mesurer en raison des tests de laboratoire coûteux et lents.

Dans ce projet de fin d'études, nous avons abordé une question cruciale dans le domaine de la recherche scientifique liée à l'écologie industrielle. Notre travail s'est concentré sur la modélisation moléculaire, en développant des modèles RQSP basés sur des réseaux de neurones artificiels multicouches (ARN-PMC) pour prédire trois propriétés critiques des hydrocarbures purs : la température critique ( $T_c$ ), la pression critique ( $P_c$ ) et le volume critique ( $V_c$ ). Ces prédictions sont réalisées uniquement à partir des informations structurelles des molécules d'hydrocarbures étudiées, en utilisant des techniques avancées d'intelligence artificielle.

La méthode de développement utilisée dans ce projet comprend plusieurs étapes basées sur des études antérieures. Les données expérimentales, provenant de diverses sources, incluent environ 400 molécules d'hydrocarbures purs de différentes catégories industrielles, avec plusieurs isomères de position. Ces données ont été divisées en deux ensembles : un pour l'apprentissage et l'autre pour les tests. Un grand nombre de descripteurs moléculaires ont été calculés en ligne avec le logiciel Dragon 7. Plusieurs méthodes statistiques, y compris la régression linéaire multiple (RLM), ont été utilisées pour réduire ce nombre et ne conserver que les descripteurs pertinents. Les meilleurs modèles RQSP-RNA ont été obtenus par essais et erreurs en utilisant l'algorithme d'apprentissage "BFGS". Les architectures optimales étaient "20-14-1" pour  $T_c$ , "20-28-1" pour  $P_c$  et "20-16-1" pour  $V_c$ . Chaque modèle a ensuite été analysé et testé à l'aide de paramètres statistiques tels que le coefficient de détermination ( $R^2$ ), la Racine Carrée de l'Erreur Quadratique Moyenne (REQM) et l'Erreur Relative Absolue Moyenne (ERAM).

Les modèles RQSP-RNA proposés dans ce projet ont démontré une excellente précision, avec des valeurs de  $R^2$  allant de 0,9981 pour  $V_c^{\text{app}}$  à 0,9539 pour  $P_c^{\text{Test}}$ . La robustesse et la capacité prédictive de ces modèles ont également été confirmées par diverses erreurs (REQM, ESP, ERM et ERAM), avec des résultats très prometteurs : 11,3976 K pour  $T_c$ , 1,4673 bar pour

Pc et 16,8448 cm<sup>3</sup>/mol pour Vc. Pour les erreurs relatives restantes (ESP, ERM et ERAM), les résultats sont également satisfaisants, n'excédant pas 7,1 % pour les trois paramètres dans les deux ensembles de données. Plus précisément, les erreurs varient de 1,0811 % pour l'ERAM de Tc<sup>app</sup> à 7,0387 % pour l'ESP de Pc<sup>Test</sup>. Comparés à d'autres modèles similaires dans la littérature, les modèles RQSP-RNA développés ici ont montré des performances supérieures pour la prédiction des paramètres Tc, Pc et Vc, fournissant ainsi de meilleurs résultats.

Les modèles RQSP-RNA développés dans ce projet ne se limitent pas à des familles spécifiques telles que les alcanes ou les alcènes, mais englobent toutes les familles d'hydrocarbures dans un même modèle. Ces modèles peuvent être utilisés dans l'industrie et l'ingénierie pétrolière, ainsi que dans d'autres domaines liés aux hydrocarbures purs, pour estimer leurs Tc, Pc et Vc à partir de leur structure moléculaire, en utilisant des descripteurs moléculaires simples et faciles à calculer.

L'ensemble des travaux présentés dans ce projet a largement permis d'atteindre l'objectif principal : développer des modèles utilisant des méthodes computationnelles d'intelligence artificielle et des informations structurelles des molécules d'hydrocarbures pour prédire certaines de leurs propriétés critiques.

En vertu de ce travail, on peut tirer quelques perspectives qui peut être des informations utiles pour les prochaines recherches dans ce domaine :

- L'application de la même méthodologie aux mélanges pétroliers représenterait une excellente perspective, à condition que les données nécessaires, les ressources logistiques et la collaboration avec les services concernés soient disponibles.
- Nous pouvons également étendre l'idée de ce travail à d'autres techniques de modélisation récentes, comme les Machines à Vecteurs de Support (SVM).
- Pour faciliter le développement de modèles de type RQSP-RNA, il est recommandé d'utiliser des logiciels et outils informatiques spécifiques. Par exemple, le logiciel commercial Dragon 7 peut calculer et traiter environ 5270 descripteurs moléculaires distincts. De même, Matlab est un outil puissant utilisé par de nombreux ingénieurs pour développer des modèles neuronaux et d'intelligence artificielle. Il offre des interfaces intégrées standard et la possibilité de programmer des scripts personnalisés pour plus de liberté, de choix et de créativité dans la modélisation. Ces logiciels permettent de produire des programmes plus stables et des résultats plus précis et reproductibles

---

# *Références bibliographiques*

---

**Abdel-Aal, H. K., & Alsahlawi, M. A. (Eds.). (2013).** Petroleum economics and engineering. CRC Press.

**Abdul Jameel, A. G., Van Oudenhoven, V., Emwas, A. H., & Sarathy, S. M. (2018).** Predicting octane number using nuclear magnetic resonance spectroscopy and artificial neural networks. *Energy & fuels*, 32(5), 6309-6329

**Albahri, T. A. (2003).** Structural group contribution method for predicting the octane number of pure hydrocarbon liquids. *Industrial & engineering chemistry research*, 42(3), 657-662

**Albahri, T. A., Riazi, M. R., & Alqattan, A. A. (2002).** Octane number and aniline point of petroleum fuels. *Fuel Chemistry Division Preprints*, 47(2), 710.

**Alqaheem, S. S., & Riazi, M. R. (2017).** Flash points of hydrocarbons and petroleum products: Prediction and evaluation of methods. *Energy & Fuels*, 31(4), 3578-3584.

**Ammar, M. Y. (2007).** "Mise en œuvre de réseaux de neurones pour la modélisation de cinétiques réactionnelles en vue de la transposition batch/continu". Thèse de doctorat en Génie des Procédés et de l'Environnement, Ecole doctorale : Transfert, Dynamique des Fluides, Energétique & Procédés, Institut national polytechnique de Toulouse, France. Disponible en ligne : <http://ethesis.inp-toulouse.fr/archive/00000607/>.

**Ammi, Y. (2016).** "Modélisation des phénomènes de transport lors d'osmose inverse". Thèse de Doctorat en génie des procédés, Option : Modélisation des phénomènes de transport en sciences de l'eau et d'environnement, Laboratoire de Biomatériaux et Phénomènes de Transport (LBMPT), Département de génie des procédés et environnement, Université Yahia Fares de Médéa. Médéa, Algérie.

**Amri, F. (2017).** The relationship amongst energy consumption (renewable and non-renewable), and GDP in Algeria. *Renewable and Sustainable Energy Reviews*, 76, 62-71.

**Andrade, C. H., Pasqualoto, K. F., Ferreira, E. I., & Hopfinger, A. J. (2010).** 4D-QSAR: perspectives in drug design. *Molecules*, 15(5), 3281-3294.

**Apergis, N., & Payne, J. E. (2012).** Renewable and non-renewable energy consumption-growth nexus: Evidence from a panel error correction model. *Energy economics*, 34(3), 733-738.

**Arab, M., Bracene, R., Roure, F., Zazoun, R. S., Mahdjoub, Y., & Badji, R. (2015).** Source rocks and related petroleum systems of the Chelif Basin,(western Tellian domain, north Algeria). *Marine and Petroleum Geology*, 64, 363-385.

**Atlas, R. M., & Bartha, R. (1992).** Hydrocarbon biodegradation and oil spill bioremediation. In *Advances in microbial ecology*(pp. 287-338). Springer, Boston, MA.

**Bagheri, M., Borhani, T. N. G., & Zahedi, G. (2012).** Estimation of flash point and autoignition temperature of organic sulfur chemicals. *Energy Conversion and Management*, 58, 185-196

**Balaban, A. T. (1982).** Highly discriminating distance-based topological index. *Chemical Physics Letters*, 89(5), 399-404.

**Balaban, A. T. (1985).** Applications of graph theory in chemistry. *Journal of chemical information and computer sciences*, 25(3), 334-343.

- Basheer, I. A., & Hajmeer, M. (2000).** Artificial neural networks: fundamentals, computing, design, and application. *Journal of microbiological methods*, 43(1), 3-31
- Baughman, D. R., & Liu, Y. A. (1995).** *Neural networks in bioprocessing and chemical engineering.* Academic press.
- Beale, M. H., Hagan, M. T., & Demuth, H. B. (2012).** *Neural network toolbox™ user's guide.* In R2012a, The MathWorks, Inc., 3 Apple Hill Drive Natick, MA 01760-2098, [www.mathworks.com](http://www.mathworks.com).
- Belaid, F., & Youssef, M. (2017).** Environmental degradation, renewable and non-renewable electricity consumption, and economic growth: Assessing the evidence from Algeria. *Energy Policy*, 102, 277-287.
- Belghit, C. (2019).** "Nouvelles corrélations pour la détermination des propriétés physiques et thermodynamiques des hydrocarbures et des fractions pétrolières", Thèse de doctorat en Génie des Procédés, Option : Chimie Industrielle, Université 08 Mai 1945 Guelma, Département de Génie des Procédés, Laboratoire des Silicates, Polymère et des Nanocomposites (LSPN), Guelma.
- Berkane, M. (2010).** "Estimation et analyse de mouvement par approches neuronales". Thèse de doctorat en Images et Systèmes, Ecole doctorale : Electronique, Electrotechnique, Automatique Transfert. Laboratoire CREATIS-LRMN, CNRS UMR 5220, Inserm U6. Institut National des Sciences Appliquées de Lyon, France. Disponible en ligne : <http://theses.insalyon.fr/publication/2010ISAL0069/these.pdf>.
- Bishop, C. M. (1995).** *Neural networks for pattern recognition.* Oxford university press, New York, USA
- Bitam, S. (2018).** "Modélisation de l'activité thérapeutique à partir de la structure moléculaire". Thèse de Doctorat en génie des procédés, Laboratoire de Biomatériaux et Phénomènes de Transport (LBMPT), Département de génie des procédés et environnement, Université Yahia Fares de Médéa. Médéa, Algérie.
- Bonachera, F. (2011).** "Les triplets pharmacophoriques ou Développement et applications". Thèse de doctorat en Chimie théorique, physique et analytique. Ecole doctorale biologie santé, Université des sciences et technologies de Lille 1, France. Disponible en ligne : <https://pdfs.semanticscholar.org/1da9/5b49c738012b8b229ec0765cdc84410481d5.pdf>.
- Bonchev, D. (1991).** *Chemical graph theory: introduction and fundamentals (Vol. 1).* CRC Press.
- Borhani, T. N. G., Afzali, A., & Bagheri, M. (2016).** QSPR estimation of the auto-ignition temperature for pure hydrocarbons. *Process Safety and Environmental Protection*, 103, 115-125.
- Bosque, R., Sales, J., Bosch, E., Roses, M., García-Alvarez-Coque, M. C., & TorresLapasió, J. R. (2003).** A QSPR study of the p solute polarity parameter to estimate retention in HPLC. *Journal of chemical information and computer sciences*, 43(4), 1240-1247.
- Charton, M. (2002).** *Advances in Quantative Structure-Property Relationships.* Elsevier.

**ChemOffice (2012).** PerkinElmer Informatics. CambridgeSoft. 100 CambridgePark Drive, Cambridge, MA 02140. [www.cambridgesoft.com](http://www.cambridgesoft.com).

**Cherkasov, A., Muratov, E. N., Fourches, D., Varnek, A., Baskin, I. I., Cronin, M., ... & Consonni, V. (2014).** QSAR modeling: where have you been? Where are you going to?. *Journal of medicinal chemistry*, 57(12), 4977-5010.

**Chitour, C. E. (1999).** Raffinage: les propriétés physiques des hydrocarbures et des fractions pétrolières. Office des publications universitaires.

**Chitita, S. (2017).** Modélisation de molécules organiques hétérocycliques biologiquement actives par des méthodes QSAR/QSPR. Recherche de nouveaux médicaments. Thèse de doctorat en Chimie. Faculté des sciences, Université Moulay Ismaïl. Meknès, Maroc.

**Constantinou, L., & Gani, R. (1994).** New group contribution method for estimating properties of pure compounds. *AIChE Journal*, 40(10), 1697-1710.

**Cramer, R. D., Patterson, D. E., & Bunce, J. D. (1988).** Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *Journal of the American Chemical Society*, 110(18), 5959-5967.

**Cybenko, G. (1989).** Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4), 303-314.

**Daubert, T. E., & Danner, R. P. (1997).** API Technical Data Book Petroleum Refining: American Petroleum Institute (API): Washington, DC.

**Dearden, J. C., Cronin, M. T., & Kaiser, K. L. (2009).** How not to develop a quantitative structure–activity or structure–property relationship (QSAR/QSPR). *SAR and QSAR in Environmental Research*, 20(3-4), 241-266.

**Dong, J., Cao, D. S., Miao, H. Y., Liu, S., Deng, B. C., Yun, Y. H., ... & Chen, A. F. (2015).** ChemDes: an integrated web-based platform for molecular descriptor and fingerprint computation. *Journal of cheminformatics*, 7(1), 60.

**Dreyfus, G. (2005).** Neural networks: methodology and applications. Springer Science & Business Media.

**Dreyfus, G., Martinez, J. M., Samuelides, M., Gordon, M. B., Badran, F., Thiria, S., & Hérault, L. (2002).** Réseaux de neurones-Méthodologie et applications. (No. BOOK). Eyrolles.

**Dreyfus, G., Martinez, J. M., Samuelides, M., Gordon, M. B., Badran, F., & Thiria, S. (2008).** Apprentissage statistique : Réseaux de neurones : Cartes topologiques Machines à vecteurs supports, eyrolles.

**Duch, W., Swaminathan, K., & Meller, J. (2007).** Artificial intelligence approaches for rational drug design and discovery. *Current pharmaceutical design*, 13(14), 1497-1508.

**Dudek, A. Z., Arodz, T., & Gálvez, J. (2006).** Computational methods in developing quantitative structure-activity relationships (QSAR): a review. *Combinatorial chemistry & high throughput screening*, 9(3), 213-228.

- Dudley, B. (2018).** BP statistical review of world energy. BP Statistical Review, London, UK, accessed Aug, 6, 2018.
- Dupuy, A. (2015).** "Formation professionnalisante PRO/RCP1 Groupe 01, Ingénieurs procédés en raffinage et pétrochimie hydrocarbures et produits pétrolières", Centre IAP, Hassi Messaoud.
- Elsharkawy, A. M., & Elkamel, A. (2000).** Compressibility factor for sour gas reservoirs. In SPE Asia Pacific Oil and Gas Conference and Exhibition. Society of Petroleum Engineers.
- Elsharkawy, A. M., Hashem, Y. S. K. S., & Alikhan, A. A. (2001).** Compressibility factor for gas condensates. *Energy & fuels*, 15(4), 807-816.
- Engel, T., & Gasteiger, J. (Eds.). (2018).** Applied chemoinformatics: achievements and future opportunities. John Wiley & Sons.
- Fahim, M. A., Al-Sahhaf, T. A., & Elkilani, A. (2009).** Fundamentals of petroleum refining. Elsevier.
- Fukui, K. (1970).** Theory of orientation and stereoselection. In *Orientation and Stereoselection* (pp. 1-85). Springer, Berlin, Heidelberg.
- Ghahramani, Z. (2015).** Probabilistic machine learning and artificial intelligence. *Nature*, 521(7553), 452-459.
- Goh Gevrey, M., Dimopoulos, I., & Lek, S. (2003).** Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological modelling*, 160(3), 249- 264.
- Goh, A. T. (1995).** Back-propagation neural networks for modeling complex systems. *Artificial Intelligence in Engineering*, 9(3), 143-151.
- Goulon-Sigwalt-Abram, A. (2008).** "Une nouvelle méthode d'apprentissage de données structurées: applications à l'aide à la découverte de médicaments" Thèse de doctorat en Electronique, Université Pierre et Marie Curie-Paris VI, France. Disponible en ligne : <https://pastel.archives-ouvertes.fr/pastel-00004103/>.
- Gurney, K. (1997).** *An Introduction to Neural Networks*, Routledge, ISBN 1-85728-673-1 London.
- Guthrie, V. B., & Guthrie, V. B. (1960).** Petroleum products handbook (Vol. 33). New York/Toronto/London: McGraw-Hill.
- Hagan, M. T., Demuth, H. B., & Beale, M. H. (1996).** Neural network design, PWS Pub. Co., Boston, 3632
- Hamadache, M. (2014).** "Modélisation en vue de la prédiction ou de la corrélation de l'activité toxicologique à partir de la structure moléculaire". Thèse de Doctorat en génie des procédés, Laboratoire de Biomatériaux et Phénomènes de Transport (LBMPT), Département de génie des procédés et environnement, Université Yahia Fares de Médéa. Médéa, Algérie.
- Hamadache, M., Amrane, A., Hanini, S., & Benkortbi, O. (2018).** Multilayer perceptron model for predicting acute toxicity of fungicides on rats. *International Journal of Quantitative Structure-Property Relationships (IJQSPR)*, 3(1), 100-118.

- Hansen, P. J., & Jurs, P. C. (1988).** Chemical applications of graph theory. Part I. Fundamentals and topological indices. *Journal of Chemical Education*, 65(7), 574.
- Hassoun, M. H. (1995).** Fundamentals of artificial neural networks. MIT press
- Hawkins, D. M. (2004).** The problem of overfitting. *Journal of chemical information and computer sciences*, 44(1), 1-12.
- Haykin, S. (1998).** *Neural Networks: A Comprehensive Foundation*, Prentice Hall PTR. Upper Saddle River, NJ, USA.
- Haykin, S. (2010).** *Neural Networks and Learning Machines*, 3/E. Pearson Education India.
- Higo, J. I., & Gō, N. (1989).** Algorithm for rapid calculation of excluded volume of large molecules. *Journal of computational chemistry*, 10(3), 376-379.
- Jain, A. K., Mao, J., & Mohiuddin, K. M. (1996).** Artificial neural networks: A tutorial. *Computer*, 29(3), 31-44.
- Jensen, F. (2007).** *Electron correlation methods*. Jensen F. Introduction to computational chemistry. 2nd ed. New Jersey: John Wiley & Sons, 133-144.
- Karelson, M. (2000).** *Molecular descriptors in QSAR/QSPR (Vol. 230)*. New York: WileyInterscience.
- Khaouane, L. (2013).** "Etude et Modélisation de la Biosynthèse des Antibiotiques à partir de Différentes Souches Productrices – Cas de Pleuromutiline". Thèse de Doctorat en génie des procédés, Laboratoire de Biomatériaux et Phénomènes de Transport (LBMPT), Département de génie des procédés et environnement, Université Yahia Fares de Médéa. Médéa, Algérie.
- Khouas, R., Ziane, F., & Khaouane, L. (2020).** Développement des modèles de type RQSPRNA pour la prédiction des propriétés critiques des hydrocarbures.
- Kikuchi, O. (1987).** Systematic QSAR procedures with quantum chemical descriptors. *Quantitative Structure-Activity Relationships*, 6(4), 179-184.
- Kraim, K. (2009).** "Etudes de QSAR sur des activités biologiques utilisant des produits d'origines naturels". Thèse de doctorat en génie des procédés, Option : Chimie industrielle, Département de génie des procédés, Université 08 mai 1945 Guelma. Guelma, Algérie.
- Krose, B., & Smagt, P. V. (1996).** *An Introduction to Neural Network*. The University of Amsterdam. Amsterdam, Netherland.
- Labute, P. (2000).** A widely applicable set of descriptors. *Journal of Molecular Graphics and Modelling*, 18(4-5), 464-477.
- Lahiouel, Y. (1995a).** "Estimation de la température d'ébullition et de facteur acentrique des composés organiques par les méthodes de contribution de groupes et les indices topologiques". Mémoire de Magister en Chimie Industrielle, Option : Conception Assistée par Ordinateur, Ancien Institut de Chimie Industrielle, Ancien Centre Universitaire de Guelma. Guelma, Algérie.
- Laidi, M. (2009).** "Approche Neuronale Pour l'Estimation du coefficient de Diffusion Moléculaire Apparent dans un Milieu Poreux Anisotrope (Bois)". Mémoire de Magister en Génie mécanique, Option : Energétique, Ancien Faculté des Sciences et de la Technologie, Université Yahia Fares de Médéa. Médéa, Algérie.

- Laidi, M. (2013).** "Etude et Modélisation d'un Réfrigérateur Solaire". Thèse de Doctorat en génie des procédés, Laboratoire de Biomatériaux et Phénomènes de Transport (LBMPT), Département de génie des procédés et environnement, Université Yahia Fares de Médéa. Médéa, Algérie.
- Lewis, D. F. V., Ioannides, C., & Parke, D. V. (1994).** Interaction of a series of nitriles with the alcohol-inducible isoform of P450: computer analysis of structure—activity relationships. *Xenobiotica*, 24(5), 401-408.
- Lucea, M. (2006).** "Modélisation dynamique par réseaux de neurones et machines à vecteurs supports : contribution à la maîtrise des émissions polluantes de véhicules automobiles". **Thèse de doctorat en Electronique, Université Pierre et Marie Curie-Paris VI, France. Disponible en ligne : <https://pastel.archives-ouvertes.fr/pastel-00001943>.**
- Martins, J. P. A., Barbosa, E. G., Pasqualoto, K. F., & Ferreira, M. M. (2009).** LQTAQSAR: a new 4D-QSAR methodology. *Journal of chemical information and modeling*, 49(6), 1428-1436.
- Masseron, J. (1990).** Petroleum economics. Editions Technip.
- Mauri, A., Consonni, V., & Todeschini, R. (2017).** Molecular descriptors. Handbook of Computational Chemistry, 2065-2093.
- Mauri, A., Consonni, V., Pavan, M., & Todeschini, R. (2006).** Dragon software: An easy approach to molecular descriptor calculations. *Match*, 56(2), 237-248.
- Mihalić, Z., & Trinajstić, N. (1992).** A graph-theoretical approach to structure-property relationships. *Journal of chemical education*, 69(9), 701-712.
- Minsky, M. (1961).** Steps toward artificial intelligence. *Proceedings of the IRE*, 49(1), 8-30.
- Navajas, C., Poso, A., Tuppurainen, K., & Gynther, J. (1996).** Comparative Molecular Field Analysis (CoMFA) of MX Compounds using different Semi-empirical Methods: LUMO Field and its Correlation with Mutagenic Activity. *Quantitative Structure-Activity Relationships*, 15(3), 189-193.
- Nieto-Draghi, C., Fayet, G., Creton, B., Rozanska, X., Rotureau, P., de Hemptinne, J. C., ... & Adamo, C. (2015).** A general guidebook for the theoretical prediction of physicochemical properties of chemicals for regulatory purposes. *Chemical reviews*, 115(24), 13093-13164.
- Oussar, Y. (1998).** "Réseaux d'ondelettes et réseaux de neurones pour la modélisation statique et dynamique de processus". Thèse de doctorat en Robotique, Université Pierre et Marie Curie Paris VI, France. Disponible en ligne : <https://hal.archives-ouvertes.fr/pastel-00000677>
- Pan, Y., Jiang, J., & Wang, Z. (2007).** Quantitative structure–property relationship studies for predicting flash points of alkanes using group bond contribution method with back-propagation neural network. *Journal of hazardous materials*, 147(1-2), 424-430.
- Pao, H. T., & Fu, H. C. (2013).** Renewable energy, non-renewable energy and economic growth in Brazil. *Renewable and Sustainable Energy Reviews*, 25, 381-392.
- Pigram, G. M., & MacDonald, T. R. (2001).** Use of neural network models to predict industrial bioreactor effluent quality. *Environmental science & technology*, 35(1), 157-162.

**Pratim Roy, P., Paul, S., Mitra, I., & Roy, K. (2009).** On two novel parameters for validation of predictive QSAR models. *Molecules*, 14(5), 1660-1701.

**Randić, M. (1975).** Characterization of molecular branching. *Journal of the American Chemical Society*, 97(23), 6609-6615.

**Randić, M. (1991).** Generalized molecular descriptors. *Journal of Mathematical Chemistry*, 7(1), 155-168.

**Ravi, M., Hopfinger, A. J., Hormann, R. E., & Dinan, L. (2001).** 4D-QSAR analysis of a set of ecdysteroids and a comparison to CoMFA modeling. *Journal of chemical information and computer sciences*, 41(6), 1587-1604.

**Rebas, O. (2006).** "Etablissement de nouvelles corrélations de contribution de groupes avec interactions pour l'estimation des propriétés physiques des hydrocarbures et de leurs mélanges", Mémoire de Magister en Génie des Procédés, Option : Génie Chimique, École Nationale Polytechnique (ENP), Département Génie Chimique, Alger.

**Rebas, O. (2016).** "Évaluation des propriétés des pétroles bruts et des fractions pétrolières par les méthodes de contribution de groupes avec interactions moléculaires", Thèse de doctorat en Génie des Procédés, Option : Génie Chimique, Ecole Nationale Polytechnique (ENP), Département Génie Chimique, Laboratoire de Valorisation des Energies Fossiles, Alger.

**Riazi, M. R. (2005).** Characterization and properties of petroleum fractions (Vol. 50). ASTM international.

**Riazi, M. R., & Roomi, Y. A. (2001).** Use of the refractive index in the estimation of thermophysical properties of hydrocarbons and petroleum mixtures. *Industrial & engineering chemistry research*, 40(8), 1975-1984.

**Riazi, M. R., & Roomi, Y. A. (2001).** Use of the refractive index in the estimation of thermophysical properties of hydrocarbons and petroleum mixtures. *Industrial & engineering chemistry research*, 40(8), 1975-1984

**ROUBEHIE FISSA, M. (2021).** *Estimation de quelques propriétés physico-chimiques des hydrocarbures en fonction de leurs Propriétés (RQSP) en utilisant la méthode de Réseaux de Neurone Artificiel (RNA)* (Doctoral dissertation).

**Roubehie Fissa, M., Lahiouel, Y., Khaouane, L., & Hanini, S. (2019).** QSPR estimation models of normal boiling point and relative liquid density of pure hydrocarbons using MLR and MLP-ANN methods. *Journal of Molecular Graphics and Modelling*, 87, 109-120.

**Roubehie Fissa, M., Lahiouel, Y., Khaouane, L., & Hanini, S. (2023).** Development of QSPR-ANN models for the estimation of critical properties of pure hydrocarbons. *Journal of Molecular Graphics and Modelling*, 121, 108450.

**Roy, K., Kar, S., & Das, R. N. (2015).** A primer on QSAR/QSPR modeling: fundamental concepts. Springer.

**Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1985).** Learning internal representations by error propagation (No. ICS-8506). California Univ San Diego La Jolla Inst for Cognitive Science.

- Samarasinghe, S. (2006).** Neural networks for applied sciences and engineering: from fundamentals to complex pattern recognition. Auerbach publications, Taylor & Francis, Group. Boca Raton New York, USA. ISBN : 978-0-8493-3375-0. ;
- Sarkhosh, M., Ghasemi, J. B., & Ayati, M. (2012).** A quantitative structure-property relationship of gas chromatographic/mass spectrometric retention data of 85 volatile organic compounds as air pollutant materials by multivariate methods. *Chemistry Central Journal*, 6(2), S4.
- Saunders, R. A., & Platts, J. A. (2004).** Scaled polar surface area descriptors: development and application to three sets of partition coefficients. *New Journal of Chemistry*, 28(1), 166-172.
- SIGMA. (2020).** Laboratoire SIGnaux, Modèles, Apprentissage statistique. 10, Rue Vauquelin 75231 PARIS Cedex 05, France. Site Officiel : <https://www.neurones.espci.fr/>. (Consulté le 10/02/2020).
- Si-Moussa, C. (2010).** "Modélisation des équilibres de phases à pressions élevées par les réseaux de neurones artificiels". Thèse de doctorat. École nationale supérieure polytechnique. Alger. Algérie.
- Speight, J. G. (2015).** Handbook of petroleum product analysis. John Wiley & Sons.
- StatSoft 2007 .** STATISTICA (data analysis software system) vs 8.0. StatSoft Inc, Tulsa, OK.
- Statsoft. (2020a).** Statsoft.documentation STATISTICA Help SANN Overviews. Disponible en ligne : [:https://documentation.statsoft.com/STATISTICAHelp.aspx?path=SANN/Overview/SANNOverviewsActivationFunctions](https://documentation.statsoft.com/STATISTICAHelp.aspx?path=SANN/Overview/SANNOverviewsActivationFunctions). (Consulté le 14/01/2020).
- Statsoft. (2020b).** Concepts Fondamentaux en Statistique- Réseaux de Neurones- Fonctions d'Activation. Disponible en ligne : <http://www.statsoft.fr/concepts-statistiques/reseaux-deneurones-automatisees/reseaux-de-neurones-automatisees.htm#fonctions>. (Consulté le 14/01/2020).
- Sugden, S. (1924).** VI.—The variation of surface tension with temperature and some related functions. *Journal of the Chemical Society, Transactions*, 125, 32-41.
- Sushko, I., Novotarskyi, S., Körner, R., Pandey, A. K., Rupp, M., Teetz, W., ... & Tetko, I. V. (2011).** Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information. *Journal of computer-aided molecular design*, 25, 533-554.
- Sutton, R. P. (1985).** Compressibility factors for high-molecular-weight reservoir gases. In SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers.
- Suzuki, K. (Ed.). (2011).** Artificial neural networks: methodological advances and biomedical applications. BoD—Books on Demand.
- Tauler, R., Walczak, B., & Brown, S. D. (2009).** Comprehensive chemometrics: chemical and biochemical data analysis. Elsevier.
- Tetko, I. V., Gasteiger, J., Todeschini, R., Mauri, A., Livingstone, D., Ertl, P., ... & Tanchuk, V. Y. (2005).** Virtual computational chemistry laboratory—design and description. *Journal of computer-aided molecular design*, 19(6), 453-463.

**The MathWorks, Inc. USA (2020)**, Matlab software, Version 2020b. lien: [https://www.mathworks.com/campaigns/products/trials.html?s\\_tid=hp\\_trial\\_download](https://www.mathworks.com/campaigns/products/trials.html?s_tid=hp_trial_download).

**TIBCO Statistica 14.0.0 (2020)**. TIBCO Software Inc, Palo Alto, CA, USA. *Dostupno na*. lien : <https://docs.tibco.com/products/tibco-statistica-14-0-0>. Consulté le 03/03/2024.

**Todeschini, R., & Consonni, V. (2008)**. Handbook of molecular descriptors (Vol. 11). John Wiley & Sons.

**Todeschini, R., & Consonni, V. (2009)**. Molecular descriptors for chemoinformatics: volume I: alphabetical listing/volume II: appendices, references (Vol. 41). John Wiley & Sons.

**Todeschini, R., & Consonni, V. (2009)**. Molecular descriptors for chemoinformatics: volume I: alphabetical listing/volume II: appendices, references (Vol. 41). John Wiley & Sons.

**Topliss, J. G., & Edwards, R. P. (1979)**. Chance factors in studies of quantitative structureactivity relationships. *Journal of Medicinal Chemistry*, 22(10), 1238-1244.

**Tugcu, C. T., Ozturk, I., & Aslan, A. (2012)**. Renewable and non-renewable energy consumption and economic growth relationship revisited: evidence from G7 countries. *Energy economics*, 34(6), 1942-1950.

**Urbani, D. (1995)**. "Méthodes statistiques de sélection d'architectures neuronales : application à la conception de modèles de processus dynamiques". Thèse de doctorat en Microélectronique et Microinformatique, Université Pierre et Marie Curie-Paris VI, France. Disponible en ligne : [https://www.neurones.espci.fr/Theses\\_PS/Urbani\\_D/these.pdf](https://www.neurones.espci.fr/Theses_PS/Urbani_D/these.pdf).

**Vatani, A., Mehrpooya, M., & Gharagheizi, F. (2007)**. Prediction of standard enthalpy of formation by a QSPR model. *International Journal of Molecular Sciences*, 8(5), 407-432

**Vidal, J. (1997)**. Thermodynamique : application au génie chimique et à l'industrie pétrolière. Editions Technip.

**Viswanadhan, V. N., Ghose, A. K., Revankar, G. R., & Robins, R. K. (1989)**. Atomic physicochemical parameters for three dimensional structure directed quantitative structureactivity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics. *Journal of chemical information and computer sciences*, 29(3), 163-172.

**Wauquier, J. P. (1998)**. Raffinage du pétrole (Le). Tome 1. Pétrole brut. Produits pétroliers. Schémas de fabrication (Vol. 1). Editions Technip.

**Weininger, D. (1988)**. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*, 28(1), 31-36.

**Wessel, M. D., & Jurs, P. C. (1994)**. Prediction of reduced ion mobility constants from structural information using multiple linear regression analysis and computational neural networks. *Analytical Chemistry*, 66(15), 2480-2487.

**White, H. (1992)**. Artificial neural networks: approximation and learning theory. Blackwell Publishers, Inc

**Wiener, H. (1947).** Structural determination of paraffin boiling points. *Journal of the American Chemical Society*, 69(1), 17-20.

**Yaws, C. L., & Gabbula, C. (2003).** *Yaws" Handbook of thermodynamic and physical properties of chemical compounds*. Knovel

**Zhou, Z., & Parr, R. G. (1990).** Activation hardness: new index for describing the orientation of electrophilic aromatic substitution. *Journal of the American Chemical Society*, 112(15), 5720-5724.