République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

جامعة المدية
UNIVERSITE DE MEDEA

Faculté de la technologie
Département de Génie des Procédés et Environnement

MEMOIRE DE FIN D'ETUDES

En vue de l'obtention du Diplôme de Master en Génie des Procédés
Option : Pollution Chimique et Gestion Environnementale

Thème

Modélisation semi empirique et neuronale pour la prédiction de la solubilité des solutés solides dans les fluides supercritiques

Réalisé par :

- **BOUREK AbdelFettah**
- **BELKACEM Mohamed**

Proposé et dirigé par :

- **Dr. LAIDI Maamar**
- **Pr. HANINI Salah**

La soutenance aura lieu le **20/05/2017.**

Promotion : 2016/2017

Remerciements

*Au terme de ce Mémoire nous tenons à remercier tout naturellement en premier lieu **ALLAH** qui nous a donné la force, le courage et la patience de bien mener ce travail.*

*Ce travail a été réalisé sous la direction de Monsieur **M. LAIDI**, Maître de Conférences au Pôle Universitaire Yahia Farès de Médéa, qu'il trouve l'expression de notre profonde reconnaissance pour sa compétence, ses encouragements et ses précieux conseils durant le déroulement de ce travail.*

*Un hommage particulier au co-promoteur de cette étude **Mr. S. HANINI** qu'il trouve ainsi l'expression de notre profonde reconnaissance pour sa compétence, ses encouragements et ses précieux conseils durant le déroulement de ce travail.*

*Nos remerciements vont aussi à Monsieur **C. SI MOUSSA** Maître de Conférences au Pôle Universitaire Yahia Farès de Médéa, pour leur aide et encouragements.*

*Nous remercions Monsieur **C. SI MOUSSA**, qui a bien voulu assurer la Présidence de notre jury ainsi que Madame **L. KHAOUANE** pour avoir accepté de juger ce modeste travail.*

Nos sentiments vont également à toutes les personnes qui ont contribué de près ou de loin à la réalisation de ce travail.

Dédicaces

Je dédie ce mémoire à :

***Mon père**, qui peut être fier et trouver ici le résultat de longues années de sacrifices et de privations pour m'aider à avancer dans la vie. Puisse Dieu faire en sorte que ce travail porte son fruit ; Merci pour les valeurs nobles, l'éducation et le soutient permanent venu de toi.*

***Ma mère**, qui a œuvré pour ma réussite, de par son amour, son soutien, tous les sacrifices consentis et ses précieux conseils, pour toute son assistance et sa présence dans ma vie, reçois à travers ce travail aussi modeste soit-il, l'expression de mes sentiments et de mon éternelle gratitude.*

***Mes frères** et **ma sœur** qui n'ont cessé d'être pour moi des exemples de persévérance, de courage et de générosité.*

*A tous les membres de **ma famille.***

*Mes **enseignants** à l'**UYFM** qui doivent voir dans ce travail la fierté d'un savoir bien acquis.*

*À mes **collègues** et **amis** à l'université Dr. Yahia Farès et ma promotion génie des procédés **PCGE** (2016/2017).*

À tous qui m'ont encouragé et soutenu dans mes moments les plus difficiles.

Abdelfettah

Dédicaces

Je dédie ce mémoire à :

***Mon père**, qui peut être fier et trouver ici le résultat de longues années de sacrifices et de privations pour m'aider à avancer dans la vie. Puisse Dieu faire en sorte que ce travail porte son fruit ; Merci pour les valeurs nobles, l'éducation et le soutient permanent venu de toi.*

***Ma mère**, qui a œuvré pour ma réussite, de par son amour, son soutien, tous les sacrifices consentis et ses précieux conseils, pour toute son assistance et sa présence dans ma vie, reçois à travers ce travail aussi modeste soit-il, l'expression de mes sentiments et de mon éternelle gratitude.*

***Mes frères** et **ma sœur** qui n'ont cessé d'être pour moi des exemples de persévérance, de courage et de générosité.*

*A tous les membres de **ma famille.***

*Mes **enseignants** à l'**UYFM** qui doivent voir dans ce travail la fierté d'un savoir bien acquis.*

*À mes **collègues** et **amis** à l'université Dr. Yahia Farès et ma promotion génie des procédés **PCGE** (2016/2017).*

À tous qui m'ont encouragé et soutenu dans mes moments les plus difficiles.

Mohamed

الملخص:

المقاربة المعرفية والحرارية للتنبأ بالذوبانية

الهدف من هذا البحث المتواضع هو توقع ذوبانية الأملاح الصلبة في السوائل فوق الحرجة. مجموعة متكونة من 107 مركب صلب تم إستعمالها في النمذجة الشبه التجريبية والعصبية للذوبانية. نموذج جديد يرتكز على الكثافة تم تقديمه في هذا البحث، النتائج المتحصل عليها أكدت أن النموذج الجديد بمتوسط انحراف 7.45% كان أكثر دقة من المعادلات التجريبية والشبه تجريبية المضافة حديثا. استعملنا أيضا الواسمات الجزيئية كمداخل في النمذجة العصبونية لتوقع ذوبانية الأملاح الصلبة في السوائل فوق الحرجة، النتائج المتحصل عليها أكدت أن الواسمات ذات البعد الثاني بمتوسط انحراف 12.92% كان أكثر دقة من الشبكات العصبونية الأخرى.

كلمات مفتاحية: الذوبانية، السوائل فوق الحرجة، المعادلات الشبه تجريبية، النمذجة، الشبكة العصبونية، QSAR

Résumé :

Approche cognitive et thermodynamique pour la prédiction de la solubilité

Le but de cette modeste recherche est de prédire la solubilité des solutés solides dans les fluides supercritiques. Un ensemble de 107 composés solides ont été utilisé pour la modélisation semi-empirique et neuronale de la solubilité. Un nouveau modèle fondé sur la densité a été proposé dans cette recherche, les résultats ont montré que le nouveau modèle avec un écart moyen d'environ 7,45% était plus précis que les meilleure équation empirique et semi-empirique publiés récemment. Les descripteurs moléculaires sont utilisés comme des entrées dans la modélisation neuronale pour prédire la solubilité des solutés solides dans les fluides supercritiques, les résultats ont montré que le réseau de neurones utilisant les descriptrices 2 dimensions avec un écart moyen d'environ 12,92% était plus précis que les autres réseaux neuronaux.

Mots clés : solubilité, fluides supercritiques, équation semi-empirique, modélisation, réseau de neurones, QSAR.

Abstract :

Cognitive and thermodynamic approach for the prediction of solubility

The aim of this modest research is to predict the solubility of solid solutes in supercritical fluids. A set of 107 solid compounds was used for the semi-empirical and neuronal modeling of solubility. A new model based on density was proposed in this research, the results showed that the new model with an average deviation of about 7.45% was more accurate than the best empirical and semi-empirical equation published recently. Molecular descriptors are used as inputs in neuronal modeling to predict the solubility of solid solutes in supercritical fluids, the results showed that the neural network using 2 dimensional descriptors with an average deviation of about 12.92% was more accurate than the other neural network.

Key words: solubility, supercritical fluids, semi - empirical equation, modeling, neural network, QSAR.

Table des matières I
Liste des figures VI
Liste des tableaux VII
Nomenclature et abréviation VIII
Introduction générale 1

Chapitre I : La solubilité des solutés solides dans le CO_2 supercritique.

I.1. Introduction 4
I.2. Définitions 4
I.2.1. Solubilité 4
I.2.2. Fluide supercritique 4
I.2.2.1. Dioxyde de carbone et autres fluides 6
I.3. Nature d'équilibre 6
I.4. Condition thermodynamique des équilibres 7
I.4.1. Degré de liberté 7
I.4.2. Règle de phase 7
I.4.3. Condition d'équilibre solide – fluide 8
I.5. Fugacité 8
I.6. Règles de mélange 10
I.6.1. Règle de mélange de van der Waal conventionnelle 12
I.6.2. Règle de mélange de Mukhopadhyay et Rao 13
I.6.3. Règle de mélange de Panagiopoulos et Reid (rm-PR) 13
I.6.4. Règle de mélange d'Adachi et Sugi (AS) 14
I.6.5. Règle de mélange avec les grandeurs d'excès 15
I.7. La solubilité des solutés dans les fluides supercritiques 16
I.8. Les modèles de prédiction de la solubilité 17
I.8.1. Les modèles semi empirique 17
I.8.1.1. Modèle de Chrastile 18
I.8.1.2. Modèle de del Valle et Aguilera 18
I.8.1.3. Modèle de Bartle 19
I.8.1.4. Modèle de Mendez-Santiago et Teja 19
I.8.1.5. Modèle de Kumar et Johnston 20
I.8.2. Les équations d'état 20

I.8.2.1. Equation d'état du gaz parfait – loi des états correspondants 21
I.8.2.2. Equations d'état cubique 23
I.8.2.2.1. L'équation d'état de Redlich et Kwong (RK) 24
I.8.2.2.2. L'équation d'état de Soave Redlich et Kwong (SRK) 25
I.8.2.2.3. L'équation d'état de Peng et Robinson (PR) 26
I.8.2.2.4. L'équation d'état de Patel-Teja-Valderrama (PTV) 27
I.8.2.2.5. L'équation de Peng et Robinson modifiée (MPR) 28
I.8.2.3. Equations d'état non cubique (basée sur la théorie de perturbation) 29
I.8.2.3.1. Contribution des segments 30
I.8.2.3.2. Contribution des chaînes 31
I.8.2.3.3. Contribution d'association 31
I.8.2.3.4. Equation d'état de PAZUKI 31
I.8.2.3.5. Equation d'état PC-SAFT 32
I.9. Intelligence artificielle 33
I.9.1. Définition 33
I.9.2. Réseau de neurone artificiel 33
I.9.3. Algorithmes génétiques 33
I.9.4. Les systèmes experts 34
I.9.5. La Logique floue 34
I.10. Conclusion 35

Chapitre II : Le réseau de neurone artificiel.

II.1. Introduction 36
II.2. Historique 37
II.3. Fondements Biologiques 37
II.4. Neurone Formel 39
II.5. Analogie avec le modèle biologique 40
II.6. Réseaux de neurones artificiels 41
II.7. Catégories des réseaux de neurones artificiels 42
II.7.1. Réseaux de neurones multicouches statiques 42
II.7.2. Réseaux de neurones dynamiques 43

II.8. Apprentissage des réseaux de neurones 43

II.8.1. Types d'apprentissages 44

II.8.1.1. Apprentissage non supervisé 44

II.8.1.2. Apprentissage forcé ou imposé 45

II.8.1.3. Apprentissage supervisé 45

II.8.2. Algorithme de la rétro-propagation 46

II.8.2.1. Résumé de l'algorithme de la rétro-propagation 50

II.9. Fonctions d'activation 50

II.10. Procédure de modélisation neuronale 51

II.10.1. Sélection des données 52

II.10.2. Collecte des données 53

II.10.3. Séparation des bases de données 53

II.10.4. Normalisation des données 53

II.10.5. Choix du type de réseau de neurones 54

II.10.6. Choix du nombre de neurones dans la couche cachée 54

II.10.7. Choix de la fonction de transfert 54

II.10.8. Estimation des paramètres ou apprentissage 54

II.10.9. Validation et généralisation 55

II.11. Conclusion 55

Chapitre III : Modélisation de la solubilité des solutés solides dans le CO_2 supercritique.

III.1. Contribution I : modélisation empirique et semi-empirique de la solubilité des solutés solides dans le dioxyde de carbone

III.1.1. Introduction 56

III.1.2. Elaboration de la base de données 56

III.1.2.1. Introduction 56

III.1.2.2. Structure de la base de données 56

III.1.3. L'ajustement des paramètres 58

III.1.3.1. L'algorithme de l'ajustement des paramètres 59

III.1.4. Les modèles de calcul de la solubilité 60

III.1.4.1. Les modèles basés sur la densité du solvant 60

III.1.4.1.1. Présentation des résultats trouvées 60

III.1.4.1.2. Discussion des résultats trouvées 64

III.1.4.2. Les modèles proposés (série 01) 65

III.1.4.2.1. Présentation des résultats trouvées 65

III.1.4.2.2. Discussion des résultats trouvées 66

III.1.4.3. Les modèles proposés (série 02) 67

III.1.4.3.1. Présentation des résultats trouvées 67

III.1.4.3.2. Discussion des résultats trouvées 69

III.1.5. Comparaison entre les meilleures modèles 70

III.1.6. Validation 74

III.1.7. Conclusion 76

III.2. Contribution II : modélisation neuronale de la solubilité des solutés solides dans le CO_2 supercritique

III.2.1. Introduction 78

III.2.2. La méthode QSAR 78

III.2.2.1. Introduction 78

III.2.2.2. Relations activité – structure 79

III.2.2.2.1. Idéologie 79

III.2.2.2.2. Objectifs du développement du modèle QSAR 80

III.2.2.2.3. Prédiction de l'activité / propriété / toxicité 81

III.2.2.3. Développement historique de QSAR 81

III.2.2.4. Applications de QSAR 86

III.2.3. Logiciels et interfaces utilisés 87

III.2.3.1. DRAGON 7.0 87

III.2.3.2. Algorithme génétique 88

III.2.4. La méthodologie de l'analyse QSAR 90

III.2.4.1. Les descripteurs moléculaires 90

III.2.4.2. La base de données 90

III.2.4.3. Réduction des variables 90

A. Réduction par DRAGON 7.0 91

B. Réduction par l'algorithme génétique 91

III.2.5. Modélisation par le réseau de neurone artificiel 92

III.2.5.1. Structure de la base de données 92

III.2.5.2. Conception de la topologie de RN 93

III.2.5.2.1. Choix des entrées et des sorties 94

III.2.5.2.2. Préparation de la base de données 94

III.2.5.2.3. Normalisation des données 94

III.2.5.3. Résultats et discussion 95

III.2.6. Equation du RNA 99

III.2.7. Validation 99

III.2.7. Conclusion 101

Conclusion générale 103

Chapitre I : La solubilité des solutés solides dans le CO_2 supercritique.

Figure I.1: Diagramme PT d'une substance pure	5
Figure I.2. Les étapes de formation de modèle moléculaires à partir de théorie associative SAFT.	30
Figure I.3. Le modèle moléculaire dans un mélange de deux substances selon l'équation d'état PC-SAFT.	32

Chapitre II : Le réseau de neurone artificiel.

Figure II.1. Neurone biologique.	38
Figure II.2. Le neurone formel.	39
Figure II.3. Réseau multicouche MLP.	43
Figure II.4. Apprentissage non supervisé.	45
Figure II.5. Apprentissage supervisé	45
Figure II.6. Réseau de neurones multicouche	47

Chapitre III : Modélisation de la solubilité des solutés solides dans le CO2 supercritique.

Figure III.1. Le nombre des systèmes corrélé et le nombre des paramètres pour chaque modèle.	64
Figure III.2. Le nombre des systèmes corrélés et le nombre des paramètres pour chaque modèle.	67
Figure III.3. Le nombre des systèmes corrélé et le nombre des paramètres pour chaque modèle.	69
Figure III.5. Le nombre des systèmes corrélé et le nombre des paramètres pour chaque modèle.	72
Figure III.5. Diagramme de dispersions de modèle Si-Moussa et al. (Eq. (9))	73
Figure III.6. Diagramme de dispersions de modèle proposé, série 01 (Eq. (12))	73
Figure III.7. Diagramme de dispersions de modèle proposé, série 02 (Eq. (17))	74
Figure III.8. Solubilité expérimentale de la curcumine in $scCO_2$ et prédire par les modèles en fonction de la densité du fluide supercritique (graphe 2D)	76
Figure III.9. Solubilité expérimentale de la curcumine in $scCO_2$ et prédire par les modèles en fonction de la densité du fluide supercritique et la pression (graphe 3D)	76
Figure III.10. La méthode QSAR visent à trouver la corrélation entre les propriétés structurelles des produits chimiques et leur activité.	80
Figure III.11. Un cercle représente le pourcentage de chaque type des descripteurs dans la base de données de logiciel DRAGON 7.0	88
Figure III.12. Méthodologie de calcul de la solubilité par les réseaux de neurones.	93
Figure III.13. Le nombre des systèmes corrélés et le nombre d'entrées pour chaque réseau.	98
Figure III.14. Solubilité expérimentale de la curcumine in $scCO_2$ et prédire par le RNA 2D2D en fonction de la densité du fluide supercritique (graphe 2D)	100
Figure III.15. Solubilité expérimentale de la curcumine in $scCO_2$ et prédire par le RNA 2D2D en fonction de la densité du fluide supercritique et la pression (graphe 3D)	100

Chapitre II : Le réseau de neurone artificiel.

Tableau II.1. Correspondance entre les modèles neuronaux biologiques et artificiels. 41

Tableau II.2. Fonctions de transferts usuelles, a=f(t). 51

Chapitre III : Modélisation de la solubilité des solutés solides dans le CO2 supercritique.

Tableau III.1. La liste des composés utilisés dans la modélisation 57

Tableau III.2. Modèles basés sur la densité pour la corrélation de solutés solides dans des fluides supercritiques. 61

Tableau III.3. Résultats de calcul des AARD et R² pour les modèles sélectionnés 62

Tableau III.4. Modèles proposés pour la corrélation de solutés solides dans des fluides supercritiques (série 01). 65

Tableau III.5. Résultats de calcul des AARD et R² pour les modèles proposés (série 01) 65

Tableau III.6. Modèles proposés pour la corrélation de solutés solides dans des fluides supercritiques (série 02). 67

Tableau III.7. Résultats de calcul des AARD pour les modèles proposés (série 02) 68

Tableau III.8. Résultats de calcul des AARD pour les meilleurs modèles sélectionnés 70

Tableau III.9. Le vecteur de validation ($\boldsymbol{\alpha}$ [pente], $\boldsymbol{\beta}$ [interception de y] et $\boldsymbol{R}$ [coefficient de corrélation]) calculé pour les trois modèles sélectionnés. 74

Tableau III.10. Les valeurs de ($\boldsymbol{T}(°\boldsymbol{K})$, $\boldsymbol{P}(\boldsymbol{MPa})$, $\boldsymbol{\rho_1}(\boldsymbol{kg/m^3})$ et $\boldsymbol{y^{exp}}$) sont mesuré par (Zhan et al., 2017) et les valeurs ($\boldsymbol{y_{M9}^{cal}}$, $\boldsymbol{y_{M12}^{cal}}$, $\boldsymbol{y_{M17}^{cal}}$) sont calculés dans cette étude. 75

Tableau III.11. Les paramètres ainsi que leur valeur idéale et seuil comme employé dans la fonction de remise en forme de l'outil d'algorithme génétique interne. 89

Tableau III.12. Liste des classes des descripteurs introduit dans le logiciel DRAGON 7.0 90

Tableau III.13. Représentation des résultats de réduction par le logiciel DRAGON 7.0. 91

Tableau III.14. Représentation des résultats de réduction par l'algorithme génétique. 91

Tableau III.15. Topologies des RNAs utilisés 93

Tableau III.16. Architecture des RNAs utilisés 95

Tableau III.17. Résultats des AARDs et R² pour les 107 système (solutés-scCO$_2$) 96

Tableau III.18. Les valeurs de ($\boldsymbol{T}(°\boldsymbol{K})$, $\boldsymbol{P}(\boldsymbol{MPa})$, $\boldsymbol{\rho_1}(\boldsymbol{kg/m^3})$ et $\boldsymbol{y^{exp}}$) sont mesuré par (Zhan et al., 2017) et les valeurs des descripteurs (SpMax4_Bh(s), SM06_AEA(dm), nN, LLS_01) calculé par le logiciel DRAGON7.0 99

Nomenclature

b_i	Biais du neurone.
f	Fugacité
$g_n^{hs}(\eta)$	Fonction de distribution radiale moyenne.
{e ; (n_m) ; s}	Nombre de neurones dans la couche d'entrée, (e).
	Nombre de neurones dans la 1ère couche cachée, (n).
	Nombre de neurones dans la 2ème couche cachée, (m).
	Nombre de neurones dans la couche de sortie, (s).
P_c	Pression critique, (bar).
P^{sub}	Pression de sublimation, (bar).
R	Coefficient de corrélation.
S	Solubilité.
T_c	Température critique, (K).
T_r	Température réduite, (K).
μ_k / kT	Contribution de la perturbation dans l'équation PC-SAFT.
v	Variance.
v_i^s	Volume molaire, (cm^3/mole).
$w_{k,j}$	Poids synaptiques du neurone entre le neurone, (k) et le neurone, (j).
y_2	Fraction molaire du soluté solide.
$\alpha(Tr)$	Fonction dans le paramètre attractive d'une équation d'état**.**
ϕ	Coefficient de fugacité.
ρ_1	Densité du solvant, (kg/m^3).
ρ_{fs}	Masse volumique du fluide à l'état SC, (kg/m^3).
ρ_{liq}	Masse volumique du fluide à l'état liquide, (kg/m^3).
μ	Potentiel chimique, (J).
σ_{ii}	Diamètre du segment, (A°).

Abréviation

2D	2 Dimension
3D	3 Dimension
AG	Algorithme génétique
RNA	Réseau de Neurones Artificiels
Qsar	Quantitative Structure – Activity Relationship
Sar	Structure – Activity Relationship

Introduction générale

Avant qu'un médicament soit disponible à ses récepteurs, il doit être dissous dans le liquide biologique entourant la solubilité des récepteurs et par conséquent, est un sujet important en sciences pharmaceutiques. Il y a un problème de solubilité avec près de 40 % des candidats médicaments, et toute tentative de prédire la solubilité est très important dans le processus de découverte de médicaments d'enquêtes. La plus ancienne règle pour la prédiction de la solubilité est "like dissolves like". Cette règle est appliquée à toutes équations de solubilité, (Jouyban, 2010). La solubilité est définie comme la quantité maximale d'un médicament dissous dans un volume donné d'un solvant/solution.

La technologie d'extraction de fluide supercritique (SFET) a attiré beaucoup d'attention dans la séparation et la purification des substances naturelles. En particulier, le CO_2 supercritique ($scCO_2$) offre des caractéristiques intéressantes en tant que solvant non toxique, non inflammable, à faible coût et facile à traiter, avec des propriétés critiques modérées, (Zabihi *et al.*, 2010; Zhu *et al.*, 2015). Afin de développer un système efficace d'extraction ou de purification supercritique, il est nécessaire d'avoir une idée approfondie du comportement de la phase et des propriétés d'équilibre de la substance cible dans les milieux supercritiques, (Ismadji *et al.*, 2011).

Le principe des méthodes QSAR (*Quantitative Structure Activity Relationship*) consiste comme leurs noms l'indiquent à mettre en place une relation mathématique à l'aide de méthodes d'analyse de données, reliant des propriétés moléculaires microscopiques appelées descripteurs, à un effet expérimental (activité biologique, toxicité, affinité pour un récepteur), pour une série de composés chimiques similaires. Le point de départ de telles méthodes se construit sur la définition des descripteurs moléculaires empiriques ou théoriques. Ces dernières prennent en compte des informations sur la structure et les caractéristiques physico-chimiques des molécules, comme dans ce cas présent, les perturbateurs endocriniens.

Les descripteurs théoriques se déclinent eux-mêmes en sous-classes de descripteurs : les constitutionnels, basés sur la composition chimique de la molécule, les topologiques, obtenus à partir de la structure bidimensionnelle (table de connectivité des atomes de la molécule), les géométriques, évalués à partir de la structure tridimensionnelle et les quantiques, issus de la structure électronique de la molécule. Les détails concernant les descripteurs quantiques sont disponibles dans l'ouvrage de Karelson, (Karelson, 2000).

Enfin, le lien entre les descripteurs et la base de données est déterminé grâce à des outils d'analyse comme les régressions linéaires multiple (MLR), les régressions aux moindres carrés partiels (PLS), les arbres de décisions, les réseaux de neurones, et les algorithmes génétiques.

L'apprentissage connexionniste est une discipline scientifique qui recouvre plusieurs aspects d'études mathématiques, statistiques et algorithmiques. Les systèmes d'apprentissage connexionnistes (ou réseaux de neurones artificiels [RNA]) sont des systèmes numériques permettant la modélisation de processus généraux par l'établissement de modèles fonctionnels. Ceux-ci sont identifiés à partir des entrées-sorties du processus par des algorithmes dits "d'apprentissage" qui s'apparentent à des techniques d'estimation statistiques. Nés en informatique dans le domaine de l'intelligence artificielle, ils ont connu depuis le début des années 80 un développement intensif dû au succès rencontré dans une très large gamme d'applications. Les réseaux de neurones artificiels offrent une panoplie de techniques adaptatives pour de nombreux problèmes génériques : la classification, le classement, la modélisation, la prévision. Parmi les domaines où ils ont été utilisés, on peut citer le traitement du signal, de l'image, de la parole, le contrôle, la prévision, l'optimisation combinatoire, le diagnostic, la fouille de données, ...etc. Il s'agit donc d'un outil général pouvant être utilisé dans des cadres différents avec une grande souplesse. Les modèles RNAs apparaissent comme une alternative ou plus souvent comme un complément aux techniques classiques. Les recherches dans le domaine sont extrêmement actives actuellement. Plusieurs laboratoires tant industriels ou universitaires participent à cette recherche en Europe, États-Unis et au Japon. Différentes voies ont déjà été explorées et les réseaux de neurones artificiels sont devenus un outil incorporé dans différents logiciels commercialisés pour les statistiques, la fouille de données, le traitement du signal et l'identification de systèmes.

L'objectif du présent mémoire est de développer deux méthodologies pour prédire la solubilité des solutés solides dans les fluides supercritiques, la première c'est de développer un modèle semi-empirique ou on a utilisé les variables intensives comme des entrées et la deuxième c'est de développer un modèle basé sur la théorie des réseaux de neurones artificiels ou on a utilisé les variables intensives et les descripteurs moléculaires comme des entrés.

Dans ce contexte notre travail est organisé en trois chapitres suivis d'une synthèse bibliographique aussi exhaustive que possible :

Dans le premier chapitre, après une introduction générale, nous rappelons les notations de la solubilité, les fluides supercritiques et les modèles de prédiction de la solubilité.

Dans le deuxième chapitre, nous parlons brièvement sur le réseau de neurones artificiel, les étapes de l'utilisation d'un réseau neuronale.

Dans le troisième chapitre, nous attaquons le but de ce travail qui est la modélisation, et nous avons divisé ce chapitre en 2 contributions. La première contribution vise la modélisation semi-empirique, el la deuxième contribution vise la modélisation neuronale pour prédire la solubilité des solutés solides dans ls fluides supercritiques.

Pour chaque système, une comparaison, d'une part entre les valeurs des modèles introduits dans ce travail, de réseau de neurones artificiel (RNA) et les données expérimentales est réalisée en termes d'erreurs.

Enfin une conclusion générale qui rassemble les résultats obtenus dans cette étude et des perspectives qui peuvent être envisagées pour le futur.

CHAPITRE 01 :
La solubilité des solutés solides dans le CO_2 supercritique.

I.1. Introduction

L'industrie chimique a connu un développement étonnant dans ces dernières années notamment dans ces procédés de point de vue technologique et ça se voie dans la qualité d'équipements mis en œuvre ainsi que l'utilisation de l'outil informatique dans les différentes étapes de l'étude (simulation, modélisation, …).

L'industrie pharmaceutique ne sorte pas de cette règle et on peut la définir comme un ensemble des procédés qui visent à transformer les matières premières aux produits finis c'est-à-dire des médicaments qui doivent être de bonnes qualité, efficaces et compétents lorsqu'on les met sur marché, ceci signifié qu'il faut avoir un bon procédé de production c'est à dire maîtriser et optimiser tous les paramètres qu'ils peuvent être trouvés, y compris la solubilité des soluté solides dans les fluides plus précisément les fluides supercritiques qui possèdent beaucoup d'avantages.

L'objectif de ce premier chapitre c'est de présenter brièvement les modèles qui calcul la solubilité.

I.2. Définitions

I.2.1. Solubilité

De façon générale, la solubilisation d'une substance consiste à dissoudre celle-ci dans un solvant donné. Cette solubilisation ou dissolution n'est possible que si les liaisons de type solvant-solvant et soluté-soluté peuvent être remplacées par des liaisons mixtes solvant soluté.

Ces dernières peuvent être de différentes natures : ion-dipôle, dipôle-dipôle, liaison hydrogène, liaison Van der Waals ou complexation. Quant à la solubilité c'est la concentration à laquelle la solution est en équilibre avec le solide que l'on a dissout. La solubilité se définit toujours par rapport à une solution de composition donnée, à température et pression fixées, en équilibre avec une phase cristalline de composition et structure données également. A l'équilibre, le potentiel chimique du cristal (μ_c) est égal à celui de la solution (μ_2), (Lafferrere, 2002).

I.2.2. Fluide supercritique

Un fluide sur sa température et sa pression critique, présentant un bon pouvoir solvant. Dans la plupart des applications, le dioxyde de carbone est utilisé pur ou ajouté avec un co-solvant.

L'intérêt principal des fluides supercritiques est lié à leurs propriétés « réglables », qui peuvent être facilement modifiées en surveillant la pression et la température : bon pouvoir solvant à des densités élevées (température proche de la température critique et pression beaucoup plus que la pression critique) à très faible pouvoir solvant à Faible densité (température proche ou supérieure de la température critique et pression plus basse pression critique).

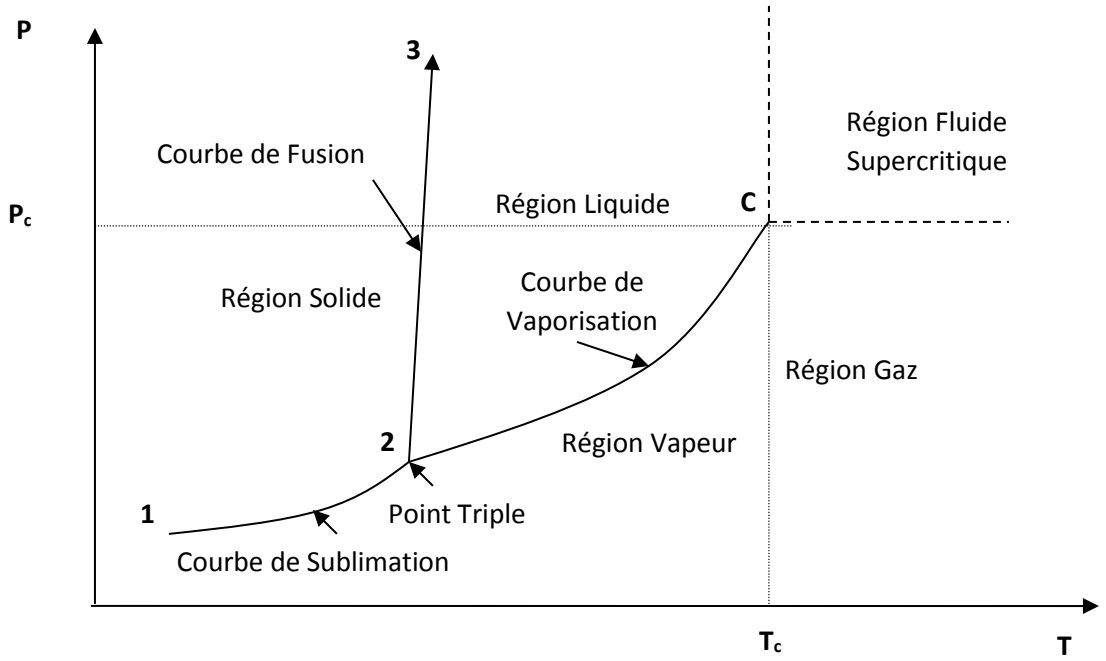


Figure I.1: Diagramme PT d'une substance pure

I.2.2.1. Dioxyde de carbone et autres fluides

Le dioxyde de carbone : Le CO2 est un fluide supercritique très attractif pour plusieurs raisons :

- Très bon marché et abondant sous forme pure (alimentaire) dans le monde entier.
- Non inflammable et non toxique
- Respectueux de l'environnement, en tant que gaz non polluant et que la majeure partie du CO2 est fabriquée à partir de flux de déchets (principalement des usines d'engrais des effluents gazeux).
- Température critique à 31 ° C, permettant des opérations à une température proche de la température ambiante, évitant toute altération du produit.
- Pression critique à 74 bar, entraînant une pression de service « acceptable », généralement entre 100 et 350 bar.

Cependant, le dioxyde de carbone se comporte toujours comme un solvant "non polaire" qui dissout sélectivement les lipides qui sont des composés insolubles dans l'eau comme les huiles végétales, le beurre, les graisses, les hydrocarbures, etc.

Le dioxyde de carbone ne dissout pas les composés hydrophiles comme les sucres et les protéines, Les espèces minérales comme les sels, les métaux, etc.

Co-solvant : solvant organique ajouté au fluide principal (généralement du dioxyde de carbone) pour modifier sa puissance de solvant vis-à-vis des molécules « polaires » car le fluide lui-même ne peut dissoudre que des molécules « non polaires ». En général, le co-solvant est choisi parmi les alcools à chaîne courte, les esters ou les cétones. Pour des raisons évidentes, l'éthanol est préféré dans de nombreux cas car il est abondant et bon marché sous des formes pures (qualité alimentaire, grade pharmacopée), pas dangereux pour l'environnement et pas très toxique.

Le dioxyde de carbone est disponible soit dans des bouteilles commerciales ou des sphères de volumes limités, soit en vrac à partir de réservoirs contenant entre 4 et 30 tonnes de gaz liquéfié à une pression de 18 bar et à une température autour de -18 ° C. Ces réservoirs sont alimentés directement à partir de camions cryogéniques comme indiqué sur l'image ci-dessous. Le prix est très sensible à la quantité livrée car le transport est le principal coût parce que le dioxyde de carbone est disponible à très bas prix dans diverses usines chimiques (principalement des usines d'ammoniac).

I.3. Nature d'équilibre

L'équilibre est une condition statique où les propriétés macroscopiques ne subissent aucune variation avec le temps. Ceci implique un équilibre de tous les potentiels ou gradients, qui pourraient causer leur variation, (Abdallah el hadj, 2009).

Cependant, à l'échelle microscopique, les conditions ne sont pas statiques, les molécules qui forment une phase à un instant donné ne sont plus les même à un autre instant, les molécules ayant des vitesses élevées et qui sont près de l'interface arrivent à vaincre les tensions superficielles et passent à l'autre phase. Cependant le taux de passage dans les deux directions est le même, (Abdallah el hadj, 2009).

I.4. Condition thermodynamique des équilibres

La plus grande partie du domaine de l'équilibre des phases est régie par deux relations théoriques fondamentales : La première est la règle de phase, qui définit l'allure générale du diagramme des phases, la seconde est la formule de Clapeyron qui détermine la pente des courbes de ce diagramme.

I.4.1. Degré de liberté

Il est nécessaire, pour définir complètement un système, d'indiquer les valeurs numériques d'un certain nombre de variables, que l'on choisit par exemple parmi la pression, la température, l'énergie interne, l'entropie et les concentrations des différentes phases.

Parmi les variables, on distingue les variables extensives (qui dépendent de la masse) et les variables intensives dites variables d'état.

Un équilibre entre phase ne dépendant pas des quantités réelles des diverses phases en présence, on ne tiendra compte que des variables intensives, comme la température, la pression et la concentration pour le décrire.

On peut faire varier un certain nombre de ces variables de manière indépendante, mais les valeurs des autres sont déterminées par les valeurs choisies et par les conditions thermodynamiques d'équilibre. Alors, le degré de liberté ou variance du système est défini par le nombre de variables d'état qu'on peut faire varier de manière indépendante sans modifier le nombre de phase, (Hadj-Kali, 2004).

I.4.2. Règle de phase

La règle de phase proposée par Gibbs fournit une relation générale entre le nombre ν de degré de liberté d'un système, le nombre de phase φ et le nombre de constituants C. Elle exprime qu'on a toujours pour un système non réactionnel :

$$v = C + 2 - \varphi \qquad \text{I.1}$$

Par exemple, pour un système binaire (C = 2) en équilibre solide- fluide (φ = 2), la règle de phase indique que la variance est égale à : v =2. Ceci signifie que l'on peut faire varier de manière

indépendante deux variables d'état sans modifier le nombre de phases. Ces variables sont en général choisies parmi la température, la pression et la composition.

I.4.3. Condition d'équilibre solide – fluide

Pour chaque constituant *i* du mélange, la condition d'équilibre thermodynamique entre phase est donnée par l'égalité des potentiels chimiques des phases qui se réduit, par définition de la fugacité, à l'égalité des fugacités dans chaque phase pour chaque constituant. Dans le cas d'un équilibre solide-fluide, (Prausnitz *et al.*, 1986a).

$$\mu_i^s = \mu_i^f \qquad \text{I.2}$$

Le problème fondamental est de relier ces fugacités aux compositions de chacune des phases. Dans ce qui suit, nous négligeons les effets dus aux forces de surface, gravité, champs électriques ou magnétiques et membranes semi-perméables. Alors, la fugacité d'un constituant dans un mélange ne dépend que de la température, de la pression et de la composition de ce mélange. On exprimera les compositions de solide dans la phase fluide (dans notre cas : un fluide supercritique) par la fraction molaire, y_i. Si la pression est faible et le mélange idéal (peu d'interactions entre les molécules présentes dans le mélange).

I.5. Fugacité

Cette propriété, qui signifie la tendance d'un système à occuper un plus grand volume, ou bien la tendance à la dispersion a été introduite par *LEWIS* pour tenir compte l'écart entre le comportement du gaz parfait et celui du gaz réel.

Le potentiel chimique d'un gaz idéal (ou parfait) est donné par :

$$\mu^{gp}(T, p) = \mu^{\circ}(T) + RT \ln \frac{P}{P^{\circ}} \qquad \text{I.3}$$

Pour simplifier la notation la pression standard (P^0 = 1 bar) sera omise dorénavant (il faut travailler en bar, donc) et les dépendances avec *T* ou *P* aussi. Dans le cas d'un mélange de gaz parfaits, pour chaque constituant :

$$\mu_i^{gp} = \mu_i^0 + RT \ln y_i p \qquad \text{I.4}$$

Où $y_i p = p_i$ est la pression partielle du constituant i.

Un fluide réel (gaz ou liquide) n'obéit pas à cette relation linéaire entre le potentiel chimique et le logarithme de la pression, mais il serait convenable de garder une forme équivalente pour les équations, en remplaçant la pression par une fugacité, soit une sorte de pression efficace. Pour une substance pure, la fugacité *f* est définie par :

$$\mu(T, p) \equiv \mu^*(T) + RT \ln \frac{f}{p^\circ} \qquad \text{I.5}$$

Où μ^* est le potentiel chimique quand la fugacité vaut 1 bar.

Bien sûr, la fugacité est fonction de la température et de la pression, car elle prend en compte la non idéalité du fluide. Le coefficient de fugacité est le rapport entre la fugacité et la pression, donc une grandeur sans dimension :

$$\phi \equiv \frac{f}{p} \qquad \text{I.6}$$

Pour un gaz idéal, $f = p$ et $\phi = 1$. Jusqu'ici, un fluide réel à l'état pur a été considéré, et la non idéalité introduite sous forme de fugacité concerne les écarts par rapport au gaz idéal sous effet de la pression et de la température.

Dans la situation plus générale d'un mélange de fluides réels, la composition affecte aussi les écarts à l'idéalité. La fugacité de chaque constituant i est définie par,(Ashour *et al.*, 2000):

$$\mu_i = \mu_i^* + RT \ln f_i \qquad \text{I.7}$$

Comme les phases en équilibre sont obligatoirement à la même température, la condition d'équilibre de phases, équation (I.2), correspond à une égalité de fugacités :

$$f_i^s = f_i^{SF} \qquad \text{I.8}$$

i= 1, NC, avec NC : nombre des composés

Où f_i^s est la fugacité du composé *i* dans le solide(s) et f_i^{SF} est la fugacité de *(i)* dans la phase supercritique. Cette fugacité peut être calculée par la relation suivante :

$$f_i^{SF} = y_i \phi_i^{SF} p \qquad \text{I.9}$$

Où ϕ_i^{SF} est le coefficient de fugacité de composé *(i)* dans la phase supercritique, et *P* est la pression du système.

Le coefficient de fugacité peut être calculé par la relation suivante :

$$RT \ln \phi_i = \int_V^{\infty} \left[\left(\frac{\partial P}{\partial n_i} \right)_{T,V,nj \neq n_i} - \frac{RT}{V} \right] dV - \ln Z \qquad \text{I.10}$$

Où *R* est la constante des gaz, *V* est volume total du système, n_i et n_j sont les nombres de moles des composés *i* et *j*, et *Z* c'est facteur de compressibilité.

La fugacité du composé i comme une phase solide pure est :

$$f_i^s = P_i^{sub} \phi_i^{sub} \exp \left[\int_{P^{sub}}^{P} \left(\frac{v_i^s}{RT} \right) dP \right] \qquad \text{I.11}$$

Où P^{sub} est la pression de sublimation de solide pure à la température de système, v_i^s est le volume molaire de solide pure, ϕ_i^{sub} est le coefficient de fugacité du solide pure.

Le mot fugacité est lié à la tendance du composant de *"fuir"* la phase en question. Ainsi dans l'exemple de l'eau liquide en équilibre avec sa vapeur, si la fugacité de l'eau liquide est égale à celle de la vapeur, les tendances de fuite sont identiques et le système se trouve en équilibre. Si la fugacité de la vapeur est supérieure à celle de l'eau liquide, alors la vapeur condense ; dans le cas opposé, l'eau s'évapore.

I.6. Règles de mélange

Comme nous l'avons signalé au début de ce chapitre, les équations d'état ont été développées initialement pour des fluides purs. Puis, leurs utilisations sont prolongées pour les mélanges. Cette extension nécessite l'introduction de règles de mélange permettant le calcul des paramètres de mélange en fonction de ceux des corps purs. Cette introduction fait intervenir à travers ces règles la régression ou l'ajustement des paramètres d'interactions binaires (dans le cas

d'un mélange binaire) à partir des données expérimentales dans le but d'améliorer la prédiction obtenue par ces équations d'état. Dans ce qui suit on va détailler les différentes règles de mélanges les plus utilisées dans les processus de calcul.

I.6.1. Règle de mélange de Van der Waal conventionnelle

La règle de mélange la plus répandue est sans doute la règle de mélange conventionnelle où la règle de Van der Waals, (Kwak and Mansoori, 1986) qui permet de calculer les coefficients de mélange a_m et b_m par les deux relations :

$$a_m = \sum_i^{nc} \sum_j^{nc} y_i y_j a_{ij} \qquad \text{I.12}$$

$$b_m = \frac{1}{2} \sum_i^{nc} \sum_j^{nc} y_i y_j b_{ij} \qquad \text{I.13}$$

Où nc est le nombre des composés dans le mélange, a_{ij} et b_{ij} sont des paramètres qui peuvent être obtenus à partir de règle de mélange :

$$a_{ij} = \sqrt{a_i a_j}\left(1 - K_{ij}\right) \qquad \text{I.14}$$

$$b_{ij} = \frac{b_i + b_j}{2}\left(1 - l_{ij}\right) \qquad \text{I.15}$$

Avec a_i, b_j sont les paramètres de corps pur et comme on peut le constater cette règle est la règle de Van der Waals à deux paramètres ajustables (VdW2). Généralement, l_{ij} est pris égal à zéro, ce qui permet d'écrire le second coefficient sous la forme :

$$b = \sum_{i=1}^{nc} y_i b_i \qquad \text{I.16}$$

Et c'est la règle de Van der Waal à un paramètre ajustable (VdW1).

Les deux paramètres K_{ij} et l_{ij} sont des paramètres d'interactions binaires ajustables, ils expriment les forces d'interaction exercées entre les différents types de particules i et j qui constituent le mélange. Ils sont usuellement déterminés par la régression des données expérimentales. Leurs valeurs sont généralement comprises entre -1 et +1 et ils affectent la

précision de prédiction des phases des équilibres. Si la valeur K_{ij} est supérieur à +1 la valeur de paramètre a_{ij} devient négative ce qui signifie que le terme attractif donne une contribution positive lorsqu'on calcule la pression qui fait violer les postulats de la théorie cinétique des gaz.

I.6.2. Règle de mélange de Mukhopadhyay et Rao

Mukhopadhyay et Rao ont constaté que l'augmentation des paramètres ajustables binaires n'augmente pas nécessairement la précision de calcul. Ils ont donné une règle de mélange avec un seul paramètre d'interaction. Cette règle de mélange (Mukhopadhyay and Rao, 1993) est immergée à partir de la supposition que dans un mélange supercritique où les molécules ont une large différence de tailles et la probabilité d'avoir lieu une interaction entre deux molécules dépend de surface des molécules plus que la fraction molaire de ces derniers (compositions). De plus, leurs forces répulsives augmentent considérablement. Dans cette règle le co-volume inverse *(b_{ij})* a été introduit dans le paramètre d'énergie d'attraction *(a)* en vue de prendre en compte l'asymétrie.

$$a = \sum_{i}^{nc}\sum_{j}^{nc} y_i y_j a_{ij}\left(\frac{b}{b_{ij}}\right)^{m_{ij}} \qquad \text{I.17}$$

$$b = \sum_{i=1}^{nc} y_i b_i \qquad \text{I.18}$$

Où :

$$a_{ij} = \sqrt{a_i a_j} \quad \text{et} \quad b_{ij} = \sqrt{b_i b_j} \qquad \text{I.19}$$

Le choix des paramètres $m_{ii} = m_{jj} = 1$ diminue la sensibilité de paramètre ajustable m_{ij} et rétréci les paramètres ajustables à un seul paramètre. De plus cette règle est développée à être utilisée pour la prédiction les propriétés des mélanges solide-fluide supercritiques donnant ainsi des avantages par rapport les autres règles, (Mukhopadhyay and Rao, 1993).

I.6.3. Règle de mélange de Panagiopoulos et Reid (rm-PR)

Panagiopoulos et Reid (Panagiotopoulos and Reid, 1986) ont utilisé les deux paramètres ajustables K_{ij}, K_{ji}, ce qui signifie que l'effet de composé *i* sur le composé *j* est déférent de l'effet

de composé *j* sur le composé *j*. ceci est juste pour les interactions solide-fluide supercritiques où il y a une grande déférence entre les tailles des molécules et les entre les forces d'interactions. La forme de règle des mélanges pour le calcul de a_{ij} pour les systèmes binaires peut être écrite :

$$a_{ij} = \sqrt{a_i a_j}\left[1 - k_{ij} + (k_{ij} - k_{ji})y_i\right] \qquad \text{I.20}$$

Avec y_i est la fraction molaire des composés de mélange. On peut constater que cette équation n'est pas valable pour les mélanges avec plus de deux composés. Cette règle a trois caractéristiques :

- si $k_{ij} = k_{ji}$ on trouve la règle de mélange de (VdW).
- Le paramètre d'interaction entre les deux composés *i* et *j*, approche k_{ij} comme y_i, et la fraction molaire y_i approche à zero.
- Appliquons cette règle pour calculer le paramètre a_m on trouve une expression cubique pour a_m.

La justification théorique de ces règles de mélange provient de l'équation d'état du Viriel donnée par l'équation (I-21) En effet, la mécanique statistique montre que le second *B* et le troisième coefficient *C''* du Viriel ne dépend que de la température :

$$Z = \frac{Pv}{RT} = 1 + \frac{B''}{v} + \frac{C''}{v^2} + \cdots \qquad \text{I.21}$$

Où les coefficients *B'', C''*, … sont appelés respectivement second, troisième, … coefficient du Viriel et sont associés respectivement aux interactions deux à deux, trois à trois, …. Pour un fluide pur, les coefficients du Viriel ne dépendent que de la température. Dans le cas des mélanges ils s'écrivent en fonction de la composition du mélange sous la forme :

$$B'' = \sum\sum x_i x_j \, B_{ij}(T) \qquad \text{I.22}$$

$$C'' = \sum\sum\sum x_i \, x_j x_k C_{ijk}(T) \qquad \text{I.23}$$

D'autre part, en faisant un développement limité, par exemple de l'équation de Van der Waals en puissance de *(b/v)*, on peut la réécrire sous la forme :

$$\frac{Pv}{RT} = \frac{1}{1 + b/v} - \frac{a}{vRT} = 1 + \sum_{n=0}^{\infty} (b/v)^n - \frac{a}{vRT} \qquad \text{I.24}$$

Ainsi, une simple identification avec l'équation du Viriel, donne, (Hadj-Kali, 2004):

$$B\left(x_i,T\right)=\sum\sum x_i x_j B_{ij}(T)=b-\frac{a}{RT} \qquad \text{I.25}$$

Notons que ces règles de mélange ne sont applicables qu'aux mélanges présentant une déviation relativement faible par rapport à l'idéalité ; car comme nous venons de le voir le développement limité n'est valable qu'à faibles densités. Cependant, jusqu'à présent, ce sont ces règles de mélange qui sont les plus utilisées à n'importe quelle densité, du moment où la mécanique statistique ne fournit pas encore suffisamment d'information valide à densités élevées pour pouvoir vérifier la crédibilité des résultats obtenus.

I.6.4. Règle de mélange d'Adachi et Sugi (AS)

Ces deux auteurs (Adachi and Sugie, 1986) ont modifié le paramètre a_{ij} en introduisant un deuxième paramètre d'interaction ajustable, dans l'expression de calcul de paramètre a_m :

$$a_m=\sum_i^{nc}\sum_j^{nc} y_i y_j a_{ij} \qquad \text{I.26}$$

Où a_{ij} est donné selon Adachi et Sugi :

$$a_{ij}=\sqrt{a_i a_j}\left[1-k_{ij}-\lambda_{ij}(y_i-y_j)\right] \qquad \text{I.27}$$

Le deuxième constant de mélange est donné comme suit :

$$b_m=\frac{1}{2}\sum_i^{nc}\sum_j^{nc} y_i y_j b_{ij} \qquad \text{I.28}$$

Avec :

$$b_{ij}=\left(\frac{b_i+b_j}{2}\right)\left(1-l_{ij}\right) \qquad \text{I.29}$$

D'où la règle de mélange d'Adachi et Sugi est une règle à trois paramètres d'interaction binaires ajustables, k_{ij} , λ_{ij} et l_{ij} (AS3).

I.6.5. Règle de mélange avec les grandeurs d'excès

La plupart des solutions à intérêt industriel sont polaires ou présentent une importante déviation par rapport à l'idéalité et sont généralement décrites par les modèles de coefficient

d'activité. Vidal en 1978 et plus tard Huron et Vidal en 1979 ont introduit une nouvelle génération de règles de mélange combinant une équation d'état avec un modèle de coefficient d'activité. Cette combinaison est réalisée par le biais de l'expression de la fonction de l'enthalpie libre d'excès calculée à partir d'une équation d'état pour développer une règle de mélange qui reproduit le même résultat, (Hadj-Kali, 2004):

$$G^E = RT\left(\ln\phi - \sum_i x_i \ln\phi_i\right) \qquad \text{I.30}$$

Wong et Sandler (Wong and Sandler, 1992) ont proposé une règle de mélange pour l'équation d'état cubique qui peut être écrite :

$$b_m = \sum\sum x_i x_j (b - a/RT)_{ij}[1 - \sum x_i a_i / b_i RT - A_\infty^E(x)/\Omega RT] \qquad \text{I.31}$$

D'autres chercheurs (Kurihara *et al.*, 1987) ont proposés une autre règle de mélange (KTK) basée sur la notion de séparation de l'énergie libre d'excès de Gibbs :

$$g^E = g_{RS}^E + g_{RES}^E \qquad \text{I.32}$$

Avec g_{RS}^E est l'énergie libre de Gibbs pour la solution et g_{RES}^E est l'énergie d'excès résiduelle. Kurihara et al. Ont dérivé une expression pour la contribution de solution, régulière en utilisant une équation d'état la règle de mélange est donnée :

$$a_m = \sum_i \sum_j x_i x_j (a_{ii} a_{jj})^{0.5} - (\tau - \phi) g_{RES}^E / \ln[(b-\phi)/(b-\tau)] \qquad \text{I.33}$$

$$g_{RES}^E = RT x_1 x_2 [\eta_1 + \eta_2 (x_1 - x_2) + \eta 3 (x_1 - x_2)^2] \qquad \text{I.34}$$

Il faut signaler que ce type des règles de mélanges a été développé pour calculer les équilibres liquide-vapeur des mélanges à haute pression.

D'autres approches plus empiriques ont simplement rajouté au paramètre *(a)* de la règle classique de Van der Waal une variation par rapport à la composition en gardant le deuxième paramètre *(b)* inchangé, il s'agit par exemples des règles de mélange de Schwartzentruber et Renon :

$$K_{ij} = K_{ij} + l_{ij} \times \frac{m_{ij} x_{ij} - m_{ji} x_j}{m_{ij} x_{ij} + m_{ji} x_j}\left(x_i + x_j\right) \qquad \text{I.35}$$

Avec : $K_{ij} = K_{ji}$; $l_{ij} = -l_{ji}$; $m_{ij} = l_{ij} - m_{ij}$; $K_{ij} = l_{ij} = 0$.

Les règles de mélanges suscitées sont celles les plus utilisées dans les différentes étapes de calcul de propriétés des mélanges, ils existent d'autres règles mais elles restent inférieur à celles mentionnées ci-dessus de point de vue précision de la prédiction (stryjek-Vera, Mathias-Klotz-Prausnitz …), ce qu'on peut constater des travaux précédents et ce qui a été publié c'est qu'on ne peut pas dire qu'une équation d'état mieux qu'une autres, ou bien une règle de mélange mieux qu'une autre, tous dépend de système à traiter, d'où lors de prédiction de solubilité des nouvelles substances à intérêt pharmaceutique rien n'est universel et le choix d'une équation d'état ou une règle de mélange ou ainsi une méthode d'estimation des propriétés fait partie de l'étape d'optimisation (minimisation de fonction d'objet).

I.7. La solubilité des solutés dans les fluides supercritiques

La solubilité du soluté solide n'est que la fraction molaire ou la fraction massique de ce solide dans la phase supercritique, et elle peut être calculée à partir de la relation thermodynamique suivant :

$$y_i = \left(\frac{P^{sub}}{P} \right) \qquad \text{I.36}$$

Où y_i est la fraction molaire du solide dans la phase supercritique et c'est la solubilité idéale. La solubilité réelle est définie comme suit :

$$y_i = \left(\frac{P^{sub}}{P} \right) E \qquad \text{I.37}$$

Avec E est le facteur de poiyting est donné par la relation suivante :

$$E = \frac{\phi_i^{sub} \exp\left(\frac{v_i^S}{RT} (P - P^{sub} \right)}{\phi_i^{FS}} \qquad \text{I.38}$$

Ce facteur est introduit pour apporter une correction due à la compression du fluide, car la pression d'équilibre du mélange est en général différente de la pression de saturation qu'aurait le composant solide pur à la même température, (Prausnitz *et al.*, 1986a).

I.8. Les modèles de prédiction de la solubilité

La solubilité est une information très importante et cruciale pour le design des procédés pharmaceutiques, surtout ceux qui comportent des procédures de séparation (extraction, chromatographie, fractionnement...) et de traitement des solides (micronisation, micro encapsulation…). La modélisation des données expérimentales de la solubilité dans la phase supercritique peut aussi apporter une meilleure compréhension de phénomènes de dissolution qui peut être utilisée dans la prédiction de solubilité à des pressions et températures intéressantes après la mesure d'un nombre minimum de données expérimentales qui peut accélérer le développement des procédés des fluides supercritiques.

Lors de la modélisation on devrait maintenir dans l'esprit que la solubilité des systèmes peut être catégorisée en trois groupe, le premier un seul type du soluté dans un fluide supercritique le deuxième un mélange de solutés dans un fluide supercritique et le dernier un soluté dans un mélange de fluides supercritiques.

I.8.1. Les modèles semi empirique

La solubilité des solutés solides dans les phases supercritiques est une propriété thermo physique très importante qui nécessite d'être modélisée comme un premier pas pour le développement des technologies supercritiques. La corrélation de solubilité peut être conventionnellement mise en œuvre en utilisant la densité de fluide supercritique, d'où beaucoup de modèles semi empiriques appelé autrement les modèles de densité existent, on cite à titre d'exemple :

I.8.1.1. Modèle de Chrastile

Chrastile (1982) a proposé un modèle pour la modélisation de solubilité des substances solides dans les phases supercritiques, ce dernier il relie la solubilité de soluté à la densité de fluide supercritique avec la supposition que l'association de soluté *(S)* avec N molécules de fluide (C) donne un complexe solvaté (SCN) qui est en équilibre avec le gaz. L'expression du modèle peut être donnée, (Chrastil, 1982):

$$S = \rho^{A_2} \exp(\frac{A_1}{T} + A_0) \qquad \text{I.39}$$

Où, A_2 est le nombre d'association qui représente le nombre moyen des molécules de CO_2 dans le complexe solvate, le paramètre a dépend de température de vaporisation et les enthalpies de solvatation de soluté *($A_1 = \frac{\Delta H}{T}$)*, tandis que A_0 dépend de poids moléculaire de soluté (Ms) et ceux du fluide supercritique (Mscf), la transformation logarithmique de l'équation donne :

$$LnS = A_0 + \frac{A_1}{T} + A_2 Ln\rho \qquad \text{I.40}$$

Où les paramètres A_0, A_1 et A_2 sont des constants qui peuvent être déterminés à partir des données expérimentales, (Chrastil, 1982).

I.8.1.2. Modèle de del Valle et Aguilera

del Valle et Aquilera (1988) ont trouvé que la variation de *I*, défini comme suit $I = A_0 + \frac{A_1}{T}$, en fonction de *(1/T)* est linéaire et la pente de cette droite est équivalente à A_1 de l'équation de Chrastile (I.40), et pour récompenser le changement de l'enthalpie en fonction de température ils ont proposé la modification de l'équation de Chrastile :

$$S = \rho^{A_2} \exp(\frac{A_3}{T^2} + \frac{A_1}{T} + A_0) \qquad \text{I.41}$$

I.8.1.3. Modèle de Bartle

Un autre modèle empirique proposé, (Bartle *et al.*, 1991) qui décrive la relation entre la solubilité et la densité, notant qu'il y a une relation peut être existe entre le facteur de poiyting *(E)* et la densité de solvant :

$$\ln E = K_0 + K_1 \rho \qquad \text{I.42}$$

Où E définie comme $E = y_2 P / P_2^S$ avec P_2^S est la pression de sublimation de soluté, les paramètres K_0, K_1 peuvent être obtenus par la régression des données de solubilité expérimentales. Cette équation permet la corrélation des solubilités des solides à chaque isotherme. Cependant, cette application est strictement limitée aux systèmes où la pression de sublimation est disponible. Pour cette raison Bartle *et al*. Ont réécris cette expression en introduisant une pression de référence P_{ref} = 1bar, et une densité de référence $\rho_{ref} = 700 kg/m^3$, connaissant que $P_2^S = A + \frac{B}{T}$ l'équation (I.42) devient :

$$\ln \frac{y_2 P}{P_{ref}} = C_0 + C_1(\rho - \rho_{ref}) + \frac{C_2}{T} \qquad \text{I.43}$$

Les paramètres C_1, C_2 peuvent être obtenus par la régression des données expérimentales de y_2. A différentes valeurs de température et pression le paramètre C_0 peut être utilisé pour estimer L'enthalpie de vaporisation, (Bartle *et al.*, 1991).

I.8.1.4. Modèle de Mendez-Santiago et Teja

C'est le modèle de densité le plus utilisé (Méndez-Santiago and Teja, 1999), cette méthode nécessite la pression de sublimation de solide. Ils ont proposé un modèle basé sur la théorie des solutions diluées dans lequel le coefficient de poiyting est donné en fonction de densité de solvant. L'équation du modèle :

$$T \ln E = T \ln \frac{y_2 P}{P_2^{sub}} = A + B\rho_1 \qquad \text{I.44}$$

Ou *T* est la température absolue, *E* facteur de poiyting, y_2 est la fraction molaire de soluté, *P* est la pression totale du système, P_2^{sub} est la pression de sublimation du soluté, ρ_1 est la densité de fluide supercritique, *A* et *B* sont des constantes de régression. Le facteur *E* représente le rapport entre la solubilité actuelle (réelle) et idéale, cette dernière est calculée selon la loi des gaz parfait (P/P^{sub}), (Prausnitz *et al.*, 1986a). Avantage de ce modèle c'est que les données expérimentales de solubilité pour différentes isothermes peuvent être représentés par une seule droite ou la régression donne des constantes *A* et *B* indépendant de la température. Dans le plus part des cas la pression de sublimation n'est pas disponible, c'est pour cette raison que Mandèz et Teja ont proposé une modification de leur expression pour l'utilisation dans le cas où P^{sub} est inconnue en remplaçant ce dernier par les constantes de l'équation d'Antoine :

$$T \ln y_2 P = A' + B' \rho_1 + C'^T \qquad \text{I.45}$$

Avec A', B' et C' sont des paramètres de régression, (Méndez-Santiago and Teja, 1999).

I.8.1.5. Modèle de Kumar et Johnston

Kumar et Johnton ont observé qu'il y a une relation linéaire entre le logarithme de solubilité de soluté solide et la densité de solvant l'expression de modèle (Kumar and Johnston, 1988) est la suivante :

$$Lny_2 = b_0 + b_1 + \rho \frac{b_2}{T} \qquad \text{I.46}$$

Où les paramètres b_n (n=1,3) sont des paramètres obtenus par la corrélation des données expérimentales. Cependant, le paramètre b_2 relatif à l'enthalpie totale $\Delta H_{total} = -b_2 R$, (Kumar and Johnston, 1988).

I.8.2. Les équations d'états

De par leur origine historique et par la prédominance historique de la pétrochimie en Génie Chimique, les équations d'état jouent un rôle central dans la conception des unités de Génie des Procédés. Les articles foisonnent dans la littérature et les revues sont nombreuses. Sans intention

d'exhaustivité, cette partie présente le développement d'équations d'état qui ont le plus marqué la discipline du Génie des Procédés.

L'utilisation d'une équation d'état pour reproduire les propriétés volumétriques et thermodynamiques des corps purs et des mélanges de corps purs fait l'objet de recherche par les thermodynamiciens depuis plus d'un siècle. En effet, depuis que Van der Waals en 1873 a proposé sa très célèbre équation d'état cubique, un nombre très important de publications a vu le jour. Avec le développement de l'informatique, l'utilisation d'expressions analytiques de plus en plus performantes permet la prédiction des propriétés thermodynamiques avec une grande précision permettant d'améliorer la conception des procédés de Génie Chimique tels que la distillation, l'absorption, …etc. Néanmoins, ce n'est que pour les gaz qu'on a mis ces relations sous une forme assez précise, non seulement en partant des données expérimentales concernant les paramètres *P*, *V*, *T*, mais aussi grâce à des considérations théoriques sur la constitution des atomes et des molécules et des interactions qui restent faibles dans le cas des gaz, (Hadj-Kali, 2004).

En général, une équation d'état est développée initialement pour les substances pures et elle s'écrit sous la forme : $f(P,V,X,T) = 0$. Où X représente le vecteur composition (fraction molaire) ; et $X = 1$ pour un Corps pur.

Toute équation d'état doit vérifier le critère suivant aux coordonnées critiques (c'est-à-dire pour $T = T_c$, $P = P_c$ et $V = V_c$) :

$$\left(\frac{\partial P}{\partial V}\right)_{T=Tc} = \left(\frac{\partial^2 P}{\partial V^2}\right)_{T=Tc} = 0 \qquad \text{I.47}$$

Ceci est dû au fait que l'isotherme critique possède une pente horizontale et un point d'inflexion au point critique.

I.8.2.1. Equation d'état du gaz parfait – loi des états correspondants

La première équation d'état reliant les propriétés macroscopiques d'un système est la loi des gaz parfait découlant de la loi de Boyle-Mariotte, (Hadj-Kali, 2004):

$$P\,v = RT \qquad \text{I.48}$$

Avec v représentant le volume molaire du gaz.

Nous avons aussi vu que la thermodynamique statistique permet d'établir cette loi en séparant les contributions cinétiques et potentielles de l'énergie du système et en faisant la simple hypothèse que l'énergie d'interaction entre les molécules est nulle.

Pour un gaz réel, on peut décrire commodément les écarts par rapport à l'état parfait en introduisant un facteur de compressibilité Z :

$$P v = ZRT \qquad \text{I.49}$$

Van der Waals a constaté que le facteur de compressibilité pour des fluides simples avec des molécules sphériques comme les gaz rares et le méthane ne dépend que des paramètres réduits *Tr* et *Pr* définis de la manière suivante :

$$T_r = \frac{T}{T_c} \qquad \text{et} \qquad P_r = \frac{P}{P_c} \qquad \text{I.50}$$

Où T_c et P_c sont la température critique et la pression critique respectivement.

Alors, toutes les substances ayant les mêmes coordonnées réduites (*Tr, Pr*) auront forcement le même facteur de compressibilité Z : c'est la loi des états correspondants.

Cependant, pour des molécules non sphériques, on a constaté qu'il y a une déviation par rapport à la loi des états correspondants, d'où la nécessité d'introduire un troisième paramètre qui tient compte de la non sphéricité de la molécule. Plusieurs types de paramètres ont été proposés dont le facteur de compressibilité critique Z_c par Meissner et Seferian et le paramètre α_c proposé par Reidel, mais celui qui a connu le plus grand succès est le facteur acentrique de Pitzer ω défini par :

$$\omega = -\left(\log P_r^{sat}\right)_{T_C=0.7} = 1 \quad \text{Avec} \quad P_r^{sat} = \frac{P^{sat}}{P_C} \qquad \text{I.51}$$

Le facteur de compressibilité s'exprime alors par la relation suivante :

$$Z = Z^{(0)}\left(T_r, P_r\right) + \omega Z^{(1)}\left(T_r, P_r\right) \qquad \text{I.52}$$

Le premier terme de cette égalité ($Z^{(0)}$) représente le facteur de compressibilité pour une molécule sphérique, alors que $Z^{(1)}$ exprime la déviation par rapport à la sphéricité idéale de la molécule. Pitzer et ses collaborateurs ont réparti dans un tableau les valeurs de $Z^{(0)}$ et $Z^{(1)}$ en fonction des coordonnées réduites T_r et P_r, alors qu'Edmister propose les mêmes valeurs sous forme de diagrammes, (Prausnitz *et al.*, 1986).

I.8.2.2. Equations d'état cubique

L'équation des gaz parfaits n'est utilisable que dans un domaine de pression restreint (celui de basses pressions), pour cette raison cette équation a été modifiée sous la forme proposée par Hirn, (Hadj-Kali, 2004):

$$(P+\pi)(v-b)=RT \qquad \text{I.53}$$

Où *b* est appelé *covolume* représente la part du volume molaire qui n'est pas réellement disponible à cause de la présence d'autres molécules, et π est un terme correctif de pression dû à l'attraction mutuelle des molécules, appelé pression interne.

La première équation d'état à fournir une description quantitative des deux phases vapeur et liquide et une prédiction de la transition entre phases est celle de Van der Waals en 1873 qui a établi théoriquement que la pression interne est inversement proportionnelle au carré du volume molaire, d'où son équation d'état :

$$\left(P+\frac{a}{v^2}\right)(v-b)=RT \qquad \text{I.54}$$

Ou bien :

$$P=\frac{RT}{(v-b)}-\frac{a}{v^2} \qquad \text{I.55}$$

Dans cette dernière équation, le premier terme associé au covolume b caractérise la répulsion et le second associé à la pression interne l'attraction. Si cette formule ne représente pas parfaitement les isothermes d'un fluide, elle ne conduit pas à des absurdités physiques, d'où sa validité et son application pour des calculs approximatifs.

Pour la détermination des paramètres de cette équation (et ceux des autres équations d'état du même type), il existe deux alternatives :

1. La première consiste à choisir des paramètres qui permettent de reproduire des données expérimentales, généralement la pression de vapeur et la densité liquide ou vapeur. Avec l'équation de Van der Waals dont les coefficients ne dépendent pas de la température, une seule température suffit, pour les autres équations, par contre, où ces paramètres seront fonction de la température, il faudra régresser ces paramètres sur un intervalle de température.

2. La deuxième procédure, qui reste la plus recommandée, utilise la condition du point critique, donnée par l'équation (I-47), pour obtenir les valeurs de ces paramètres aux points critiques a_c et b_c ; et ensuite applique une correction qui dépendra de la température (c'est-à-dire $a\,(T)=a_c \times \alpha\,(T)$), égale à l'unité au point critique et ajusté pour permettre de mieux reproduire tout l'intervalle de température.

Hormis son caractère historique, l'équation de Van der Waals n'est pas quantitativement précise. Par exemple, elle prédit que le facteur de compressibilité critique :

$$Z_C = \frac{P_c v_c}{RT_c} \qquad \text{I.56}$$

Z_C est égal à 3/8 = 0,375 pour tous les fluides ; alors que sa valeur pour différents hydrocarbures varie de 0,24 à 0,29, et cet intervalle est plus grand si d'autres fluides non hydrocarbures sont considérés. De même, les pressions de saturation prédites sont incorrectes. A cause de ces insuffisances, de nombreux chercheurs ont tenté de modifier la forme de l'équation de Van der Waals afin d'améliorer sa précision.

La deuxième alternative, (Hadj-Kali, 2004) est la plus recommandée et utilisée pour la prédiction de solubilité des solutés solides dans les fluides supercritiques.

Après cette petite revue sur les origines des équations d'état et les méthodes d'utilisation on passe maintenant à la description des étapes de calcul de la solubilité des solides à l'aide d'une équation d'état cubique qui a pour but de prédire le coefficient de fugacité du solide dans l'autre phase (supercritique) nécessaire dans ce processus.

I.8.2.2.1. L'équation d'état de Redlich et Kwong (RK)

Dans le but de mieux représenter les propriétés volumétriques des fluides, Redlich et Kwong ont proposé une modification empirique du terme d'attraction de Van der Waals en introduisant également une variation du paramètre ***(a)*** avec la température. Ceci permet d'améliorer le calcul de propriétés de corps purs, (Redlich and Kwong, 1949).

$$P = \frac{RT}{(v-b)} - \frac{\alpha \times a}{v(v+b)} \qquad \text{I.57}$$

(Avec $\alpha = T^{-1/2}$ pour l'équation de Redlich-Kwong)

Exprimée en fonction du facteur acentrique, l'équation de Redlich-Kwong a une forme cubique en posant :

$$A = \frac{aP}{(RT)^2} \qquad \text{et} \qquad B = \frac{bP}{RT}$$

$$Z^3 - Z^2 + \left(A - B - B^2\right)Z - AB = 0 \qquad \text{I.58}$$

La représentation de cette équation sur un diagramme de corps pur illustre bien les trois racines de cette équation sur l'isotherme d'équilibre liquide–vapeur : deux racines correspondent aux points de bulle et de rosée, tandis que la troisième située dans la région liquide – vapeur n'a pas de signification physique, (Hadj-Kali, 2004).

I.8.2.2.2. L'équation d'état de Soave Redlich et Kwong (SRK)

Il a été remarqué que les équations d'état cubique à deux paramètres ne peuvent représenter avec une précision satisfaisante à la fois le comportement volumétrique et les tensions de vapeur, et qu'il convient de choisir une priorité. C'est ce qu'a compris Soave en 1972 qui a appliqué l'équation d'état de Redlich-Kwong au calcul des tensions de vapeur des hydrocarbures en modifiant l'expression du paramètre α devenu fonction de la température et du facteur acentrique de Pitzer ω, (Soave, 1972) :

$$P = \frac{RT}{(v-b)} - \frac{a(T)}{v(v+b)} \qquad \text{I.59}$$

Où les constants des corps pures *a(T)* et *b* pour cette équation, (Soave, 1972) :

$$a(T) = 0.42747\,\frac{R^2 T_c^{\,2}}{P_c}\,\alpha(T_r, w) \qquad \text{I.60}$$

$$b = 0.08664\,\frac{RT_c}{P_c} \qquad \text{I.61}$$

Avec :

$$\alpha(Tr, w) = \left[1 + \left(0.480 + 1.574\omega - 0.176\omega^2\right) \times \left(1 - T_r^{0.5}\right)\right]^2 \qquad \text{I.62}$$

Et sous forme cubique :

$$Z^3 - Z^2 + \left(A - B - B^2\right)Z - AB = 0 \qquad \text{I.63}$$

L'expression du coefficient de fugacité est donnée en logarithme :

$$\ln\phi_i = \frac{b_i}{b_m}(Z-1) - \ln(Z-B) - \frac{A}{B}\left(2\frac{\sqrt{a_i}}{\sqrt{a_m}} - \frac{b_i}{b_m}\right)\ln\left(1+\frac{B}{Z}\right) \qquad \text{I.64}$$

L'équation de Soave-Redlich-Kwong conserve la même valeur du facteur de compressibilité critique (Zc = 0,333) qui reste donc trop élevée par rapport aux valeurs expérimentales constatées.

I.8.2.2.3. L'équation d'état de Peng et Robinson (PR)

En 1976, Peng et Robinson apportent en plus de la modification du paramètre α, une nouvelle expression pour le terme d'attraction. Le progrès est sensible pour la tension de vapeur et surtout pour le calcul des masses volumiques en phase liquide. En outre, le facteur de compressibilité critique est mieux estimé (Zc = 0,307), (Peng and Robinson, 1976).

$$P = \frac{RT}{(v-b)} - \frac{\alpha \times a}{v(v+b)+b(v-b)} \qquad \text{I.65}$$

Avec :

$$\alpha = \left[1 + \left(0.37464 + 1.5422\omega - 0.26992\omega^2\right) \times \left(1 - T_r^{0.5}\right)\right]^2 \qquad \text{I.66}$$

Et en termes de facteur de compressibilité :

$$Z^3 - (1-B)Z^2 + \left(A - 2B - 3B^2\right)Z - \left(AB - B^2 - B^3\right) = 0 \qquad \text{I.67}$$

Où les constants des corps purs a et b pour cette équation :

$$a = 0.45724\frac{R^2T_c^2}{P_c}\alpha(T_r, w) \qquad \text{I.68}$$

$$b = 0.07780\frac{RT_c}{P_c} \qquad \text{I.69}$$

Pour l'équation de PR, l'expression de coefficient de fugacité est donnée par :

$$\ln\phi_i = \frac{b_i}{b_m}(Z-1) - \ln(Z-B) - \frac{A}{2\sqrt{2}B}\left(\frac{2\sum_{j=1}^{n} y_j a_{ij}}{a_m} - \frac{b_i}{b_m}\right)\ln\left(\frac{Z+(1+\sqrt{2})B}{Z+(1-\sqrt{2})B}\right) \quad \text{I.70}$$

Les équations de Peng-Robinson (PR) et de Soave-Redlich-Kwong (SRK) sont largement utilisées dans l'industrie, particulièrement pour le raffinage et la simulation des réservoirs. Leurs avantages résident dans le fait qu'elles suivent un processus de calcul successive pour le calcul de coefficient de fugacité mais pour ce faire elles nécessitent les coordonnées critiques et le facteur acentrique qui sont généralement indisponible pour les nouvelles substances, un temps de simulation relativement court et surtout qu'elles conduisent à une bonne estimation, (Valderrama, 2003).

I.8.2.2.4. L'équation d'état de Patel-Teja-Valderrama (PTV)

Abbott (1979), dans une étude comparative des équations d'état cubique, a conclu que la forme la plus adéquate doit faire intervenir des paramètres supplémentaires. De nouvelles équations ont été développées, toutes en convergeant vers les équations SRK ou PR lorsque des valeurs particulières sont attribuées à ces paramètres additionnels.

L'une de ces équations qui contiennent des paramètres autres que les deux paramètres *a* et *b* est l'équation de Patel-Teja-Valderrama, (Valderrama, 1990) et son expression est donnée ci-dessous :

$$P = \frac{RT}{(v-b)} - \frac{a(T)}{v(v+b)+c(v-b)} \quad \text{I.71}$$

Où les constantes des corps purs *a*, *b* et *c* pour cette équation :

$$a = \Omega_a(Z_c)\frac{R^2 T_c^2}{P_c}\alpha(T_r, w, Z_c) \quad \text{I.72}$$

Avec:

$$\Omega_a(Z_c) = 0.66121 - 0.761057.Z_c \quad \text{I.73}$$

$$b = \Omega_b(Z_c)\frac{RT_c}{P_c} \quad \text{I.74}$$

Ou:

$$\Omega_b(Z_c) = 0.02207 + .0.20868.Z_c \quad \text{I.75}$$

$$c = \Omega_c(Z_c)\frac{RT_c}{P_c} \quad \text{I.76}$$

Avec:

$$\Omega_b(Z_c) = 0.57765 - 1.87080.Z_c \quad \text{I.77}$$

L'expression de la fonction dépendante de la température $\alpha(T_r, w, Z_c)$ pour cette équation, (Valderrama, 1990) :

$$\alpha(T_r, w, Z_c) = \left[1 + F \times \left(1 - T_r^{0.5}\right)\right]^2 \quad \text{I.78}$$

Et

$$F = \left(0.46283 + 3.58230\omega Z_c - 8.19417(\omega Z_c)^2\right) \quad \text{I.79}$$

I.8.2.2.5. L'équation de Peng et Robinson modifiée (MPR)

La fonction alpha *(α)* de l'équation d'état de Peng et Robinson a été modifiée (Gasem *et al.*, 2001) pour donner l'expression suivante :

$$\alpha(Tr) = \exp\left[(2 + 0.836Tr) \times \left(1 - T_r^m\right)\right]^2 \quad \text{I.80}$$

Et

$$m = \left(0.134 + 0.508\omega - 0.0467\omega^2\right) \quad \text{I.81}$$

Tous les autres paramètres utilisés sont ceux de l'équation PR (*a*, *b*…).

Les équations citées sont les équations d'état cubiques les plus utilisées dans la prédiction de la solubilité des solutés solides dans les phases supercritiques et d'autres paramètres, il existe d'autres équations mais elles ne sont pas assez robustes pour donner des résultats comparables à celles données par les équations suscitées.

- **Les modifications de $\alpha(Tr)$**

Les axes de recherche**s** sur les équations d'état cubiques après les contributions de Peng et Robinson et celles de Soave suivis trois voies principales, (Valderrama, 2003):

(i) Modification de fonction $\alpha(Tr)$ dans les équation**s** PR et SRK.

(ii) Modification du volume dépend de terme attractif.

(iii) L'utilisation de troisième paramètre relatif au corps purs *(c).*

Ces modifications de la fonction dépendante $\alpha(Tr)$ dans le terme attractif des deux équations (SRK et PR) ont été principalement proposées pour améliorer les capacités de prédiction des différentes propriétés des corps pures et des mélanges.

I.8.2.3. Equations d'état non cubique (basée sur la théorie de perturbation)

Les équations de type SAFT (Statistical Associated Fluid Théory) ou, basées sur les théories de perturbation connaissent un développement important depuis 15 ans (Müller and Gubbins, 2001). Elles évaluent l'énergie libre de *Helmoltz* comme la somme de plusieurs contributions formalisant une représentation des molécules et de leur interaction : contribution de monomère, contribution de liaison de ces monomères pour former une molécule ; contribution d'association intermoléculaire entre sites d'interaction. Ce formalisme est proche de celui imaginé en mécanique moléculaire et l'un des atouts de l'approche (SAFT) est le fait que les prédictions d'une telle équation peuvent être confrontées à des simulations moléculaires utilisant le même formalisme, (Hadj-Kali, 2004). Cette gamme contient des équations d'état combinées avec la théorie associative telle que l'équation d'état de CPA (Cubic Plus Association) et celle de Pasuki et des équations d'état complexes et non cubiques à fondement théorique basé sur thermodynamique statistique telles que les équations d'état SAFT (Statistical Association Fluid Théory), PC-SAFT (Perturbed Chain Statistical Association Fluid Théory), VR-SAFT. Dans cette partie on va donner une revue sur ces équations, ses origines et ses bases théoriques.

L'équation d'état SAFT (Chapman *et al.*, 1989) est considérée comme une méthode qui combine entre la théorie de perturbation thermodynamique de Wertheim, (Wertheim, 1984) pour l'association des fluides avec l'idée de formuler une équation d'état. Dès le début, SAFT n'est pas

une équation d'état rigide mais plutôt une méthode qui permet l'incorporation l'effet de l'association dans une théorie donnée.

L'équation d'état d'un fluide associatif est une perturbation (Müller and Gubbins, 2001) donnée en terme de l'énergie molaire de Helmholtz résiduelle définie comme la différence entre l'énergie molaire de Helmholtz totale et celle d'un gaz idéal à température et densité constantes :

$$a^r(T,\rho) = a(T,\rho) - a^{gazidéal}(T,\rho) \qquad \text{I.82}$$

SAFT suppose qu'il y a trois types de contributions au potentiel intermoléculaire :

La contribution des segments individuels, la contribution des chaînes qui peuvent se former à partir des segments, la contribution due à la possibilité que certains segments forment des sites d'associations avec d'autres molécules. La figure I.2 montre les étapes de formations supposées par SAFT.

L'énergie molaire de Helmholtz résiduelle est définie comme la somme de ces contributions :

$$a^r = a^{seg} + a^{chain} + a^{assoc} \qquad \text{I.82a}$$

I.8.2.3.1. Contribution des segments

Dans cette théorie les constructions de base pour les molécules sont des segments sphériques durs qui interagissent sous l'effet de deux forces attractive et répulsive (dispersion), et la contribution des segments relative aux interactions aléatoires et non directionnelles. Ces segments peuvent être des atomes, des groupes fonctionnels, une répétition d'une unité dans un polymère ou bien molécule entière.

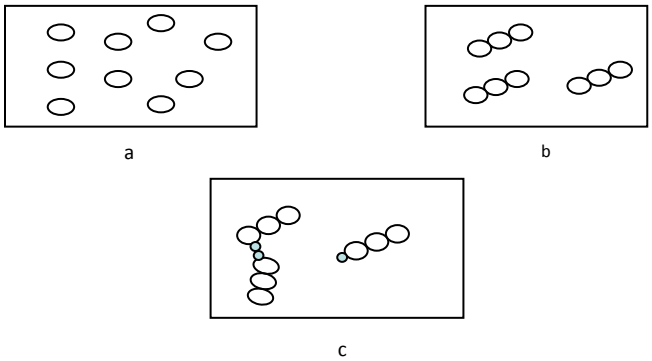


Figure I.2. Les étapes de formation de modèle moléculaires à partir de théorie associative SAFT, (Müller and Gubbins, 2001).

I.8.2.3.2. Contribution des chaînes

Cette contribution due à la possibilité que certains segments qui contiennent des sites d'associations pour former des chaînes linéaires, par exemple un polymère qui se constitue d'une seule unité qui se répète n fois peut être représenté comme une chaîne de n segments dures (Hard sphère).

I.8.2.3.3. Contribution d'association

Dans une molécule il n'y a pas une limite des sites d'association ces derniers peuvent être des liaisons d'hydrogène, des liaisons électrostatiques…. Par exemple dans un mélange binaire où il y a un contact entre deux composés, selon SAFT il y aura un contact entre deux chaînes des segments sphérique dures différentes qui peut générer des sites d'association pour former un complexe de mélange, (Müller and Gubbins, 2001).

L'équation d'état SAFT est donnée en termes de facteur de compressibilité Z :

$$Z = 1 + \rho\left[\frac{\partial a}{\partial \rho}\right]_{T,Y} \qquad \text{I.82b}$$

Avec ρ la densité et a l'énergie molaire de Helmholtz (J/mole), il faut signaler que la contribution de l'équation d'état dans la prédiction de solubilité des solides dans les phases supercritiques est le calcul du coefficient de fugacité de soluté dans l'autre phase, le processus de calcul pour cette dernière est à la fois successif et itératif et demande un calcul énorme.

I.8.2.3.4. Equation d'état de PAZUKI

Cette nouvelle équation basée sur la théorie de perturbation (Dashtizadeh *et al.*, 2006) possède deux termes le terme répulsif et attractif selon la thermodynamique statistique (MacQuarrie, 1976), elle peut être écrite en termes de facteur de compressibilité :

$$Z = Z^{hs} + Z^{pert} \qquad \text{I.83}$$

La forme cubique de cette équation est la suivante :

$$Z^3 - Z^2 + (A - B^2 - 2B)Z - (AB + B^2) = 0 \qquad \text{I.84}$$

Avec les fonctions *alpha* et *bêta* pour le calcul des constantes des mélanges :

$$\alpha(Tr,w) = [1 + m_1(1-Tr) + m_2(1-Tr)^2 + m_3(1-Tr)^3]^2 \quad \text{I.85}$$

$$\beta(Tr,w) = [1 + n(1-Tr)] \quad \text{I.86}$$

I.8.2.3.5. Equation d'état PC-SAFT

L'équation d'état PC-SAFT (Gross and Sadowski, 2000) a été développée à partir de la théorie de perturbation comme celles citées auparavant où elle considère les molécules comme des chaînes linéaires composées des segments sphérique dures (hard Chain) qui peuvent générer des sites d'associations, d'où le principe est le même que l'équation d'état SAFT. L'expression générale de cette équation est donnée par la relation suivante :

$$\frac{A}{NkT} = \frac{A^{hc}}{NkT} + \frac{A^{pert}}{NkT} \quad \text{I.87}$$

Où *A* est l'énergie libre de Helmholtz, *N* est le nombre total des molécules, *k* est la constante de Boltzmann, *T* est la température absolue. Dans cette terminologie le premier terme (hc) est relatif aux chaînes des segments durs dus aux interactions répulsives et la partie de perturbation signifie la contribution due aux interactions attractives des molécules. L'expression de cette équation peut être écrite en termes de facteur de compressibilité :

$$Z = Z^{hc} + Z^{pert} \quad \text{I.88}$$

Par exemple dans un mélange de deux substances c'est-à-dire selon PC-SAFT un contact de deux chaînes linéaires composées des sphères où les interactions intermoléculaires sont régies par trois types d'interactions : les interactions dispersives, celles attractives entre les chaînes et des forces d'associations entre les sites sur les deux chaînes.

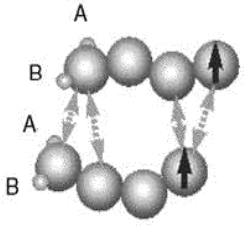


Figure I.3. Le modèle moléculaire dans un mélange de deux substances selon l'équation d'état PC-SAFT, (Tumakaka *et al.*, 2005).

I.9. Intelligence artificielle

I.9.1. Définition

L'intelligence artificielle (IA) est un terme général qui implique l'utilisation d'un ordinateur pour modéliser le comportement intelligent avec une intervention humaine minimale. L'AI est généralement accepté comme ayant commencé avec l'invention des robots. Le terme dérive du mot tchèque robota, c'est-à-dire des machines biosynthétiques utilisées comme travail forcé, (Bellazzi and Abu-Hanna, 2009).

I.9.2. Réseau de neurone artificiel

Les réseaux de neurones artificiels (RNAs) sont des modèles mathématiques non linéaires conçus dans la deuxième moitié du XXe siècle par la simulation des procédures du cerveau humain et ont été utilisés dans de nombreuses disciplines scientifiques jusqu'ici (Xu *et al.,* 2007). Ces réseaux sont capables de relier les entrées et les sorties de la plupart des systèmes non linéaires multi-variables avec n'importe quelle complexité. Les RNAs peuvent être appliquées à la reconnaissance de formes, au traitement des données et sont également considérées comme des outils puissants pour l'approximation fonctionnelle. Ces réseaux se composent d'un certain nombre d'unités de traitement simples qui sont connectées ensemble d'une manière spécifiée selon le type du réseau. Ces unités de traitement ont été inspirées des neurones biologiques et sont appelées neurones artificiels. Les signaux d'entrée sont toujours transmis à la couche d'entrée, puis transférés aux neurones dans les couches cachées et à la sortie, respectivement.

I.9.3. Algorithmes génétiques

Les algorithmes génétiques (AGs) sont des algorithmes d'optimisation numérique inspirés par la sélection naturelle et la génétique naturelle. La méthode est en générale capable d'être appliquée à une gamme extrêmement étendue de problèmes. Au contraire quelque approches leurs promesses ont rarement été survendues et elles sont utilisées pour aider à résoudre des problèmes pratiques sur une base quotidienne. Les algorithmes sont simples à comprendre et le code informatique nécessaire facile à écrire. Bien qu'il y a un nombre croissant de disciples des AGs, la technique n'a jamais attirée l'attention que par exemple, les réseaux de neurones artificiels, (Coley, 1999).

I.9.4. Les systèmes experts

Au cours de la dernière décennie, l'intérêt pour les résultats de la recherche sur l'intelligence artificielle a augmenté de plus en plus. En particulier, le domaine des systèmes basés sur la connaissance, l'un des premiers domaines de l'intelligence artificielle à être commercialement fructueux, a reçu beaucoup d'attention. L'expression système basé sur la connaissance est généralement employée pour indiquer des systèmes d'information dans lesquels une représentation symbolique de la connaissance humaine est appliquée, généralement d'une manière ressemblant au raisonnement humain, (Lucas and van der Gaag, 1991). Parmi ces systèmes basés sur la connaissance, les systèmes experts ont été les plus réussis au temps actuel. Les systèmes experts sont des systèmes qui sont capables de proposer des solutions à des problèmes spécifiques dans un domaine donné ou qui sont capables de donner des conseils, à la fois d'une manière et à un niveau comparable à celui des experts dans le domaine. Construire des systèmes experts pour des domaines d'application spécifiques est même devenu un sujet distinct appelé ingénierie des connaissances, (Lucas and van der Gaag, 1991).

I.9.5. La Logique floue

La Logique floue a été initiée en 1965 (Zadeh, 1965) par Lotfi A. Zadeh, professeur d'informatique à l'Université de Californie à Berkeley. Fondamentalement, la logique floue (LF) est une logique à valeurs multiples qui permet de définir des valeurs intermédiaires entre des évaluations classiques comme vrai / faux, oui / non, Haut / bas, etc. Des notions comme plutôt grandes ou très rapides peuvent être formulées mathématiquement et traitées par des ordinateurs, afin d'appliquer une manière plus humaine de penser dans la programmation des ordinateurs (Lejewski, 1967). Les systèmes flous ont une alternative aux notions traditionnelles d'appartenance à un ensemble et de logique qui a ses origines dans la philosophie grecque ancienne. La logique floue est devenue un outil rentable pour le contrôle et la direction de systèmes et les procédés industriels complexes, l'électronique domestique et de divertissement, ainsi que pour d'autres systèmes et applications experts.

I.10. Conclusion

La connaissance de la solubilité des solutés solides dans les phases supercritiques est importante pour le design des procédés de séparation. De plus la disponibilité des résultats expérimentaux de solubilité des nouveaux composés intéressants de point de vue pharmaceutique nous offre l'opportunité de modéliser et voir les capacités des modèles existants à prédire cette propriété dans des conditions opératoires bien déterminées.

Les modèles de détermination de la solubilité allant de nature semi empirique jusqu'aux modèles purement théoriques notamment les équations d'état cubiques, non cubiques et celles basées sur la théorie de perturbation qui sont malgré la complexité de processus de calcul et les déviations qui peuvent être enregistrées restent les meilleurs modèles de fait qu'elles sont construit sur une base théorique où la majorité des paramètres à optimiser ont un sens physique. Ce qui concerne les mélanges l'utilisation de ces modèles est assurée par les règles des mélanges.

CHAPITRE 02 :
Le réseau de neurones artificiels.

II.1. Introduction

Comment l'homme fait-il pour raisonner, parler, calculer, apprendre… ? Comme le cerveau humain est un dispositif de stockage et de traitement de données, une question se pose, est de savoir si ce cerveau peut faire des choses qu'un ordinateur ne peut pas et ne pourra jamais le faire.

Il y a des centaines d'exemples qui montrent à la fois combien on peut espérer de la modélisation du système nerveux, mais aussi combien il sera difficile d'imaginer et de comprendre les divers aspects des problèmes de perception. Il parait donc naturel d'essayer de comprendre comment les systèmes biologiques sont capables de telles performances, et si possible, de s'inspirer de leurs principes pour imaginer de nouveaux algorithmes ou de nouvelles machines plus efficaces que ceux dont nous disposons actuellement.

On considère qu'un cerveau est un ensemble, très complexes, de neurones interconnectés et que chaque neurone a un comportement simple et modélisable.

Sous le terme de réseaux de neurones artificiels on regroupe un certain nombre de modèles dont l'intention est d'imiter quelques fonctions du cerveau humain en reproduisant certaines de ses structures de base.

L'approche d'ingénierie pour l'informatique neuronale utilise des modélisations issues de la recherche biologique afin de concevoir des architectures informatiques pouvant accomplir des taches spécifiques telles que l'apprentissage (connaissances, comportement), la reconnaissance des formes, on peut les utilisées dans l'industrie, télécommunication et informatique, finance, pour commander un robot, capter une image, la numériser, détecter un objet mobile, capter le son d'une voix au milieu du bruit ambiant : voici quelques exemples qu'on pourra rencontrer dans la science de l'informatique.

La propriété fondamentale des réseaux de neurones multicouches est l'approximation parcimonieuse qui peut s'énoncer de la façon suivante : Toute fonction bornée suffisamment régulière peut être approchée uniformément, avec une précision arbitraire, dans un domaine fini de l'espace de ses variables, par un réseau de neurones comportant une couche de neurones cachées en nombre fini, possédant tous la même fonction d'activation, et un neurone de sortie linéaire, (Jodoin, 1994).

II.2. Historique

Les premiers à proposer un modèle sont deux américains de Chicago, le neurophysiologiste Warren Mccullers et le logicien Walter Pitt, qui inventent en 1943 le premier neurone formel qui portera leurs noms (neurone de Mccullers- pitt).

Quelques années plus tard en 1949, Heb propose une formulation du mécanisme d'apprentissage, sous la forme d'une règle de modification des connexions synaptiques. Cette règle repose sur le fait que si deux neurones sont connectés entre eux et activées au même temps, les connexions qui les relient doivent être renforcées, dans le cas contraire, elles ne sont pas modifiées.

Le premier réseau de neurones artificiels apparaît en 1958, grâce aux travaux de Frank Rosenblatt qui conçoit le fameux perceptron. Le perceptron possède une couche de neurones d'entrée, ainsi qu'une couche de sortie. Ce réseau parvient à apprendre à identifier des formes simples et à calculer certaines fonctions logiques, à la fin des années 60 (1969) Rosenblatt à reçut une critique violente par Marvin Minsky et Seymour Papert, qui publient un livre dans lequel ils décrivent une série de problèmes de reconnaissance des forme qu'un perceptron ne peut pas résoudre, problème dit d'un apprentissage difficile. Les travaux sur la technologie des réseaux neuronaux sont alors quasiment interrompus, il faudra alors attendre le début des années 80 (1982) pour que Hopfield démontre l'intérêt d'utiliser des réseaux récurrents ('feed-back').

En parallèle des travaux de Hopfield, Werbos conçoit son algorithme de rétropropagation de l'erreur, qui offre un mécanisme d'apprentissage pour les réseaux multicouches de type perceptron (MLP), fournissant ainsi un moyen d'entraîner les neurones des couches cachées. Cet algorithme de 'back- propagation' ne sera pourtant popularisé qu'en 1986 par Rumelhart.

II.3. Fondements Biologiques

Il est estimé que le cerveau humain contient 10^{11} neurones et 10^{14} synapses dans le système nerveux. Les neurones et les synapses constituent les éléments de base pour le traitement de l'information. On pense que le système nerveux compte plus de 1000 milliards de neurones interconnectés. Bien que les neurones ne soient pas tous identiques, leur forme et certaines caractéristiques permettent de les répartir en quelques grandes classes.

En effet, il est aussi important de savoir, que les neurones n'ont pas tous un comportement similaire en fonction de leur position dans le cerveau.

Le schéma d'un neurone biologique est représenté dans la figure II.1:

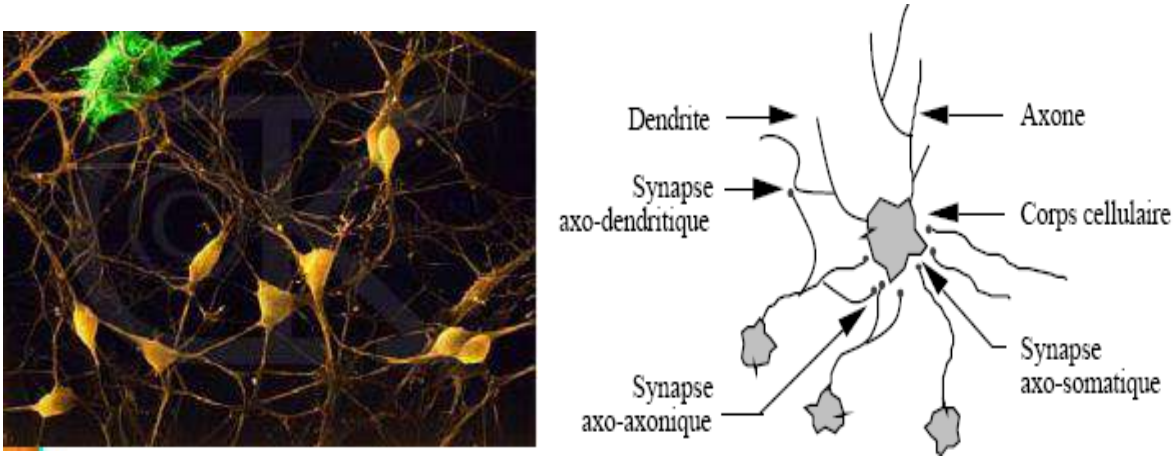


Figure II.1. Neurone biologique, (Gallinari and Gascuel, 1996; Mhaskar and Micchelli, 1994).

Le neurone est une cellule constituée principalement de trois parties qui, au regard du transfert d'information, ont un rôle fonctionnel bien défini. Ce sont :

Le corps cellulaire, les dendrites, et l'axone.

1. **Corps cellulaire :** est le centre de traitement des informations et de commande du neurone, les messages reçus à partir des dendrites sont interprétés et la réponse est envoyée à travers l'axone.
2. **Dendrites :** ce sont de fines extensions tubulaires qui se ramifient autour du neurone et forment une sorte de vaste arborescence. Les signaux envoyés au neurone sont captés par les dendrites.
3. **Axone :** c'est le long de l'axone que les signaux partent du neurone.

Contrairement aux dendrites qui se ramifient autour du neurone, l'axone est plus long et se ramifie à son extrémité où il se connecte aux dendrites des autres neurones, son rôle est de transmettre l'activité neuronale générée par le corps cellulaire aux dendrites d'autres cellules aux voisinages.

La synapse est la jonction de l'axone d'un neurone aux dendrites d'un autre neurone.

Dans un neurone l'information ne se transmet que dans un seul sens : des dendrites vers l'axone.

Le neurone va recevoir des informations venant d'autres neurones, grâce à ses dendrites, il va ensuite y avoir sommation, au niveau du corps cellulaire, de toutes ces informations et via un potentiel d'action (signal électrique) le résultat de l'analyse va transiter le long de l'axone jusqu'aux dendrites au moyen des synapses. Une synapse excitatrice contribuera à activer le neurone cible, tandis qu'une synapse inhibitrice tendra à empêcher l'activité de cet axone.

Les synapses possèdent une sorte de "mémoire" qui leur permet d'ajuster leur fonctionnement, dans ce cas les connexions synaptiques vont donc se modifier. Ainsi, la synapse va faciliter ou non le passage des influx nerveux. Cette plasticité est à l'origine des mécanismes d'apprentissage.

II.4. Neurone Formel

Le neurone formel peut être considéré comme une modélisation mathématique qui reprend les principes du fonctionnement du neurone biologique.

C'est un dispositif possédant une seul sortie et un nombre d'entrée quelconque, ce neurone fait pondérée ces entrées, suivies en général d'une non-linéarité appelée fonction d'activation. Cette fonction d'activation réalisée par le neurone formel est un échelon qui vaut 1 si la somme pondérée de ses entrées est supérieure au seuil et 0 sinon.

Donc le neurone peut être représenté comme suit figure II.2 **:**

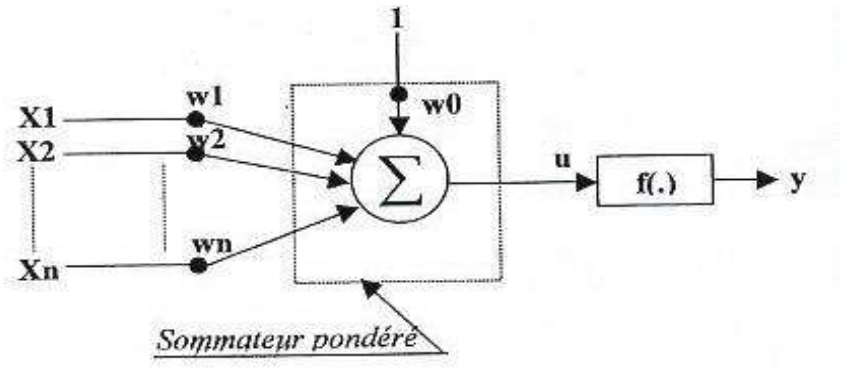


Figure II.2. Le neurone formel, (Gallinari and Gascuel, 1996).

- Les entrées : ($x_1, x_2, x_3 \ldots, x_n$) **;**
- Les poids synaptiques ($w_1, w_2 \ldots, w_n$) ;
- La valeur du seuil interne w_0 ;

L'état du neurone (la somme pondérée) :

$$u(w,x) = \sum_{i=0}^{n} wi \cdot xi \qquad \text{II.1}$$

- La fonction d'activation :

$$y = f(u) \qquad \text{II.2}$$

- La sortie du neurone est : *y.*
- La fonction d'activation :

Les fonctions d'activations les plus utilisées sont les fonctions : Rampe, Sigmoïde, Gaussienne, Heaviside, signe et la fonction à seuil.

Cette fonction est en général monotone, croissante, ainsi l'augmentation de l'entrée ne peut que faire augmenter la sortie ou la garder constante, elle ne peut jamais faire décroître le signal.

Le choix d'une fonction d'activation linéaire facilite les calculs mais le neurone perd sa robustesse. Tandis que, les neurones dont la fonction d'activation est non linéaire augmentent la capacité du réseau à approximer les fonctions complexes.

- La fonction sigmoïde : $f(u) = 1/(1+e^{-B.u})$

B : est le gai de la fonction sigmoïde.

II.5. Analogie avec le modèle biologique

L'idée des neurones artificiels est inspirée par analogie des neurones biologiques, et qui peut être résumée dans le tableau suivant :

Tableau II.1. Correspondance entre les modèles neuronaux biologiques et artificiels, (Gallinari and Gascuel, 1996)

Cerveau	**Réseau de neurones**
Neurone	Elément processeur
Dendrites	Somme pondérée
Corps cellulaire	Fonction d'activation
Axone	Fonction de sortie
Synapse	Poids

II.6. Réseaux de neurones artificiels

Un réseau de neurones artificiels est un ensemble de neurones formels associés en couches (ou sous-groupes) et fonctionnant en parallèle.

Dans un réseau, chaque sous-groupe fait un traitement indépendant des autres et transmet le résultat de son analyse au sous-groupe suivant. L'information donnée au réseau va donc se propager couche par couche, de la couche d'entrée à la couche de sortie, en passant ne soit par aucune, une ou plusieurs couches intermédiaires dites cachées. Il est à noter qu'en fonction de l'algorithme d'apprentissage, il est aussi possible d'avoir une propagation de l'information en reculons ('back propagation').

Habituellement (excepté pour les couches d'entrée et de sortie) chaque neurone dans une couche est connecté à tous les neurones de la couche précédente et de la couche suivante.

Les réseaux de neurones artificiels ont la capacité de stocker de la connaissance empirique et de la rendre disponible à l'usage. Les habiletés de traitement (et donc la connaissance) du réseau vont être stockées dans les poids synaptiques, obtenus par des processus d'adaptation ou d'apprentissage. En ce sens, les réseaux de neurones artificiels ressemblent donc au cerveau car non seulement, la connaissance est acquise au travers d'un apprentissage mais de plus, cette connaissance est stockée dans les connexions, soit dans les poids synaptiques.

II.7. Catégories des réseaux de neurones artificiels

On peut classer les réseaux de neurones en deux grandes catégories :

- Les réseaux de neurones statiques.
- Les réseaux de neurones dynamiques.

II.7.1. Réseaux de neurones multicouches statiques

Ce type de réseau est organisé généralement en couche de neurones. Chaque neurone d'une couche reçoit ses entrées à partir des neurones de la couche précédente ou tout simplement de l'entrée du réseau. Dans ce cas il n'existe pas de "feed-back" (retour) d'information c'est-à-dire l'information circule dans une seule sens, de l'entrée vers la sortie.

Ces réseaux peuvent être utilisés pour les problèmes de classification ou d'approximation de fonctions non linéaires complexes.

Les réseaux de neurones statiques les plus connus sont :

- Les réseaux multicouches (MLP : MultiLayer Perceptron).
- Les réseaux modulaires.
- Le modèle de Kohonen.

Dans notre étude on va utiliser les réseaux multicouches.

L'architecture d'un MPL est organisée en couches successives comprend une couche d'entrée, une couche de sortie et une ou plusieurs couches intermédiaires appelées couches cachées car elles ne sont pas vues à l'extérieur. Chaque couche est composée d'un certain nombre de neurone. Chaque neurone dans une couche est connecté à tous les neurones de la couche suivante tel que il n'y a pas de connexion entres les neurones de la même couche.

Un neurone ne peut envoyer son résultat qu'à un neurone situé dans une couche postérieure à la sienne figure II.3**.**

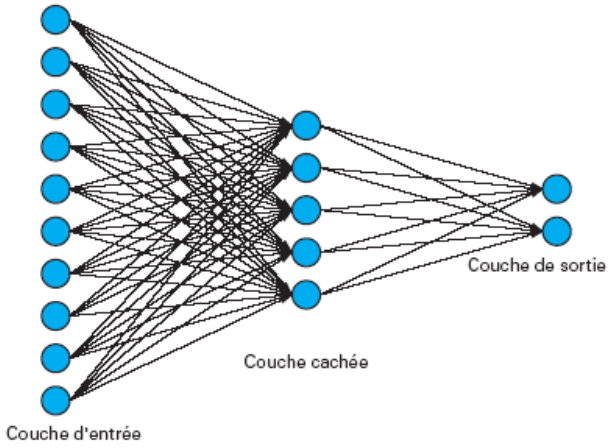


Figure II.3. Réseau multicouche MLP, (Dreyfuss *et al.*, 2002).

II.7.2. Réseaux de neurones dynamiques

Les réseaux dynamiques ou réseaux récurrents sont organisés de telle sorte que chaque neurone reçoit sur ses entrées une partie ou la totalité de l'état du réseau (sortie des autres neurones) en plus des informations externes. Pour ces réseaux l'influence entre les neurones s'exerce dans les deux sens. L'état global du réseau dépend aussi de ses états précédents. L'équation du neurone dans ce cas, est décrite par des équations différentielles ou aux différences. La structure interne de ces réseaux est exploitée pour l'identification et la commande de systèmes dynamiques.

Les réseaux de neurones dynamiques les plus connus sont :

- Les réseaux de neurones avec temps de retard.
- Les réseaux de neurones avec retour de sortie.
- Le réseau de Hopfield.

II.8. Apprentissage des réseaux de neurones

On appelle apprentissage, l'opération par la quel le réseau de neurones acquiert la capacité de réaliser certaines taches en modifiant les paramètres internes (poids synaptique), cette modification se fait en utilisant un algorithme d'adaptation paramétrique appelé "algorithme d'apprentissage".

L'apprentissage consiste alors à améliorer les performances futures du réseau en se basant sur les expériences acquises.

Plusieurs modes d'apprentissage peuvent être distinguées, ceux-ci se différencient principalement par le type d'exemples disponible pour un RNA, l'apprentissage peut être considéré comme le problème de la mise à jour des poids des connexions au sein du réseau, afin de réussir la tache qui lui est demandée. L'apprentissage est la caractéristique principale des RNA et il peut se faire de différentes manières et selon différentes règles.

L'objectif des algorithmes d'apprentissages est de minimiser l'erreur de décision effectuée par le RNA en ajustant les poids à chaque présentation d'un vecteur d'entraînement.

II.8.1. Types d'apprentissages

Les algorithmes d'apprentissages peuvent se répartir en trois classes :

- Apprentissage non supervisé.
- Apprentissage forcé ou imposé.
- Apprentissage supervisé.

II.8.1.1. Apprentissage non supervisé

Cet entraînement se fait sur la base d'informations locales existant à l'entrée. Il représente la capacité des réseaux de neurones à apprendre sana connaître la sortie désirée. Dans ce cas, le réseau organise lui-même les entrées qui lui sont présentées de façon à minimiser un critère de performance donné. Les informations d'entrée sont présentées au réseau et des catégories internes sont formées par une auto-organisation du réseau. Cet apprentissage est dit aussi compétitif du moment où il y a une compétition entre les unités du réseau à représenter le vecteur d'entrée figure II.4.

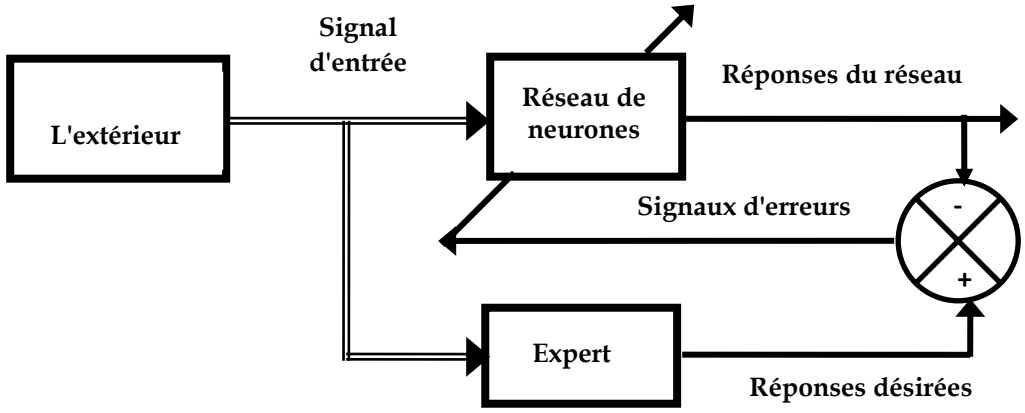


Figure II.4. Apprentissage non supervisé.

II.8.1.2. Apprentissage forcé ou imposé

Dans cet apprentissage on indique seulement au réseau si la sortie est bonne ou mauvaise en induisant ainsi une variation sélective des poids de connexions.

II.8.1.3. Apprentissage supervisé

L'apprentissage supervisé correspond au cas où l'on dispose d'exemples d'entrées et de sorties simultanément (c'est-à-dire les entrées et les sorties désirées correspondantes). On cherche alors à construire un modèle capable de prédire et donner la valeur de la sortie désirée connaissant celle de l'entrée et fournit un signal d'erreur quand une erreur est établie ; en minimisant celle-ci, on a un outil d'optimisation des paramètres internes du réseau qui sont les poids synaptiques. Cette procédure est répétée jusqu'à qu'un critère de performance soit satisfait figure II.5**.**

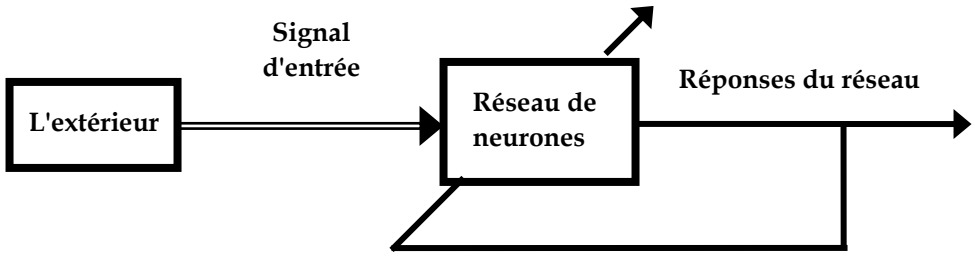


Figure II.5. Apprentissage supervisé

Le réseau de neurone génère une ou plusieurs valeurs de sorties basées sur la performance de calcul des valeurs d'entrées. Les résultats de ces calculs dépendent des valeurs des poids assignés à chaque entrée obtenue après la phase d'apprentissage.

La rétro propagation est la méthode d'apprentissage supervisé la plus utilisée pour l'entraînement des réseaux statiques et qui sera utilisée pour notre application.

Une fois l'apprentissage effectué, le réseau de neurones acquis la capacité de réaliser la tache prévue.

II.8.2. Algorithme de la rétro-propagation

La règle de la rétro-propagation est la règle la plus utilisée pour l'apprentissage supervisé des réseaux de neurones.

Le terme "rétro-propagation" est utilisé pour décrire l'apprentissage du réseau de neurones de type MLP utilisant la descente du gradient appliquée à la fonction de la somme des erreurs quadratique, mais elle peut être appliquée à n'importe quelle architecture à fonctions d'action dérivables.

Le principe de cet algorithme est de propager le signal de la couche d'entrée vers la couche de sortie (propagation avant), calculer ensuite l'erreur commise entre les sorties désirées et les sorties calculées par le réseau, et enfin de rétro propager cette erreur de la sortie vers l'entrée : ainsi, la rétro propagation de l'erreur est utilisée dans l'ajustement des poids synaptiques du réseau. Le réseau multicouche est présenté comme suit figure II.6 **:**

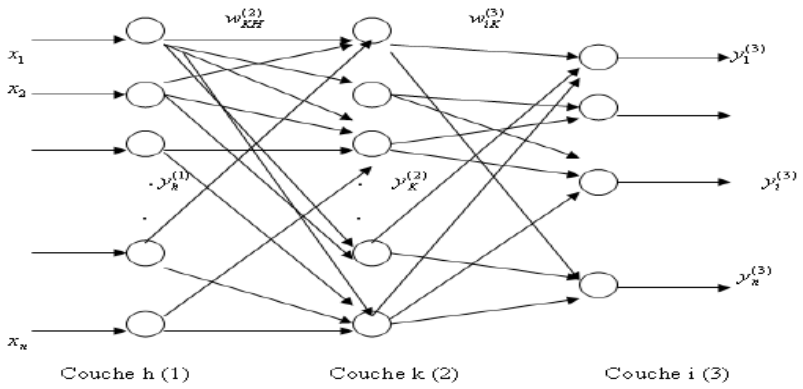


Figure II.6. Réseau de neurones multicouche

La somme pondérée peut être présentée dans la formule suivante :

$$Neti = \sum_{j=0}^{N} Wij\,Xj \qquad \text{II.3}$$

Tel que :

- *Xj :* est l'état du neurone de la couche précédente.
- Wij : Est le poids de la connexion reliant le neurone cible i au neurone émetteur *j*.
- *N* : le nombre de neurone dans la couche émettrice.

Les neurones auront à la sortie :

$$y_i = f(Net_i) = \frac{1}{(1+\exp(-Net_i))} \qquad \text{II.4}$$

Où :

$f(Net_i)$: est la fonction sigmoïde qui a pour dérivée :

$$f'(Net) = f(Net)(1-f(Net)) \qquad \text{II.5}$$

L'erreur quadratique est décrite par la formule suivante

$$E = \frac{1}{2}\sum_{i=1}^{N}(y_i - y_{des(i)})^2 \qquad \text{II.6}$$

Avec :

$y_{des(i)}$ *:* la sortie désirée du neurone *i* de la sortie.

Calculons le gradient de cette erreur par rapport au poids $w_{KH}^{(2)}$.

$$\frac{\delta E}{\delta W_{KH}^{(2)}} = \sum_{i \in couche(3)} \frac{\delta E}{\delta Net_i^{(3)}} \cdot \frac{\delta Net_i^{(3)}}{\delta w_{KH}^{(2)}} \qquad \text{II.7}$$

$$\frac{\delta E}{\delta Net_i^{(3)}} = \frac{\delta E}{\delta y_i^{(3)}} \cdot \frac{\delta y_i^{(3)}}{\delta Net_i^{(3)}} \qquad \text{II.8}$$

Sachant que :

$$\frac{\delta E}{\delta y_i^{(3)}} = (y_i^{(3)} - y_{des}) \qquad \text{II.9}$$

$$\frac{\delta y_i^{(3)}}{\delta Net_i^{(3)}} = f'(Net^{(3)}) \qquad \text{II.10}$$

De l'équations (II.9) et (II.10) la formule (II.8) devient :

$$\frac{\delta E}{\delta Net_i^{(3)}} = (y_i^{(3)} - y_{des}) f'(Net^{(3)}) \qquad \text{II.11}$$

Et pour $\frac{\delta Net_i^{(3)}}{\delta w_{KH}^{(2)}}$ on aura :

$$\begin{aligned}
\frac{\delta Net_i^{(3)}}{\delta w_{KH}^{(2)}} &= \frac{\delta}{\delta w_{KH}^{(2)}} \left[\sum_{i \in couche(2)} w_{iK}^{(3)} \, y_K^{(2)} \right] \\
&= w_{iK}^{(3)} \frac{\delta}{\delta w_{KH}^{(2)}} \left(\sum_{i \in couche(2)} y_K^{(2)} \right) \\
&= w_{iK}^{(3)} \frac{\delta}{\delta w_{KH}^{(2)}} \left[\sum_{i \in couche(2)} \left(f(w_{KH}^{(2)} \, y_h^{(1)}) \right) \right] \\
&= w_{iK}^{(3)} \, f'(Net^{(2)}) \frac{\delta}{\delta w_{KH}^{(2)}} \left[\sum_{i \in couche(2)} w_{KH}^{(2)} \, y_h^{(1)} \right] \\
&= w_{iK}^{(3)} \, f'(Net^{(2)}) \frac{\delta}{\delta w_{KH}^{(2)}} \left[\sum_{i \in couche(2)} w_{KH}^{(2)} \, y_h^{(1)} \right] \qquad \text{II.12} \\
&= w_{iK}^{(3)} \, f'(Net^{(2)}) \, y_h^{(1)}
\end{aligned}$$

En remplaçant l'équation (II.11) et (II.12) dans l'équation (II.7) on aura l'expression de gradient d'erreur :

$$\frac{\delta E}{\delta w_{KH}^{(2)}} = \sum_{i \in couche(3)} (y_i^{(3)} - y_{des}) f'(Net^{(3)}) \, w_{iK}^{(3)} \, f'(Net^{(2)}) \, y_h^{(1)} \qquad \text{II.13}$$

Par notation :

$$\delta_i^{(3)} = (y_i^{(3)} - y_{des}) f'(Net^{(3)}) \qquad \text{II.14}$$

Alors la formule (II.13) devient :

$$\frac{\delta E}{\delta w_{KH}^{(2)}} = \sum_{i \in couch(3)} \delta_i^{(3)} w_{iK}^{(3)} f'(Net^{(2)}) y_h^{(1)}$$

$$= \left[\sum_{i \in couche(3)} \delta_i^{(3)} w_{iK}^{(3)} \right] f'(Net^{(2)}) y_h^{(1)} \quad \text{II.15}$$

Et par notation :

$$\delta_K^{(2)} = \left[\sum_{i \in couche(3)} \delta_i^{(3)} w_{iK}^{(3)} \right] f'(Net^{(2)}) \quad \text{II.16}$$

Alors la formule (II.13) devient :

$$\frac{\delta E}{\delta w_{KH}^{(2)}} = \delta_K^{(2)} y_h^{(1)} \quad \text{II.17}$$

Les poids de la couche 2 seront ajustés par la règle suivante :

$$\Delta w_{KH}^{(2)} = -\alpha \delta_K^{(2)} y_h^{(1)} \quad \text{II.18}$$

Tel que α : est le pas d'apprentissage.

D'une manière générale :

$$\Delta w_{KH}^{(j)} = -\alpha \delta_K^{(j)} y_h^{(j-1)} \quad \text{II.19}$$

Où l'erreur $\delta_K^{(j)}$ vaut :

- Pour les neurones de la dernière couche :

$$\delta_K^{(j)} = (y_K^{(j)} - y_{des}) \quad \text{II.20}$$

- Pour les neurones d'une couche interne :

$$\delta_K^{(j)} = \left[\sum_{i \in couche(j+1)} \delta_i^{(j+1)} w_{iK}^{(j+1)} \right] f'(Net^{(j)}) \quad \text{II.21}$$

II.8.2.1. Résumé de l'algorithme de la rétro-propagation

Cet algorithme suit les étapes suivantes :

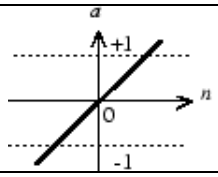
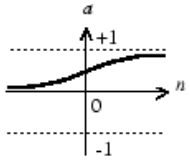
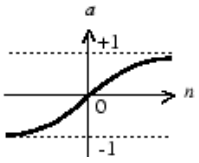
1. Initialiser les poids à des petites valeurs et les seuils du réseau.
2. Insérer à l'entrée du réseau une observation (exemple) de la base de données en forme de vecteur de caractéristique, puis calculer sa valeur d'activation et sa fonction d'activation selon les formules (II.3) et (II.4).
3. Évaluer le signal d'erreur des sorties du réseau en utilisant la formule (II.8).
4. Ajuster les poids en utilisant la formule (II.14).
5. Évaluer le signal d'erreur pour chaque couche cachée en utilisant la formule (II.12).
6. Ajuster les poids de la couche cachée en utilisant la formule (II.18).
7. Répéter les étapes de **2** à **6** pour l'ensemble des observations de la base d'apprentissage tant que le critère d'arrêt n'a pas été atteint.

II.9. Fonctions d'activation

La fonction d'activation calcule la valeur de l'état du neurone qui est transmise aux neurones avals. La fonction d'activation sert aussi à introduire une non-linéarité dans le fonctionnement du neurone. A la différence des neurones biologiques dont l'état est binaire, la plupart des fonctions de transfert sont continués, offrant une infinité de valeurs possibles comprises dans l'intervalle [0, +1] ou [-1 , +1]. Il existe de nombreuses formes possibles pour la fonction de transfert. Le tableau II.2 résume quelques fonctions de transfert utilisées dans le programme de calcul techniques MATLAB® (Hagan *et al.*, 1995).

En outre, du fait que la continuité et la différenciabilité des fonctions d'activation à tous les points sont les caractéristiques requises par les algorithmes d'optimisation actuels, les fonctions d'activation typiques qui accomplissent ces exigences sont la fonction linéaire, la fonction tangente hyperbolique et la fonction sigmoïde logarithmique. Selon Samarasinghe (2006) la fonction linéaire est plus appropriée pour les neurones de sorties alors que les fonctions de sigmoïdes sont plus appropriées pour les neurones des couches cachées.

Tableau II.2. Fonctions de transferts usuelles, a=f(t)

Nom de la fonction	Relation entré /sortie	Graphe de la fonction	Fonction MATLAB® correspondante
Linéaire	$a = n$		*purelin*
Logistique (sigmoide logarithmique)	$a = \frac{1}{1+e^{-n}}$		*logsig*
Tangente hyperbolique	$a = \frac{e^{+n}-e^{-n}}{e^{+n}+e^{-n}}$		*tansig*

II.10. Procédure de modélisation neuronale

Le modèle boite noire constitue la forme la plus primitive de modèle mathématique : il est réalisé uniquement à partir de données expérimentales ou d'observations ; il peut avoir une valeur prédictive, dans un certain domaine de validité, mais il n'a aucune valeur explicative (Dreyfuss *et al.*, 2002). A l'opposé, un modèle de connaissance, que l'on pourrait qualifier de boite blanche, est issu de l'analyse des phénomènes physiques, chimiques, biologiques, etc., dont résulte la grandeur que l'on chercher à modéliser. Ces phénomènes sont mis en équation à l'aide des connaissances théoriques disponibles au moment de l'élaboration du modèle. La démarche scientifique s'efforce, dans la mesure du possible, de construire des modèles de connaissance : l'élaboration d'un modèle de connaissance nécessite l'existence d'une théorie, alors qu'un modèle boite noire requiert seulement la disponibilité de mesures.

Entre la boite noire et le modèle de connaissance se situe le modèle semi- empirique, ou modèle boite grise, qui contient à la fois des équations résultant d'une théorie, et des équations purement empiriques, résultant d'une modélisation de type boite noire (Dreyfuss *et al.*, 2002).

Les réseaux de neurones constituent le plus souvent des modèles boite noire, néanmoins dans certaines conditions, ils peuvent être utilisés avec profit dans le cadre de modèles boite grise.

Un modèle statique est régi exclusivement par des équations algébriques ; un modèle dynamique, en revanche, obéit à des équations différentielles et à des équations algébriques (Dreyfuss *et al.*, 2002). Nous examinerons en bref, exclusivement, dans ce qui suit la conception d'un modèle statique type boite noire, réalisé à l'aide de réseaux de neurones bouclés (MLP) du fait qu'il a un lien direct avec le thème de ce travail.

L'objectif de la modélisation boite noire est de trouver, à partir des données ou des mesures disponibles, une relation déterministe, si elle existe, entre les variables d'entrées du modèle *x* et la grandeur à modéliser *y*.

La procédure de développement d'un réseau de neurone qui répond à une modélisation statique peut se résumer aux points suivants :

1. Sélection des données
2. Collecte des données
3. Séparation des bases de données
4. Normalisation des données
5. Choix du type de réseau de neurones
6. Choix du nombre de neurones dans la couche cachée
7. Choix de la fonction de transfert
8. Estimation des paramètres ou apprentissage
9. Validation et généralisation

II.10.1. Sélection des données

Lorsqu'on conçoit un modèle à partir de mesures, il est indispensable que l'ensemble des variables d'entrée soit aussi réduit que possible ; en effet, à chaque entrée du modèle correspond un nombre de paramètres égal au nombre de neurones cachés. Il est donc nécessaire

- Que la représentation des entrées soit aussi compacte que possible,
- Que la représentation des entrées contienne tous les facteurs pertinents par rapport à la grandeur à modéliser, mais seulement ceux-ci. En effet, la présence de variables qui ont, sur la grandeur à modéliser, moins d'influence que les perturbations ou le bruit, entraîne celle de paramètres superflus et introduit des variations non significatives sur les variables

d'entrées, qui constituent autant de sources d'erreur pour le modèle (Dreyfuss *et al.*, 2002).

Avant la conception d'un modèle neuronale, il convient donc de procéder à une sélection des entrées. Cette sélection revêt deux aspects :

- Réduction de la dimension de l'espace de représentation des variables du modèle,
- Élimination des entrées qui n'ont qu'une faible influence sur la sortie.

II.10.2. Collecte des données

Une fois les entrées pertinentes du modèle sont bien connues, l'objectif de cette étape est de les recueillir, à la fois pour développer (l'apprentissage) le réseau de neurones et pour le tester. La base de données reconstituée doit être représentative des données susceptibles d'intervenir en phase d'utilisation du système neuronal.

II.10.3. Séparation des bases de données

Afin de développer un modèle neuronal, il est nécessaire de disposer d'au moins deux bases de données : une base pour effectuer l'apprentissage et une autre pour valider le réseau obtenu et déterminer ses performances. On peut utiliser une troisième base de données, si elle est disponible, pour tester le réseau. Il n'y a pas de règle précise pour déterminer le partage de manière quantitative des données. Il résulte souvent d'un compromis tenant compte du nombre de données dont on dispose, l'essentiel, est que la base de données consacrées à l'apprentissage doit être la plus importante. Il est important de souligner que les sous-ensembles formés à partir de la base de données de départ devront être représentatifs de celle-ci.

II.10.4. Normalisation des données

Avant tout apprentissage, il est indispensable de normaliser et de centrer toutes les variables d'entrée. En effet, si des entrées ont des grandeurs très différentes, celles qui sont petites n'ont pas d'influence sur l'apprentissage. Il existe plusieurs approches de normalisation des données dont le choix dépend de la nature des données et de la fonction à approximer.

II.10.5. Choix du type de réseau de neurones

Comme il s'agit d'une modélisation statique type boite noire pour approcher une fonction bornée et continue, la meilleure approche neuronale envisageable est celle des MLP à une seule couche cachée.

II.10.6. Choix du nombre de neurones dans la couche cachée

Il n'existe pas de règle permettant de prévoir à l'avance le nombre de neurones dans la couche cachée. On pourrait penser qu'en plaçant un grand nombre de neurones dans la couche cachée, on obtiendrait systématiquement de meilleurs résultats. Mais il convient de prendre garde d'une part au temps de calcul (plus le nombre de neurones placés dans la couche cachée est important, plus le temps de calcul est long), d'autre part au risque de sur apprentissage (diminue la capacité de généralisation du réseau) ou de sous apprentissage lorsque leur nombre est insuffisant.

D'autre part, nous savons que le nombre des paramètres croit linéairement avec le nombre des variables d'entrées du modèle (parcimonie). Il va falloir donc trouver un compromis ente la parcimonie et le sur ou le sous apprentissage.

II.10.7. Choix de la fonction de transfert

La fonction de transfert neuronale, que l'on appelle aussi fonction d'activation, est la fonction mathématique appliquée à la sortie de chaque neurone. Là encore, il n'existe pas de méthode permettant de prévoir a priori quelle sera la fonction qui donnera les meilleurs résultats. Mais pour les MLP, il faut des fonctions de type sigmoïde pour les neurones de la couche cachée et généralement une fonction linéaire pour le neurone de la couche de sortie.

II.10.8. Estimation des paramètres ou apprentissage

Pour les modélisations statiques au moyen des MLP, il s'agit de faire un apprentissage supervisé.

Le critère d'arrêt de l'apprentissage est souvent calculé à partir d'une fonction de coût, caractérisant l'écart entre les valeurs de sortie obtenues et les valeurs désirées. Ce qui permet un arrêt adéquat de l'apprentissage pour obtenir de bonnes performances de généralisation.

L'obtention des paramètres architecturaux optimaux se fera par comparaison des performances obtenues pour différentes architectures de réseaux de neurones.

II.10.9. Validation et généralisation

La validation consiste à tester le réseau après ou au cours de l'apprentissage avec des données proches des données utilisées lors de la phase d'apprentissage. Les résultats de ce test permettront au modèle d'éviter les situations de sous ou de sur apprentissage. En revanche la généralisation s'effectue avec des données tout à fait nouvelles et différentes de la base de données de départ. Elle permet de savoir si le réseau a développé une capacité de généralisation assez performante.

II.11. Conclusion

Tout au long de ce chapitre, nous avons tenté de donner un maximum d'informations sur les réseaux de neurones d'une façon générale tout en mettant l'accent sur les réseaux de neurones non bouclés ou statiques, type MLP, vraisemblablement les plus utilisés dans les modélisations statiques du fait qu'ils sont doués de la propriété d'approximation universelle. Nous avons présenté un algorithme d'apprentissage très utilisé dans la communauté connexionniste : l'algorithme de rétro-propagation du gradient. Plusieurs travaux lui ont été consacrés et un certain nombre de modifications ont été proposées pour accélérer le temps d'apprentissage et améliorer la capacité de généralisation ainsi que les performances du réseau obtenu. Vers la fin, nous avons également présenté une procédure plus au moins détaillée d'une modélisation statique en utilisant les MLP.

Chapitre 03 :
Modélisation de la solubilité des solutés solides dans le CO2 supercritique.

CONTRIBUTION 01 :
Modélisation empirique et semi-empirique de la solubilité des solutés solides dans le dioxyde de carbone.

III.1.1. Introduction

La détermination précise de la solubilité des solutés solides dans les fluides supercritiques et sa corrélation sont essentielles pour le développement de la technologie supercritique dans les industries pharmaceutiques. Dans cette recherche, des nouvelles équations empiriques ont été proposées pour corréler la solubilité des solutés avec la température, la pression et la densité du CO_2 supercritique pur.

Dans cette contribution nous allons décrire la méthodologie suivie pour collecter la base de données de 107 composés solides, écrire les modèles testés de la littérature qui sont basés sur la densité du solvant et ceux proposés dans ce travail scindé sous deux séries, schématiser et interpréter des résultats trouvés.

III.1.2. Elaboration de la base de données

III.1.2.1. Introduction

Les données de mesures expérimentales de la solubilité des solutés solides dans les fluides supercritiques sont généralement présentées dans la littérature par les variables intensives : la température, la pression, la masse volumique du fluide supercritique en équilibre avec le soluté et la solubilité exprimé soit en unité de masse de soluté solide soluble dans un volume de solution en équilibre avec le soluté solide pur, soit en fraction molaire du soluté dans la phase supercritique. Il faut noter que les valeurs de la solubilité sont très faibles lorsqu'elles sont exprimées en fractions molaires pour les solutions considérées. Les sources des composées solides utilisées pour la modélisation par les méthodes semi-empirique et neuronales sont indiquées dans le tableau III.1. Ces données concernent donc 107 composés pour un totale de 3010 points expérimentaux. La majorité de ces composées sont des médicaments utilisés pour des différentes thérapies (antibiotique, antifongiques, anti-inflammatoires, antiépileptiques, anticholestérols, antioxydants ...etc.)

III.1.2.2. Structure de la base de données

Dans le but d'éviter les erreurs de saisies une quantité importante des données expérimentales de 107 composés sélectionnés ont été saisies dans des fichiers Excel à partir des fichiers PDF des références bibliographiques. Ces données concernent la température $T(°K)$, la pression $P(Bar)$, la densité du CO_2 $\rho_1(kg/m^3)$ et la solubilité du soluté solide dans le CO_2 supercritique y_2(en fraction molaire).

Tableau III.1. La liste des composés utilisés dans la modélisation

No.	Composés	Références
1	1,5-Naphthalenediamine	(Sparks et al., 2008)
2	3-Aminobenzoic acid	(Li et al., 2013)
3	4,4-diaminodiphenylmethane	(Khimeche et al., 2007)
4	4-Aminoantipyrine	(Chen and Chen, 2009)
5	4-Aminosalicylic acid	(Jin et al., 2014)
6	5-Fluorouracil	(Yamini et al., 2012)
7	Anastrozole	(Hojjati et al., 2009)
8	Aspirin	(Huang et al., 2004)
9	Atorvastatin	(Hojjati et al., 2007)
10	Atropine	(Yamini et al., 2001)
11	Azadirachtin	(Ismadji et al., 2011)
12	Azelaic acid	(Sparks et al., 2007)
13	Beclomethasone dipropionate	(Vatanara et al., 2005)
14	Benzocaine	(Garmroodi et al., 2004)
15	Benzoin	(Cheng et al., 2002)
16	Bisacodyl	(Asghari-Khiavi and Yamini, 2003)
17	Caffeine	(Johannsen and Brunner, 1994)
18	Capecitabine	(Yamini et al., 2012)
19	Carbamazepine	(Yamini et al., 2001)
20	Celecoxib	(Tsai et al., 2014)
21	Cetirizine	(Hezave et al., 2011)
22	Chlormezanone	(Lee et al., 2014b)
23	Cinnarizine	(Lee et al., 2014a)
24	Climbazole	(Park et al., 2009)
25	Clobetasol propionate	(Asiabi et al., 2013)
26	Clofenamic acid	(Li et al., 2015)
27	Clotrimazole	(Yamini and Moradi, 2011)
28	Clozapine	(Hosseini et al., 2010)
29	Codeine	(Yamini et al., 2001)
30	Cortisone acetate	(Dean et al., 1995)
31	Cyproheptadine	(Lashkarbolooki et al., 2013)
32	Cyproterone acetate	(Asghari-Khiavi et al., 2004)
33	Dexamethasone	(Chim et al., 2012; Tang et al., 2014)
34	Diazepam	(Yamini et al., 2001)
35	Diclofenac acid	(Hezave and Esmaeilzadeh, 2012)
36	Diflunisal	(Coimbra et al., 2008)
37	Docetaxel	(Yamini et al., 2012)
38	Dutasteride	(Yamini et al., 2010)
39	Erythromycin	(Burgos-Solórzano et al., 2004)
40	Exemestane	(Hojjati et al., 2009)
41	Finasteride	(Yamini et al., 2010)
42	Flurbiprofen	(Coimbra et al., 2006)
43	Flutamide	(Yamini et al., 2010)
44	Fluvastatin	(Hojjati et al., 2007)
45	Gabapentin	(Shojaee et al., 2013a)
46	Gemfibrozil	(Chen et al., 2010)
47	Ibuprofen	(Coimbra et al., 2006)
48	Irgacure® 2959	(Coimbra et al., 2008)
49	Isoniazid	(Heryanto et al., 2010)
50	Ketoconazole	(Yamini and Moradi, 2011)
51	Ketoprofen	(Coimbra et al., 2006)
52	Lamotrigine	(Hosseini et al., 2010)
53	Letrozole	(Hojjati et al., 2009)
54	Levonorgestrel	(Yamini et al., 2013)
55	Lovastatin	(Hojjati et al., 2007)
56	Mandelic acid	(Cheng et al., 2002)
57	Medroxyprogesterone acetate	(Asghari-Khiavi et al., 2004)
58	Mefenamic acid	(Hezave et al., 2012)
59	Megestrol acetate	(Yamini et al., 2013)
60	Menadione	(Reveco-Chilla et al., 2016)
61	Metaxalone	(Lee et al., 2014b)
62	Methimazole	(Yamini et al., 2003)
63	Methocarbamol	(Lee et al., 2014b)

64	Methylparaben	(Asghari-Khiavi and Yamini, 2003)
65	Metronidazole benzoate	(Garmroodi et al., 2004)
66	Nabumetone	(Su and Chen, 2008)
67	Naproxen	(Coimbra et al., 2006)
68	Nicotinic acid	(Kotnik et al., 2011)
69	Nifedipine	(Knez et al., 1995)
70	Niflumic acid	(Tsai et al., 2014)
71	Nimodipine	(Medina and Bueno, 2001)
72	Nitrendipine	(Zhan et al., 2015)
73	Norfloxacin	(Chim et al., 2012)
74	Ofloxacin	(Chim et al., 2012)
75	O-Nitrobenzoic acid	(Tang et al., 2012)
76	Oxymetholone	(Khamda et al., 2012)
77	Paclitaxel	(Suleiman et al., 2005)
78	Penicillin G	(Gordillo et al., 1999)
79	Penicillin V	(Ko et al., 1991)
80	Pentoxifylline	(Lee et al., 2014a)
81	Phenazopyridine	(Yamini et al., 2003)
82	Pindolol	(Weinstein et al., 2004)
83	Piracetam	(Lee et al., 2014a)
84	Piroxicam	(Shojaee et al., 2013)
85	Prednisolone	(Dean et al., 1995)
86	Prednisone	(Dean et al., 1995)
87	Progesterone	(Huang et al., 2014)
88	Propranolol	(Yamini et al., 2003)
89	Protocatechuic acid	(Paula et al., 2016)
90	Pyrocatechol	(García-González et al., 2001)
91	Quinine	(Zabihi et al., 2016)
92	Retinoic Acid	(Rosa et al., 2015)
93	Rosuvastatin	(Hojjati et al., 2007)
94	Salicylamide	(Su and Chen, 2008)
95	Salicylic acid	(Ke et al., 1996)
96	Simvastatin	(Hojjati et al., 2007)
97	Sinapic acid	(Paula et al., 2016)
98	Spironolactone	(Hezave et al., 2013)
99	Sulindac	(Hezave et al., 2012)
100	Tetramethylpyrazine	(Yongyue Sun et al., 2005)
101	Theobromine	(Johannsen and Brunner, 1994)
102	Theophylline	(Johannsen and Brunner, 1994)
103	Thymidine	(Suleiman et al., 2005)
104	Tolfenamic acid	(Li et al., 2015)
105	Triclocarban	(Park et al., 2009)
106	Uracil	(Burgos-Solórzano et al., 2004)
107	Zopiclone	(Medina and Bueno, 2001)

Avec N_i le nombre de points dans chaque système

III.1.3. Ajustement des paramètres

L'ajustement des paramètres de corrélation des données expérimentales est un problème d'optimisation multi variables. Les approches utilisées varient selon la fonction objective à minimiser et la méthode d'optimisation adoptée. La fonction objective utilisées pour la corrélation des données expérimentales qu'on trouve généralement dans les problèmes d'estimation des paramètres des modèles empiriques et semi-empiriques, sont obtenues à partir du principe des moindres carrés et du principe du maximum de vraisemblance. Il s'agit en fait d'une fonction objective suivante qui est utilisée pour les problèmes de corrélation de la

solubilité des solutés solides dans les fluides supercritiques par les modèles semi-empiriques ou par les équations d'états :

$$AARD(\%) = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{y_2^{exp}-y_2^{cal}}{y_2^{exp}}\right| \qquad \text{III.1}$$

Où : N est le nombre total des points expérimentales, y_2^{exp} et y_2^{cal} sont les valeurs de la solubilité expérimentale et calculé respectivement exprimé en fraction molaire.

III.2.3.1. Algorithme de l'ajustement des paramètres

L'ajustement des paramètres sélectionnés et proposés dans cette contribution a été réalisé à l'aide d'un programme MATLAB qui comporte les étapes suivantes :

1. La lecture des données expérimentales qui sont présenté dans un dossier qui contient les fichiers Excel utilisant la fonction $\boldsymbol{xlsread}$ pour former une matrice qui contient les valeurs expérimentales, la température en Kelvin, la pression en Bar, la densité de CO_2 en Kilogramme par mètre cube et la solubilité en fraction molaire, de chaque composé, et aussi un autre fichier Excel contient les masses molaires des composés pour les modèles qui nécessite les poids moléculaires (Eq.(3) et Eq.(5)).
2. Les approximations initiales des paramètres sont ajustées en minimisant la fonction objective (Eq. III.1). Nous avons utilisé dans ce cas la fonction $\boldsymbol{ga}$ de MATLAB qui utilise l'algorithme génétique.
3. Ces approximations initiales sont utilisées ensuite pour affiner les paramètres ajustables en utilisant la fonction $\boldsymbol{lsqcurvefit}$ de MATLAB qui utilise la méthode des moindres carrés non linéaires en fixant l'utilisation de l'algorithme de Levenberg-Marquardt comme option de calcul.

La combinaison entre les deux fonction précédentes ($\boldsymbol{ga}$ et $\boldsymbol{lscurvefit}$) a été exécuter 15 fois pour converger vers le minimum local.

4. Le calcul de la fonction objective (Eq. III.1), plus précisément l'erreur absolue relative moyenne, utilisant une fonction qui comporte les solubilités expérimentales et les solubilités calculées par le modèle considéré.

5. Illustration des résultats dans un graphe de régression linéaire de solubilité expérimentale en fonction de la solubilité calculé pour tous les points de la base de données, utilisant la fonction ***postreg*** de MATLAB.
6. Les étapes 1 à 5 sont réalisées pour les 107 composés de la base des données.

III.1.4. Les modèles de calcul de la solubilité

III.1.4.1. Les modèles basés sur la densité du solvant

L'application des équations d'état pour le calcul des équilibres de phases nécessite les paramètres des corps purs. Ces paramètres ne sont pas toujours disponibles, surtout pour les biomolécules et certains composés à intérêt pharmaceutiques. Même si ces propriétés sont estimées, comme c'est souvent le cas, par les méthodes de contribution des groupes, ces estimations engendrent un facteur supplémentaire d'incertitude à l'imprécision souvent constatée lors de l'application des équations d'état pour la modélisation de la solubilité des solutés solides dans un fluide supercritique.

C'est pourquoi plusieurs modèles empiriques et semi empiriques (Chrastil (1982), Ziger et Eckert (1983), Adashi et Lu (1983), del Valle et Aguilira (1988), Kumar et Johnston (1988), Bartle et al.. (1991), Yu et al.. (1994), Gordillo et al.. (1999), Mendez-Santiago et Teja (1999), Mendez-Santiago et Teja (2000), Jouyban et al.. (2002), Sauceau et al.. (2003), Sparks et al.. (2008)), basés sur la densité du solvant, ont été proposés pour la modélisation de la solubilité des solutés solides dans une phase supercritique. Une revue de la littérature de ces modèles et leur application est présentée dans le tableau III.2 ci-dessous, (Si-Moussa et al., 2016).

III.1.4.1.1. Présentation des résultats trouvés

Pour comparer la performance corrélative des modèles proposés avec d'autres modèles, un ensemble de neuf modèles ont été suggérés récemment (Eqs. (1), (2), (3), (4), (5), (6), (7), (8), (9)), (Bian et al., 2016; Si-Moussa et al., 2016; Tabernero et al., 2010) pour avoir la meilleure performance de corrélation.

Dans cette première partie, nous avons recueillie une base de données contient 3010 points expérimentaux et pour l'analyse des données nous avons utilisé un programme MATLAB pour estimer les paramètres des modèles sélectionnés précédemment.

Tableau III.2. Modèles basés sur la densité pour la corrélation de solutés solides dans des fluides supercritiques.

Modèles	Equations	N°	Références
Adachi et Lu	$lny_2 = a_0 + (a_1 + a_2\rho_1 + a_3{\rho_1}^2)ln\rho_1 + \frac{a_4}{T}$	(1)	(Adachi and Lu, 1984)
Jouyban et al	$lny_2 = a_0 + a_1P + a_2P^2 + a_3PT + a_4\frac{T}{P} + a_5ln\rho_1$	(2)	(Jouyban et al., 2002)
Sparks et al	$c_2^* = \rho_{r,1}^{(a_0+a_1+a_2\rho_{r,1}^2)}\exp(a_3 + \frac{a_4}{T_r} + \frac{a_5}{T_r^2})$ $c_2^* = \frac{c_2}{\rho_{c,1}}$; $\rho_{r,1} = \frac{\rho_1}{\rho_{c,1}}$; $T_r = \frac{T}{T_{c,1}}$	(3)	(Sparks et al., 2008)
Garlapati et Madras	$lny_2 = a_0 + \left(a_1 + a_2\rho_1 + \frac{a_3}{T} + a_4\, ln(\rho_1 T)\right)$	(4)	(Garlapati and Madras, 2010)
Bian et al	$C_2 = \rho_1{}^{(a_0+a_1\rho_1+a_2/lnT)}exp(\frac{a_3+a_4\rho_1}{T} + a_5)$ $C_2 = \frac{\rho_1+Mw_2Y_2}{Mw_1(1-Y_2)}$	(5)	(Bian et al., 2011)
Keshmiri et al	$lny_2 = a_0 + \frac{a_1}{T} + a_2P^2 + \left(a_3 + \frac{a_4}{T}\right)ln\rho_1$	(6)	(Keshmiri et al., 2014)
Khansary et al	$lny_2 = \frac{a_0}{T} + a_1P + \frac{a_2P^2}{T} + (a_3 + a_4P)ln\rho_1$	(7)	(Khansary et al., 2015)
Bian et al	$lny_2 = a_0 + \frac{a_1}{T} + \frac{a_2\rho_1}{T} + (a_3 + a_4\rho_1)ln\rho_1$	(8)	(Bian et al., 2016)
Si Moussa et al	$lny_2 = a_0 + a_1\rho_1 + a_2{\rho_1}^2 + a_3\rho_1 T + a_4\frac{T}{\rho_1} + a_5ln\rho_1$	(9)	(Si-Moussa et al., 2016)

Les modèles sélectionnés par Si-Moussa (Si-Moussa et al., 2016), mentionnés dans ce travail ont été évalués comme les meilleurs modèles corrélés pour prédire la solubilité des médicaments solides dans les fluides supercritiques.

Pour comparer la précision de chaque modèle on a calculé deux constantes : l'$AARD$ (Eq. III.1) et R^2 (Eq. III.2).

R^2 , peut-être écrit comme :

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_2^{exp} - y_2^{cal})}{\sum_{i=1}^{N}(y_2^{exp} - \sum_{i=1}^{N} y_2^{exp}/N)} \qquad \text{III.2}$$

Avec : N est le nombre total des points expérimentales, y_2^{exp}, y_2^{cal} sont les solubilités expérimentale et calculées en fraction molaire respectivement.

Les résultats de calcul des $AARD$ (%) et R^2 pour les modèles sélectionnés sont regrouper dans le tableau III.3, les colonnes représentent les modèles choisis et les lignes représentes les valeurs des $AARD$ (%) pour les 107 composés. Le meilleur résultat de corrélation pour chaque soluté solide est montré dans ce tableau mentionné en gras.

Tableau III.3. Résultats de calcul des AARD et R^2 pour les modèles sélectionnés

N°	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	6,97	5,57	6,55	7,42	3,48	**2,98**	6,62	3,81	4,01
2	2,99	4,60	**2,22**	2,80	2,67	3,30	3,02	2,69	2,74
3	35,62	**33,68**	35,51	34,95	35,70	36,04	35,15	34,87	35,76
4	**6,30**	7,13	6,38	7,35	6,81	7,59	8,84	7,82	6,73
5	1,91	4,15	**1,56**	2,18	2,29	2,26	3,53	2,29	1,92
6	7,77	6,74	**5,27**	8,58	7,23	7,69	8,17	7,83	6,72
7	6,27	10,15	**6,25**	7,25	7,00	9,26	10,06	7,65	6,97
8	3,40	5,65	**2,63**	5,66	3,02	5,39	5,44	4,65	3,31
9	5,08	11,94	3,97	6,17	4,08	9,09	10,64	5,51	**3,84**
10	19,93	19,45	16,94	18,71	15,41	16,68	16,80	**15,18**	16,24
11	3,16	5,25	2,68	3,49	2,78	**2,60**	3,70	3,43	3,05
12	21,51	17,69	21,54	21,86	16,81	16,81	**15,80**	20,22	17,04
13	9,65	**6,53**	9,26	10,03	8,20	8,08	9,14	8,14	7,66
14	11,22	**8,33**	11,06	11,43	11,17	11,05	12,55	11,16	9,67
15	5,31	4,36	4,79	5,97	3,17	3,87	6,13	4,76	**2,80**
16	9,92	**7,17**	9,98	12,23	11,03	12,96	15,06	11,02	9,16
17	**2,87**	2,88	2,91	3,09	3,02	3,20	5,42	3,24	2,90
18	7,52	11,71	**5,68**	10,63	5,98	11,07	11,44	8,57	6,95
19	12,22	15,04	12,22	15,83	12,86	17,63	18,02	15,78	**11,85**
20	1,94	2,87	1,80	2,03	1,64	1,75	2,93	1,74	**1,42**
21	4,17	8,12	4,18	4,88	4,78	5,84	11,18	5,43	**4,11**
22	4,13	4,76	**4,10**	5,14	4,19	5,61	7,34	4,97	4,13
23	4,85	4,80	4,75	6,17	**3,94**	5,63	8,35	6,04	4,12
24	3,92	3,72	3,34	4,22	2,50	2,83	5,51	3,37	**2,39**
25	14,29	12,02	14,60	15,94	12,62	14,80	16,71	12,48	**10,95**
26	5,05	6,76	5,01	5,23	**4,75**	4,97	5,99	5,04	4,80
27	3,31	8,63	2,71	4,89	**2,69**	7,03	8,34	4,54	3,71
28	24,65	17,68	23,82	24,58	18,83	22,67	25,72	18,84	**15,12**
29	12,46	**10,69**	12,71	12,08	13,09	14,28	13,31	12,43	12,68
30	13,34	23,84	**12,14**	23,68	13,57	21,66	37,53	21,31	15,40
31	8,02	**6,90**	7,13	9,07	6,92	8,85	9,96	7,43	7,04
32	11,58	10,25	10,24	11,71	**5,26**	10,32	9,70	8,62	5,55
33	7,74	**4,39**	7,73	8,27	7,32	6,48	5,89	7,53	7,42
34	7,77	5,00	7,39	7,58	4,13	7,93	8,11	4,58	**3,53**
35	8,19	13,19	5,39	7,31	**4,13**	5,81	5,84	4,24	5,01
36	12,57	10,05	13,02	13,20	11,00	15,54	14,52	11,20	**11,05**
37	7,77	9,25	8,13	8,21	7,69	7,50	8,74	7,67	**6,62**
38	5,82	10,80	**4,08**	8,17	4,19	10,01	10,26	6,94	5,05
39	13,08	**1,85**	9,42	12,98	6,25	9,23	8,29	9,56	6,09
40	23,13	26,38	**10,77**	20,68	17,07	19,91	17,71	17,05	20,60
41	17,05	15,28	16,35	19,04	12,99	19,00	21,22	13,14	**11,57**
42	5,32	5,85	**5,16**	5,46	5,18	6,19	5,97	5,40	5,23
43	19,86	**13,39**	19,77	19,72	17,06	15,90	18,84	17,54	15,63
44	8,77	6,87	8,12	10,41	5,33	11,42	12,19	6,28	**3,71**
45	10,21	11,77	10,23	10,82	**9,29**	10,19	12,58	10,45	9,61
46	**10,65**	15,42	11,56	11,81	11,42	13,98	16,48	10,86	11,35
47	8,43	11,35	8,53	11,90	8,39	10,63	9,73	11,86	**6,22**
48	2,18	5,25	**2,14**	2,35	2,48	2,77	3,12	2,49	2,39
49	8,17	6,51	6,91	8,17	5,53	5,70	6,73	5,53	**5,37**
50	13,15	18,40	**7,61**	15,16	9,27	18,89	14,67	12,68	12,20
51	8,49	6,20	8,34	10,84	6,20	9,31	9,86	8,38	**6,18**
52	3,24	3,97	3,14	4,12	**3,13**	4,32	3,71	3,97	3,53
53	15,70	18,75	15,23	19,71	14,18	19,84	19,50	18,33	**13,96**
54	9,52	11,02	10,03	9,69	9,62	**9,51**	10,52	9,65	9,54

55	4,63	7,90	**4,02**	5,33	4,09	4,88	4,25	4,95	4,29
56	3,41	4,49	**3,40**	3,84	3,93	7,57	7,75	4,11	3,56
57	23,64	26,21	25,70	24,09	25,03	27,77	26,21	24,97	**23,64**
58	4,06	5,96	3,71	3,89	3,58	3,89	6,98	3,62	**3,42**
59	23,15	23,88	19,62	22,57	20,44	19,99	**18,96**	21,31	21,10
60	1,39	1,15	0,75	1,36	0,65	1,00	1,35	0,83	**0,60**
61	22,93	**11,68**	23,69	23,34	23,05	22,18	19,93	22,59	22,89
62	14,48	**9,08**	14,26	13,73	11,19	12,43	13,49	12,24	9,46
63	4,02	5,96	4,02	4,29	**3,46**	4,28	5,13	4,21	3,82
64	5,13	**5,02**	5,09	5,47	5,97	7,24	7,47	5,84	5,21
65	13,75	14,43	**9,27**	14,56	14,35	14,80	19,48	14,85	12,86
66	2,99	9,02	2,91	3,63	2,06	5,59	5,68	3,82	**1,75**
67	3,86	4,94	3,81	4,06	**3,67**	3,84	4,34	3,74	3,77
68	21,25	**16,02**	21,16	21,37	18,80	18,83	26,28	19,37	17,68
69	16,55	**10,00**	16,35	16,35	13,54	13,15	17,14	14,49	12,16
70	2,99	3,64	1,43	2,82	1,24	**1,08**	2,65	1,27	1,15
71	7,55	8,45	7,38	7,86	5,16	6,65	10,14	7,68	**4,80**
72	7,91	9,11	7,67	9,45	**6,68**	7,55	9,68	8,70	8,25
73	29,30	13,42	29,66	33,23	**11,40**	19,66	27,33	13,59	14,93
74	8,85	10,50	8,83	9,17	**5,10**	7,80	8,42	9,00	6,64
75	3,12	6,06	**2,85**	3,53	3,52	4,10	5,23	3,77	3,12
76	3,87	3,08	3,96	4,07	3,04	3,65	4,89	3,22	**2,90**
77	8,66	9,30	8,65	12,12	**8,18**	11,04	10,08	10,85	8,47
78	23,28	**11,04**	22,30	23,48	15,51	24,05	19,19	16,32	15,57
79	3,92	5,42	3,69	9,18	**3,63**	7,10	8,30	9,11	4,97
80	5,62	6,54	5,70	6,06	**5,49**	5,75	6,61	6,05	5,61
81	7,40	7,20	7,41	8,28	7,32	11,08	10,89	7,55	**6,67**
82	8,58	14,08	8,38	8,47	8,03	16,36	13,76	8,85	**7,67**
83	3,83	3,37	3,05	4,37	2,86	3,66	5,28	3,97	**2,75**
84	17,21	15,17	14,29	17,09	13,63	14,42	14,40	**13,60**	13,86
85	18,83	38,04	**17,97**	33,36	19,53	29,66	55,77	30,38	19,03
86	27,75	36,81	26,64	34,63	24,24	22,91	48,31	35,29	**21,04**
87	6,90	10,91	**5,19**	9,41	6,57	13,23	14,51	9,84	5,92
88	16,59	**7,47**	15,61	16,01	12,66	14,05	15,17	13,22	11,29
89	9,80	5,22	9,29	10,35	6,91	6,41	**5,01**	7,18	7,46
90	6,58	10,34	6,45	8,34	**6,39**	9,94	8,52	7,44	6,57
91	14,29	14,31	13,74	19,82	**11,98**	13,32	13,37	16,91	12,88
92	6,13	11,44	5,30	6,99	5,84	6,77	**5,01**	6,81	6,49
93	3,34	7,69	**3,18**	7,20	3,60	8,84	9,83	7,00	3,82
94	2,73	7,69	2,69	3,52	2,31	5,01	4,31	3,57	**1,78**
95	3,61	3,70	3,61	3,64	3,67	3,83	3,87	**3,61**	3,63
96	6,39	9,38	6,38	9,62	6,66	11,29	12,95	9,59	**5,92**
97	10,88	19,18	10,61	26,96	**9,62**	26,95	27,01	25,12	10,44
98	15,63	20,04	8,36	15,15	**8,13**	11,41	13,75	8,41	9,68
99	7,31	12,69	7,36	8,67	8,37	9,46	10,97	9,22	**6,70**
100	2,18	10,54	2,09	3,03	2,18	2,99	**2,90**	3,37	3,63
101	16,44	4,17	16,60	16,80	16,43	4,00	**3,92**	16,59	16,25
102	5,37	**4,19**	5,32	5,30	5,03	4,82	5,09	5,26	4,61
103	23,37	22,89	**13,88**	24,42	22,48	22,41	22,11	23,63	22,88
104	4,78	7,91	4,48	4,90	**3,58**	4,31	4,81	3,70	3,60
105	5,94	**3,56**	5,82	5,72	4,55	5,17	6,05	5,34	5,15
106	21,04	18,93	21,35	22,61	**10,91**	28,67	30,47	14,18	12,80
107	5,92	5,32	5,95	6,42	**4,27**	5,45	7,26	5,48	5,77
Min AARD	1,39	1,15	0,75	1,36	0,65	1,00	1,35	0,83	**0,60**
Max AARD	35,62	38,04	35,51	**34,95**	35,70	36,04	55,77	35,29	35,76
Moyenne	9,97	10,42	9,20	11,22	8,41	10,62	12,01	9,83	**8,35**
R^2	0,9946	0,9947	0,9954	0,9939	**0,9959**	0,9947	0,9934	0,9950	**0,9959**

III.1.4.1.2. Discussion des résultats trouvés

Le tableau III.3 montre que la gamme des valeurs de AARD est entre 0.60% et 55.77%, mais quand nous prendrons les meilleures des AARD mentionné en gras, nous trouverons une gamme des valeurs entre 0.60% et 33.68% avec une moyenne de 7.15%.

L'AARD minimum (0.60%) est obtenu par l'équation de Si-Moussa et al. (Eq. (9)) pour le système N° 60 ($scCO_2$ - Menadione), et l'AARD maximum (55.77%) est obtenu par l'équation de Khansary et al. (Eq. (7)) pour le système N° 85 ($scCO_2$ - Prednisolone).

La comparaison globale entre la performance des modèles indique que le modèle de Si-Moussa et al. (Eq. (9)) a donné la meilleure corrélation, ou la gamme de l'ARRD est entre 0.60% et 35.76% avec une moyenne de 8.35%. Le second modèle qui a la meilleure corrélation est le modèle de de Bian et al. (Eq. (5)), ou la gamme de l'AARD est entre 0.65% et 35.70% avec une moyenne de 8.41%. Le troisième avec une meilleure corrélation est le modèle de Sparks et al. (Eq. (3)), ou la gamme de l'AARD est entre 0.75% et 35.51% avec une moyenne de 9.20%. Ces trois modèles ensemble représente une meilleure corrélation de 67% des 107 composés considéré.

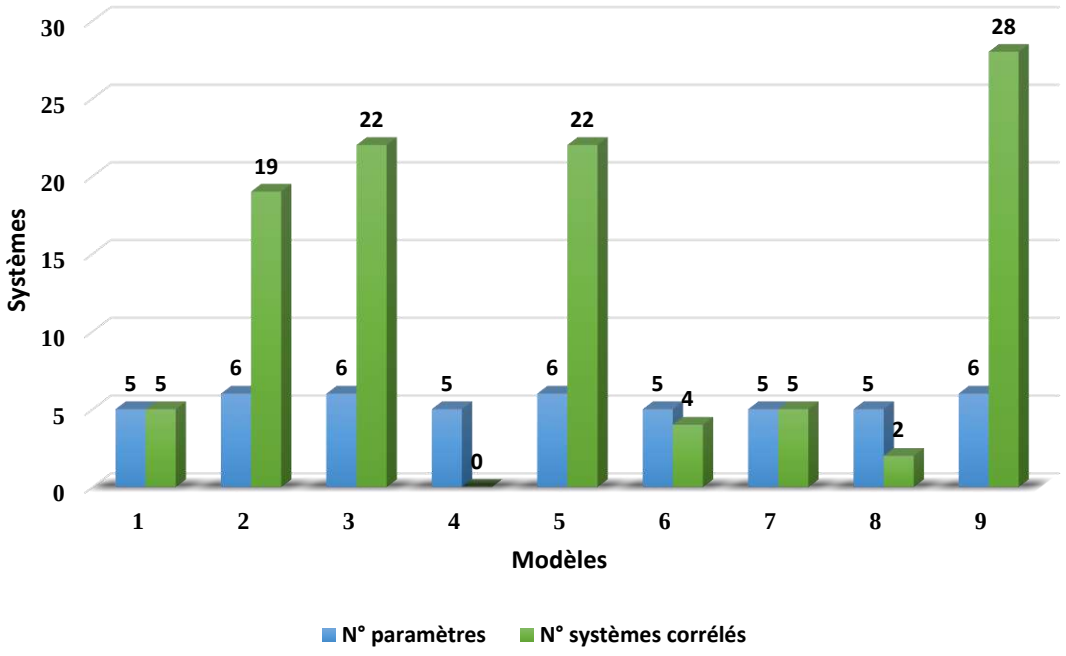


Figure III.1. Le nombre des systèmes corrélé et le nombre des paramètres pour chaque modèle.

III.1.4.2. Modèles proposés (série 01)

Dans cette série nous avons proposé une série des modèles contient des combinaisons entre les variables intensives (T, P, ρ_1) et la solubilité en fraction molaire (y_2). Le tableau III.4 représente la série 01 des modèles proposés.

Tableau III.4. Modèles proposés pour la corrélation de solutés solides dans des fluides supercritiques (série 01).

Modèles	N°	Références
$lny_2 = a_0 + a_1\rho_1 + a_2\rho_1 T + a_3 ln(\rho_1) + a_4 ln(\rho_1 T)$	10	Ce travail
$lny_2 = a_0 + (a_1 + a_2\rho_1)\, log(\rho_1) + a_3\rho_1 T + a_4 ln(\rho_1 T)$	11	Ce travail
$lny_2 = a_0 + a_1\rho_1 + a_2 T + a_3 T^2\rho_1^2 + a_4 \ln(T\rho_1)^2 + a_5 ln\rho_1$	12	Ce travail
$lny_2 = c_1(a_0 + a_1\rho_1 + a_2\rho_1^2) ln\rho_1 + c_2(b_0 + b_1 T + b_2 T^2) lnT$	13	Ce travail

III.1.4.2.1. Présentation des résultats trouvés

Les résultats de calcul des $AARD$ (%) et R^2 pour les modèles proposés (série 01) sont regroupés dans le tableau III.5, les colonnes représentent les modèles choisis et les lignes représentes les valeurs des $AARD$ (%) pour les 107 composés. Le meilleur résultat de corrélation pour chaque soluté solide est montré dans ce tableau mentionné en gras.

Tableau III.5. Résultats de calcul des AARD et R^2 pour les modèles proposés (série 01)

N°	(10)	(11)	(12)	(13)
1	**3,84**	4,08	6,54	3,88
2	2,83	2,82	**2,24**	2,70
3	34,92	34,96	**34,73**	35,15
4	7,80	7,74	**6,39**	7,69
5	2,14	2,14	2,02	**1,85**
6	8,14	8,13	**5,89**	7,98
7	6,84	6,85	**6,40**	7,27
8	4,79	4,71	**3,00**	5,49
9	5,11	5,06	**3,99**	6,70
10	**15,77**	15,83	17,09	16,57
11	3,50	3,44	**2,69**	3,53
12	20,57	20,80	21,51	**17,83**
13	**6,95**	**6,95**	9,64	7,51
14	9,49	9,60	11,25	**9,16**
15	4,78	4,79	4,75	**4,19**
16	**8,94**	8,96	10,06	8,96
17	3,09	3,09	**2,87**	2,94
18	8,76	8,71	**7,11**	9,64
19	15,21	**15,20**	15,32	15,24
20	1,71	1,72	1,80	**1,35**
21	4,78	4,79	**4,27**	4,65
22	4,71	4,68	**4,07**	4,90
23	6,01	6,04	**5,24**	5,87
24	3,43	3,45	3,35	**3,14**
25	**10,62**	10,67	14,55	11,06
26	5,06	5,09	5,01	**4,94**
27	4,43	4,43	**2,80**	5,42
28	18,60	**18,56**	22,70	19,05
29	11,75	**11,70**	12,07	12,09
30	21,50	20,54	**11,83**	20,75
31	7,48	**7,47**	7,78	7,78
32	8,74	8,98	10,18	**5,63**
33	7,65	7,85	7,90	**6,88**
34	4,06	4,18	7,46	**3,09**
35	**5,04**	5,10	5,55	5,64
36	10,55	**10,54**	12,67	12,24
37	6,55	6,56	7,75	**6,36**
38	7,13	6,98	**3,98**	9,12
39	9,59	9,66	9,75	**9,22**
40	19,70	19,70	**11,29**	22,15
41	**11,67**	11,73	16,49	14,09
42	5,44	5,45	**5,28**	5,51
43	16,05	16,34	19,77	**15,11**
44	4,31	**4,17**	8,20	5,91
45	10,45	10,46	**10,34**	10,65
46	10,75	**10,67**	12,33	11,53
47	11,67	11,23	**8,09**	10,52
48	2,30	2,30	**2,14**	2,26
49	**5,74**	**5,74**	6,85	5,90
50	13,52	13,46	**7,99**	16,30
51	8,15	**8,03**	8,49	9,44

52	3,83	3,78	**3,17**	4,12
53	17,46	17,18	**15,19**	19,55
54	9,65	**9,62**	9,92	9,81
55	5,13	5,07	**4,04**	5,77
56	3,65	3,69	**3,39**	3,50
57	24,34	24,24	**23,27**	24,09
58	3,48	**3,47**	**3,74**	3,53
59	21,93	21,98	**19,95**	22,07
60	0,98	1,05	**0,75**	0,74
61	**22,77**	22,81	22,83	23,16
62	11,00	11,10	14,04	**10,66**
63	4,21	4,24	**4,00**	4,29
64	5,24	5,25	5,00	**4,58**
65	14,47	14,50	**9,03**	13,28
66	3,66	3,67	2,90	**2,76**
67	3,74	**3,72**	4,17	4,02
68	18,58	18,93	21,23	**16,65**
69	13,68	13,85	16,32	**12,53**
70	**1,37**	1,38	1,42	1,51
71	7,80	8,09	7,33	**5,25**
72	9,04	9,28	7,63	**7,23**
73	12,25	**11,15**	29,21	18,73
74	8,84	**8,47**	8,58	9,18
75	3,52	3,54	**2,75**	3,12
76	2,90	**2,85**	3,84	3,22
77	10,91	10,87	**8,67**	10,91
78	17,43	17,84	22,10	**16,58**
79	9,18	9,16	**3,69**	8,02
80	6,04	6,05	**5,71**	5,96
81	6,81	**6,80**	7,17	7,24
82	8,95	8,47	8,37	**7,90**
83	4,06	4,11	**3,04**	3,29
84	14,25	14,26	**13,88**	14,18
85	30,93	30,29	**18,88**	31,44
86	33,53	33,37	27,27	**22,78**
87	9,14	9,15	**7,77**	8,69
88	12,30	12,51	15,82	**10,17**
89	**7,03**	7,05	8,67	7,20
90	7,47	7,68	**6,54**	7,17
91	18,45	18,90	15,85	**13,19**
92	7,01	**6,92**	8,93	7,16
93	6,53	6,46	**3,25**	7,35
94	3,52	3,57	2,73	**2,50**
95	3,61	**3,60**	3,60	3,65
96	8,61	8,51	**6,31**	9,06
97	25,58	25,54	26,99	**25,03**
98	11,17	11,12	**8,30**	11,64
99	7,52	7,51	8,48	**7,17**
100	2,90	2,93	2,53	**2,40**
101	**16,50**	**16,50**	16,83	16,53
102	4,78	4,79	5,33	**4,71**
103	23,37	23,35	**14,07**	23,03
104	3,73	**3,70**	4,60	4,13
105	**5,38**	5,43	5,80	5,39
106	15,58	16,20	21,02	**13,16**
107	5,06	**4,92**	5,93	6,47
Min *AARD*	0,98	1,05	0,75	**0,74**
Max *AARD*	34,92	34,96	**34,73**	35,15
Moyenne	9,66	9,65	**9,47**	9,54
R^2	0.9951	0.9951	0,9950	**0,9952**

III.1.4.2.2. Discussion des résultats trouvés

Le tableau III.5 montre que la gamme des valeurs de AARD est entre 0.74% et 35.15%, mais quand nous prendrons les meilleures des AARD mentionné en gras, nous trouverons une gamme des valeurs entre 0.74% et 34.73% avec une moyenne de 8.28%.

L'AARD minimum (0.74%) est obtenu par l'équation (Eq. (13)) pour le système N° 60 ($scCO_2$ - Menadione), et l'AARD maximum (35.15%) est obtenu aussi par l'équation (Eq. (13)) pour le système N° 3 ($scCO_2$ -4,4-diaminodiphenylmethane).

La comparaison globale entre la performance des modèles indique que le modèle (Eq. (12)) a donné la meilleure corrélation, ou la gamme de l'ARRD est entre 0.75% et 34.73% avec une moyenne de 9.47%. Le second modèle qui a la meilleure corrélation est le modèle (Eq. (13)), ou la gamme de l'AARD est entre 0.74% et 35.15% avec une moyenne de 9.54%. Le troisième avec une meilleure corrélation est le modèle (Eq. (11)), ou la gamme de l'AARD est entre 1.05% et 34.96% avec une moyenne de 9.65%. Ces trois modèles ensemble représente une meilleure corrélation de 92% des 107 composés considéré.

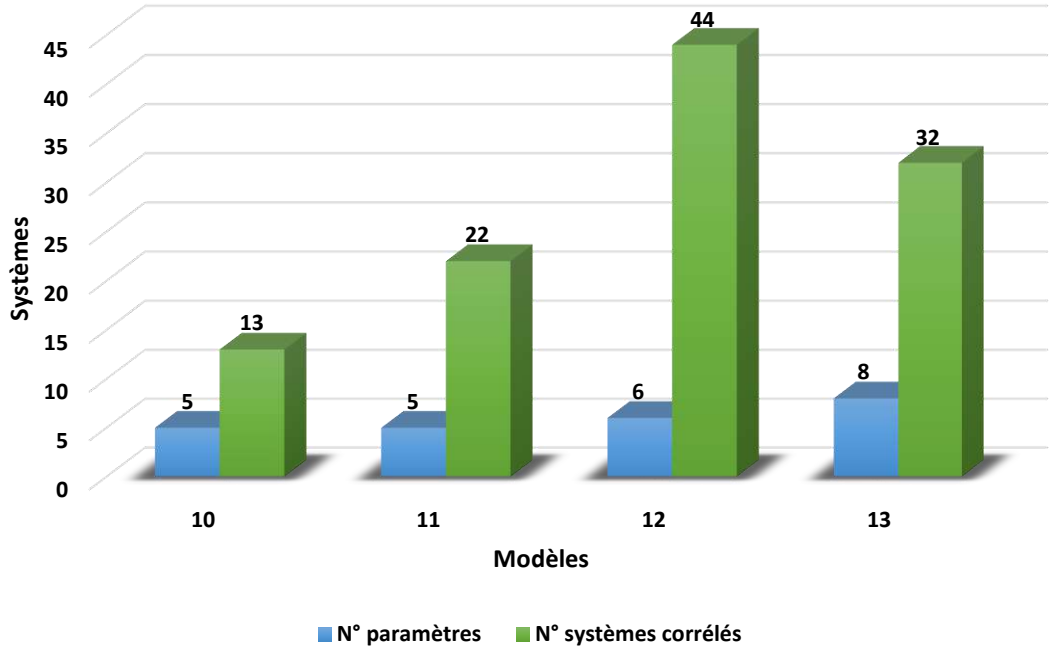


Figure III.2. Le nombre des systèmes corrélés et le nombre des paramètres pour chaque modèle.

III.1.4.3. Modèles proposés (série 02)

Dans cette série nous avons proposé une série des modèles contient des modifications aux modèles trouvés dans la littérature. Les modèles que nous avons modifiés sont : del Vall et Aguilira (1988), Gardillo et al. (1999), Khansary et al. (Eq. (7)) et Si-Moussa et al. (Eq. (9)).

Tableau III.6. Modèles proposés pour la corrélation de solutés solides dans des fluides supercritiques (série 02).

Modèles	N°	Type de changement	Références
$lny_2 = a_0 + (a_1 + a_2\rho_1 + a_3\rho_1^2)ln\rho_1 + \frac{a_4}{T} + \frac{a_5}{T^2}$	(14)	Ajoute du polynôme d'Adachi et Lu à l'équation de del Valle et Aguilira, (1988)	Ce travail
$lny_2 = a_0 + a_1\rho_1 + a_2{\rho_1}^2 + a_3\rho_1 T + a_4 T + a_5 T^2$	(15)	Changement de P par ρ_1 dans l'équation de Gardillo et al, (1999)	Ce travail
$lny_2 = \frac{a_0}{T} + a_1\rho_1 + \frac{a_{2\rho_1^2}}{T} + a_3 ln\rho_1$	(16)	Changement de P par ρ_1 dans (Eq. (7))	Ce travail
$lny_2 = a_0 + a_1\rho_1 + a_2{\rho_1}^2 + a_3\rho_1 T + a_4\frac{T}{\rho_1} + (a_5 + a_6\rho_1 + a_7\rho_1^2)ln\rho_1$	(17)	Ajoute du polynôme d'Adachi et Lu à (Eq. (9))	Ce travail

III.1.4.3.1. Présentation des résultats trouvés

Les résultats de calcul des $AARD$ (%) et R^2 pour les modèles proposés (série 02) sont regrouper dans le tableau III.7, les colonnes représentent les modèles choisis et les lignes représentes les valeurs des $AARD$ (%) pour les 107 composés. Le meilleur résultat de corrélation pour chaque soluté solide est montré dans ce tableau écrit en gras.

Tableau III.7. Résultats de calcul des AARD pour les modèles proposés (série 02)

N°	(14)	(15)	(16)	(17)
1	6,57	**2,71**	8,46	3,09
2	2,26	**2,15**	2,77	2,29
3	35,56	34,84	34,84	**31,33**
4	**6,29**	6,75	7,33	6,47
5	1,68	1,94	2,63	**1,58**
6	**5,23**	5,63	7,71	6,66
7	6,57	6,68	7,44	**6,05**
8	**2,63**	4,52	5,10	3,19
9	3,97	4,43	5,06	**3,36**
10	16,07	**15,48**	17,65	15,73
11	**2,66**	3,98	3,65	2,73
12	21,51	20,42	22,08	**16,46**
13	9,42	**6,24**	10,64	6,67
14	11,15	**7,17**	11,89	9,37
15	4,70	3,90	6,38	**2,67**
16	10,04	**6,50**	12,48	9,18
17	2,88	2,85	3,46	**2,62**
18	**5,68**	8,59	9,06	6,65
19	12,09	15,71	16,95	**10,77**
20	1,78	1,43	2,76	**1,01**
21	4,22	4,48	5,60	**4,07**
22	4,06	4,52	4,44	**3,74**
23	4,75	5,66	6,48	**3,54**
24	3,36	2,82	5,49	**2,04**
25	14,39	10,19	15,37	**9,80**
26	4,99	5,01	5,37	**4,92**
27	**2,76**	3,33	3,84	2,86
28	23,83	17,90	24,89	**10,85**
29	12,67	11,56	12,39	**10,16**
30	**11,83**	18,73	30,70	14,77
31	7,14	7,63	8,32	**6,88**
32	10,12	5,22	12,56	**4,90**
33	7,72	7,29	8,29	**6,46**
34	7,45	**3,00**	9,47	3,28
35	5,51	4,39	6,45	**4,33**
36	12,47	10,51	14,38	**10,47**
37	7,81	6,35	7,78	**6,31**
38	**3,98**	5,46	4,89	4,55
39	13,01	8,59	11,74	**3,31**
40	10,68	**10,25**	19,33	20,48
41	16,34	13,12	21,52	**11,83**
42	5,15	5,51	5,77	**4,44**
43	19,76	**13,88**	19,48	13,92
44	8,27	4,23	10,45	**3,64**
45	10,23	10,18	10,87	**9,36**
46	10,61	12,32	15,67	**8,66**
47	8,40	8,68	8,95	**4,17**
48	2,23	2,53	3,18	**2,12**
49	8,31	5,37	8,08	**5,12**
50	**8,30**	11,20	14,58	12,12
51	8,49	7,86	8,70	**4,73**
52	3,15	3,20	3,76	**2,76**
53	14,86	17,28	17,53	**13,38**
54	9,50	9,94	9,49	**8,43**
55	**4,05**	4,40	4,23	4,31
56	**3,39**	3,45	5,31	3,64
57	23,97	**22,43**	24,30	25,34
58	3,61	3,48	3,72	**3,43**
59	19,88	19,93	22,25	**14,53**
60	1,17	0,68	1,73	**0,54**
61	22,96	25,64	23,53	**22,86**
62	14,00	10,18	13,93	**8,92**
63	4,01	3,91	4,34	**3,56**
64	5,13	**3,95**	6,43	5,06
65	10,43	**8,69**	15,26	12,95
66	2,89	6,73	8,14	**1,73**
67	3,82	3,79	4,06	**3,15**
68	21,24	**16,31**	24,63	17,42
69	16,00	12,18	15,77	**11,58**
70	1,43	1,21	1,42	**0,97**
71	7,37	9,13	10,07	**4,42**
72	7,69	11,25	11,60	**6,00**
73	29,32	14,24	33,37	**8,85**
74	8,85	8,36	7,69	**1,44**
75	2,93	2,89	4,02	**2,09**
76	3,96	**2,65**	3,63	2,66
77	8,65	10,50	10,13	**8,48**
78	22,65	16,00	25,25	**13,72**
79	**3,74**	10,11	10,44	4,49
80	5,60	5,74	6,03	**5,29**
81	7,41	6,86	9,18	**5,83**
82	8,46	**7,15**	9,66	7,37
83	3,04	3,03	4,57	**2,45**
84	14,63	13,48	16,30	**12,27**
85	18,88	27,64	28,92	**14,91**
86	27,27	22,21	32,84	**19,71**
87	**5,21**	8,35	10,34	6,02
88	15,83	**9,55**	17,85	11,22
89	9,21	7,24	9,60	**6,77**
90	6,58	7,16	8,02	**3,42**
91	13,45	17,55	18,51	**12,12**
92	5,30	6,23	7,80	**3,95**
93	3,26	5,74	5,79	**3,05**
94	2,74	5,89	7,01	**1,60**
95	3,61	3,62	3,69	**2,68**
96	6,32	7,51	8,01	**5,93**
97	10,73	26,37	27,01	**9,33**
98	8,21	**7,68**	13,26	9,68
99	7,32	6,52	9,37	**6,34**
100	2,15	2,91	4,28	**1,98**
101	16,60	16,59	16,74	**16,24**
102	5,33	4,12	5,32	**3,45**
103	18,76	**13,47**	23,51	21,42
104	4,45	3,64	4,70	**3,29**
105	5,70	**4,45**	5,88	5,16
106	21,03	11,92	34,76	**9,53**
107	5,94	6,60	7,96	**3,67**
Min *AARD*	1,17	0,68	1,42	**0,54**
Max *AARD*	35,56	34,84	34,84	**31,33**

Moyenne	9,26	8,80	11,46	**7,45**	R^2	0,9951	0,9958	0,9938	**0,9964**

III.1.4.3.2. Discussion des résultats trouvés

Le tableau III.7 montre que la gamme des valeurs de AARD est entre 0.54% et 35.56%, mais quand nous prendrons les meilleures des AARD mentionné en gras, nous trouverons une gamme des valeurs entre 0.54% et 31.33% avec une moyenne de 6.97%.

L'AARD minimum (0.54%) est obtenu par l'équation (Eq. (17)) pour le système N° 60 ($scCO_2$ - Menadione), et l'AARD maximum (35.56%) est obtenu par l'équation (Eq. (14)) pour le système N° 3 ($scCO_2$ -4,4-diaminodiphenylmethane).

La comparaison globale entre la performance des modèles indique que le modèle (Eq. (17)) a donné la meilleure corrélation, ou la gamme de l'ARRD est entre 0.54% et 31.33% avec une moyenne de 7.45%. Le second modèle qui a la meilleure corrélation est le modèle (Eq. (15)), ou la gamme de l'AARD est entre 0.68% et 34.84% avec une moyenne de 8.80%. Le troisième avec une meilleure corrélation est le modèle (Eq. (14)), ou la gamme de l'AARD est entre 1.17% et 35.56% avec une moyenne de 9.26%. Ces trois modèles ensemble représente une meilleure corrélation de 100% des 107 composés considéré.

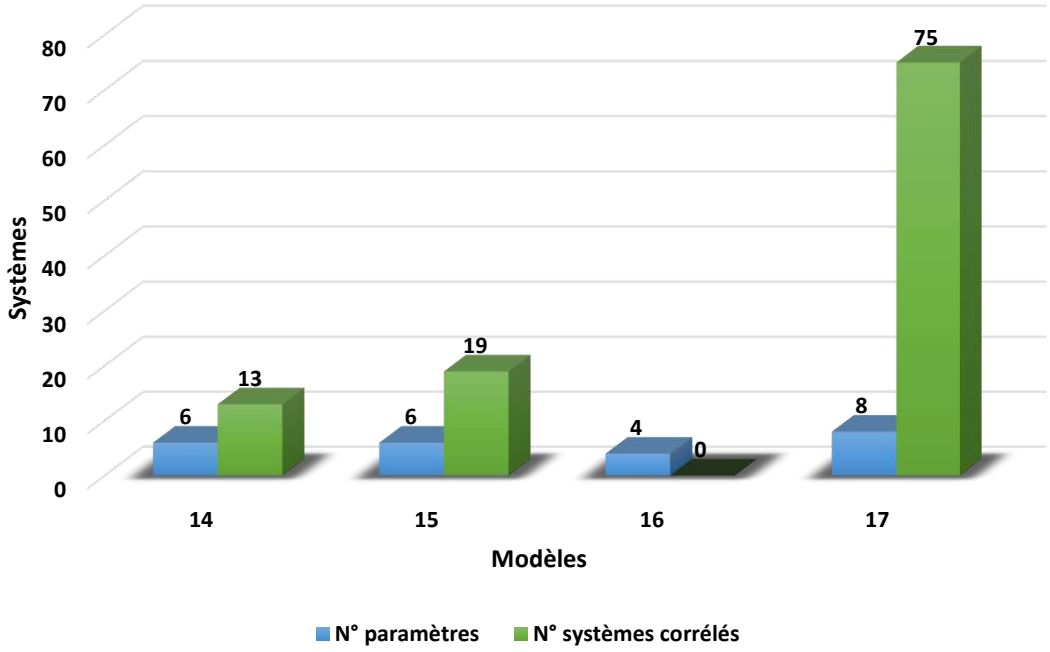


Figure III.3. Le nombre des systèmes corrélé et le nombre des paramètres pour chaque modèle.

III.1.5. Comparaison entre les meilleures modèles

D'après les résultats trouvés précédemment nous allons faire une comparaison entre les modèles qui ont la meilleure corrélation pour les 107 composés considéré. Les modèles choisis sont : le modèle de Si-Moussa et al. (Eq. (9)) sélectionné parmi la littérature (tableau III.2), (Eq. (12)) sélectionné parmi la série 01 (tableau III.4), et (Eq. (17)) sélectionné parmi la série 02 (tableau III.7).

Les résultats de calcul des $AARD$ (%) et R^2 pour les meilleurs modèles sélectionnés sont regrouper dans le tableau III.8, les colonnes représentent les modèles choisis et les lignes représentes les valeurs des $AARD$ (%) pour les 107 composés. Le meilleur résultat de corrélation pour chaque soluté solide est montré dans ce tableau mentionné en gras.

Tableau III.8. Résultats de calcul des AARD pour les meilleurs modèles sélectionnés

N°	(9)	(12)	(17)
1	4,01	6,54	**3.09**
2	2,74	**2,24**	2.29
3	35,76	34,73	**31.33**
4	6,73	**6,39**	6.47
5	1,92	2,02	**1.58**
6	6,72	**5,89**	6.66
7	6,97	6,4	**6.05**
8	3,31	**3**	3.19
9	3,84	3,99	**3.36**
10	16,24	17,09	**15.73**
11	3,05	**2,69**	2.73
12	17,04	21,51	**16.46**
13	7,66	9,64	**6.67**
14	9,67	11,25	**9.37**
15	2,8	4,75	**2.67**
16	**9,16**	10,06	9.18
17	2,9	2,87	**2.62**
18	6,95	7,11	**6.65**
19	11,85	15,32	**10.77**
20	1,42	1,8	**1.01**
21	4,11	4,27	**4.07**
22	4,13	4,07	**3.74**
23	4,12	5,24	**3.54**
24	2,39	3,35	**2.04**
25	10,95	14,55	**9.80**
26	4,8	5,01	**4.92**
27	3,71	**2,8**	2.86
28	15,12	22,7	**10.85**
29	12,68	12,07	**10.16**
30	15,4	**11,83**	14.77
31	7,04	7,78	**6.88**
32	5,55	10,18	**4.90**
33	7,42	7,9	**6.46**
34	3,53	7,46	**3.28**
35	5,01	5,55	**4.33**
36	11,05	12,67	**10.47**
37	6,62	7,75	**6.31**
38	5,05	**3,98**	4.55
39	6,09	9,75	**3.31**
40	20,6	**11,29**	20.48
41	**11,57**	16,49	11.83
42	5,23	5,28	**4.44**
43	15,63	19,77	**13.92**
44	3,71	8,2	**3.64**
45	9,61	10,34	**9.36**
46	11,35	12,33	**8.66**
47	6,22	8,09	**4.17**
48	2,39	2,14	**2.12**
49	5,37	6,85	**5.12**
50	12,2	**7,99**	12.12
51	6,18	8,49	**4.73**
52	3,53	3,17	**2.76**
53	13,96	15,19	**13.38**
54	9,54	9,92	**8.43**
55	4,29	**4,04**	4.31
56	3,56	**3,39**	3.64
57	23,64	**23,27**	25.34
58	**3,42**	3,74	3.43
59	21,1	19,95	**14.53**
60	0,6	0,75	**0.54**
61	22,89	**22,83**	22.86
62	9,46	14,04	**8.92**
63	3,82	4	**3.56**
64	5,21	**5**	5.06
65	12,86	**9,03**	12.95
66	1,75	2,9	**1.73**
67	3,77	4,17	**3.15**
68	17,68	21,23	**17.42**
69	12,16	16,32	**11.58**

70	1,15	1,42	**0.97**
71	4,8	7,33	**4.42**
72	8,25	7,63	**6.00**
73	14,93	29,21	**8.85**
74	6,64	8,58	**1.44**
75	3,12	2,75	**2.09**
76	2,9	3,84	**2.66**
77	**8,47**	8,67	8.48
78	15,57	22,1	**13.72**
79	4,97	**3,69**	4.49
80	5,61	5,71	**5.29**
81	6,67	7,17	**5.83**
82	7,67	8,37	**7.37**
83	2,75	3,04	**2.45**
84	13,86	13,88	**12.27**
85	19,03	18,88	**14.91**
86	21,04	27,27	**19.71**
87	**5,92**	7,77	6.02
88	11,29	15,82	**11.22**
89	7,46	8,67	**6.77**
90	6,57	6,54	**3.42**
91	12,88	15,85	**12.12**
92	6,49	8,93	**3.95**
93	3,82	3,25	**3.05**
94	1,78	2,73	**1.60**
95	3,63	3,6	**2.68**
96	**5,92**	6,31	5.93
97	10,44	26,99	**9.33**
98	9,68	**8,3**	9.68
99	6,7	8,48	**6.34**
100	3,63	2,53	**1.98**
101	16,25	16,83	**16.24**
102	4,61	5,33	**3.45**
103	22,88	**14,07**	21.42
104	3,6	4,6	**3.29**
105	**5,15**	5,8	5.16
106	12,8	21,02	**9.53**
107	5,77	5,93	**3.67**
Min *AARD*	0,6	0,75	**0.54**
Max *AARD*	35,76	34,73	**31.33**
Moyenne	8,35	9,47	**7.45**
R^2	0,9959	0,995	**0.9964**

Le tableau III.8 montre que la gamme des valeurs de AARD est entre 0.54% et 35.76%, mais quand nous prendrons les meilleures des AARD écrit en gras on, nous trouverons une gamme des valeurs entre 0.54% et 31.33 % avec une moyenne de 7.12%.

L'AARD minimum (0.54%) est obtenu par l'équation (Eq. (17)) pour le système N° 60 ($scCO_2$ - Menadione) et l'AARD maximum (35.76%) est obtenu par l'équation (Eq. (9)) pour le système N° 3 ($scCO_2$ -4,4-diaminodiphenylmethane).

La comparaison globale entre la performance des modèles indique que le modèle (Eq. (17)) a donné la meilleure corrélation, ou la gamme de l'ARRD est entre 0.54% et 31.33 % avec une moyenne de 7.45%. Le second modèle qui a la meilleure corrélation est le modèle (Eq. (9)), ou la gamme de l'AARD est entre 0.6% et 35.76% avec une moyenne de 8.35%. Le troisième avec une meilleure corrélation est le modèle (Eq. (12)), ou la gamme de l'AARD est entre 0.73% et 34.73% avec une moyenne de 9.47%. Le modèle (Eq. (17)) a corrélé 80 composés avec un pourcentage de 75%, le modèle (Eq. (9)) a corrélé 8 composés avec un pourcentage de 8% et le modèle de (Eq. (12)) a corrélé 19 composés avec un pourcentage de 17%.

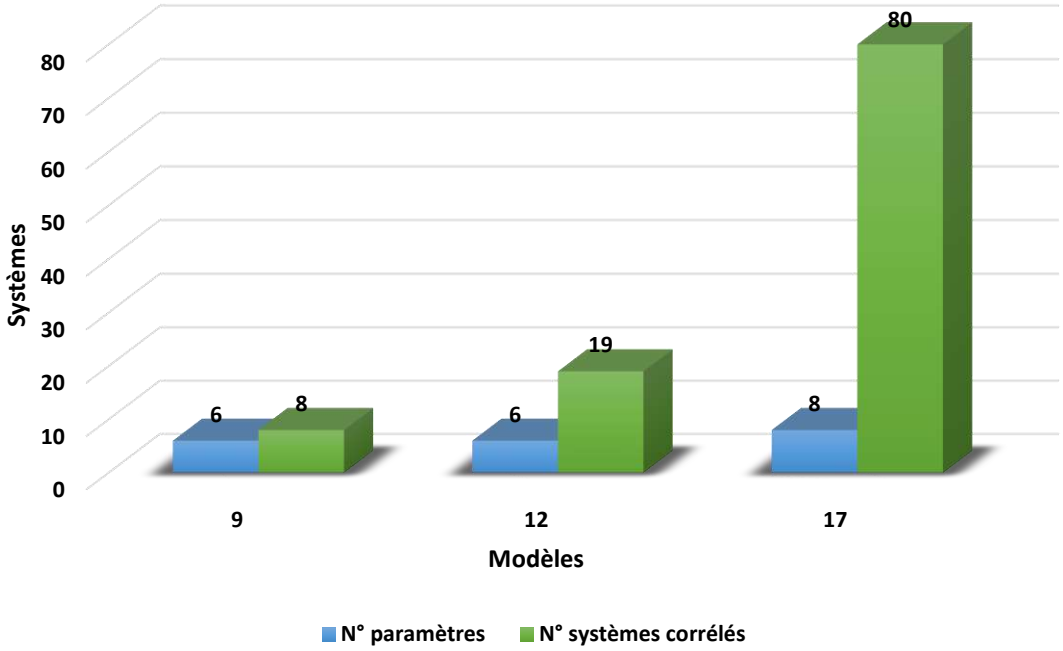


Figure III.4. Le nombre des systèmes corrélé et le nombre des paramètres pour chaque modèle.

La validation et la performance des modèles sélectionnés (Eq. (9)), (Eq. (12)) et (Eq. (17)) sont présentées sous forme des diagrammes de dispersion pour les 3010 points expérimentaux de la base de données. Les Figures III.4, III.5 et III.6 montre les diagrammes de dispersion des modèles sélectionnés. Ces diagrammes sont générés en utilisant la fonction ***postreg*** de MATLAB, dans laquelle la solubilité calculée en échelle logarithmique est tracée par rapport à la solubilité expérimentale. Le diagramme de dispersion montre la dispersion du nuage de points de l'ensemble de données autour de la première bissectrice ($log\left(y_2^{cal}\right) = log\left(y_2^{exp}\right)$), le degré de dispersion peut être évalué à partir de l'équation pour le meilleur ajustement linéaire ($log\left(y_2^{cal}\right) = \alpha . log\left(y_2^{exp}\right) + \beta$), où α est la pente et β est l'interception, qui avec le coefficient de corrélation R sont utilisés comme vecteur de validation.

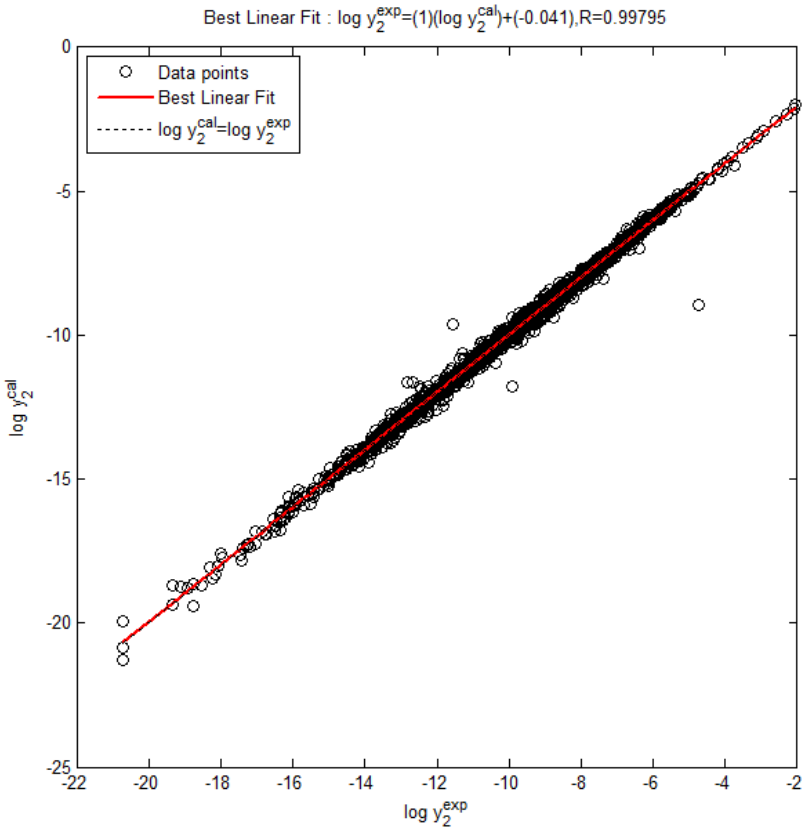


Figure III.5. Diagramme de dispersions de modèle Si-Moussa et al. (Eq. (9))

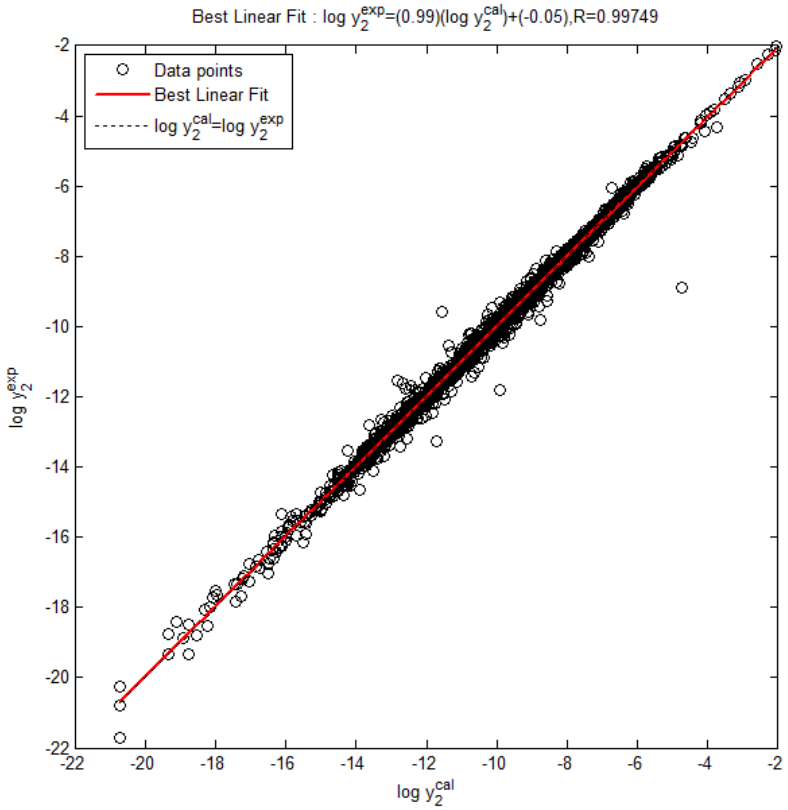


Figure III.6. Diagramme de dispersions de modèle proposé, série 01 (Eq. (12))

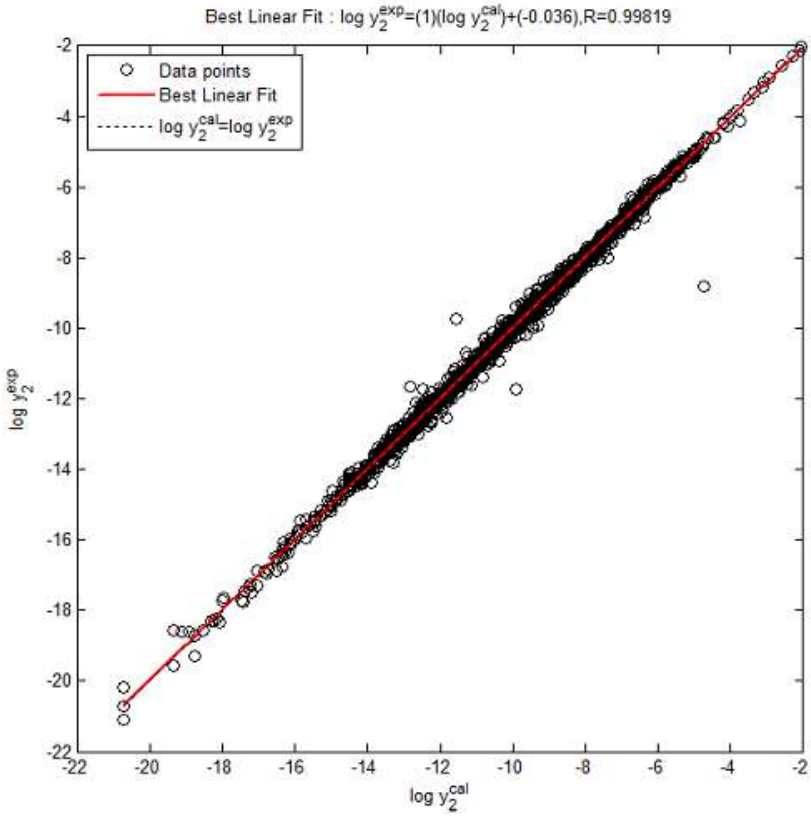


Figure III.7. Diagramme de dispersions de modèle proposé, série 02 (Eq. (17))

La performance idéale est obtenue lorsque [$\alpha = 1, \beta = 0, R = 1$]. Les valeurs de α, β et R pour les trois modèles sélectionnés sont résumées dans le tableau III.9. Selon la valeur α près de 1, la valeur β proche de 0 et R proche de 1, clairement le modèle (Eq. (17)) a le meilleur vecteur de validation et la meilleure performance de corrélation globale.

Tableau III.9. Le vecteur de validation (α [pente], β [ordonnée a l'origine] et R [coefficient de corrélation]) calculé pour les trois modèles sélectionnés.

Vecteur de validation	Modèles		
	(9)	**(12)**	**(17)**
α	0,9959	0.9950	0.9964
β	-0.0406	-0.0500	-0.0359
R	0.9979	0.9975	0.9982

III.2.6. Validation

Dans cette phase on a choisi le système (Curcumin – $scCO_2$) qui n'été pas sélectionnés parmi les 107 composés introduit dans l'étape de la modélisation pour vérifier la validation des modèles sélectionner. Le tableau III.10 représente les valeurs expérimentales et calculés pour le système sélectionné.

La curcumine est un pigment phénolique naturel extrait du curcuma et possède une bonne activité pharmacologique en anti-inflammatoire, antioxydant, hypolipidémie et antitumoral. Il a été fréquemment utilisé dans des microsphères chargées de médicaments

Tableau III.10. Les valeurs de ($T(°K)$, $P(MPa)$, $\rho_1(kg/m^3)$ et y^{exp}) sont mesuré par (Zhan et al., 2017) et les valeurs (y_{M25}^{cal}, y_{M35}^{cal}, y_{M43}^{cal}) sont calculés dans cette étude.

$T(°K)$	$P(MPa)$	$\rho_1(kg/m^3)$	$y^{exp}.10^8$	$y_{M9}^{cal}.10^8$	$y_{M12}^{cal}.10^8$	$y_{M17}^{cal}.10^8$
308,15	8	419,09	2,07	1,96	1,87	2,07
308,15	10	712,81	2,51	2,48	2,54	2,42
308,15	13	785,7	2,68	2,74	2,79	2,73
308,15	16	827,17	2,85	2,91	2,97	2,91
308,15	20	867,72	3,11	3,10	3,14	3,08
308,15	8	241,05	1,82	1,85	1,79	1,82
318,15	10	498,25	2,31	2,25	2,33	2,29
318,15	13	693,65	2,69	2,82	2,86	2,77
318,15	16	759,98	3,12	3,11	3,11	3,13
318,15	20	812,69	3,52	3,40	3,34	3,47
328,15	8	203,64	1,86	1,83	1,86	1,86
328,15	10	325,07	2,03	2,11	2,18	2,03
328,15	13	571,33	2,62	2,72	2,76	2,68
328,15	16	681,12	3,25	3,20	3,11	3,17
328,15	20	754,61	3,7	3,63	3,41	3,71

La représentation graphique entre la solubilité expérimentale en fraction molaire et celle interpolée en utilisant une nouvelle base de données par les trois meilleurs modèles en fonction de l'une de variable d'entrée (pression), permet d'une part de valider le modèle 17 est d'autre part faire une comparaison des valeurs en terme d'AARD moyenne et R^2. Les figures III.9-III.10 montrent que les points pleins estimés par le modèle 17 sont localisés au centre des objets vide (points calculés). Ces résultats montrent que le modèle 17 suit parfaitement la l'ure expérimentale, ce qui confirme l'efficacité irréfutable de ce dernier lors de la prédiction de la solubilité en fraction molaire en cas d'interpolation.

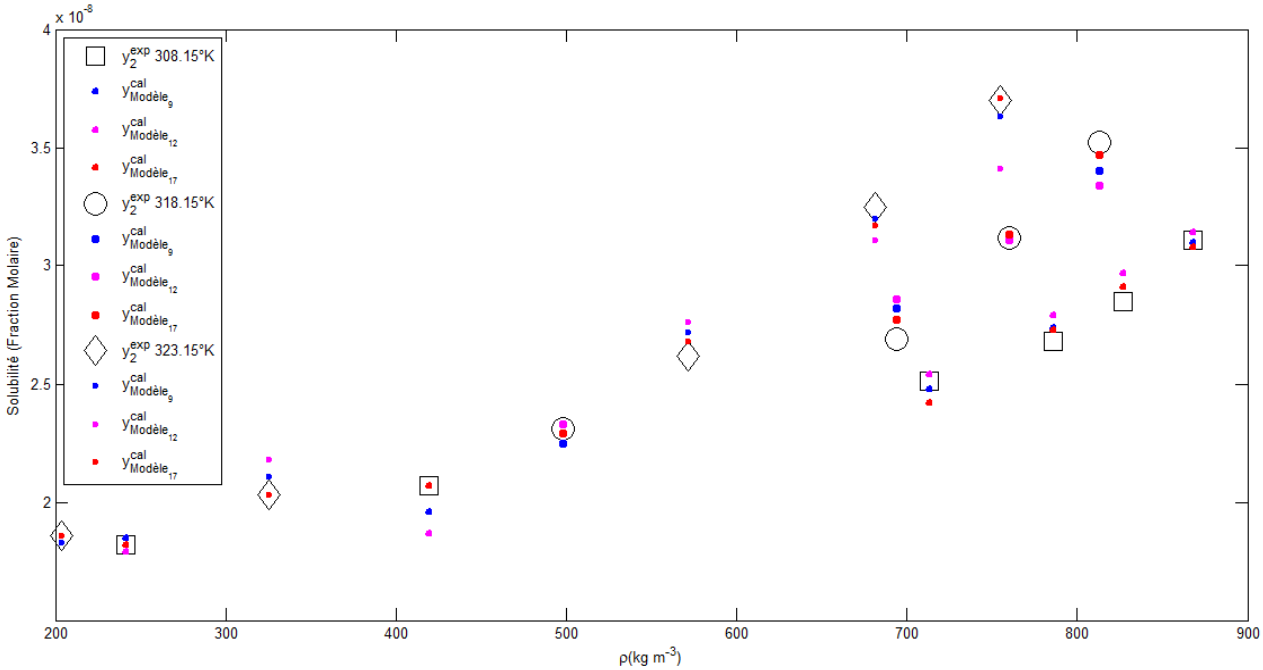


Figure III.8. Solubilité expérimentale de la curcumine in $scCO_2$ et prédire par les modèles en fonction de la densité du fluide supercritique (graphe 2D)

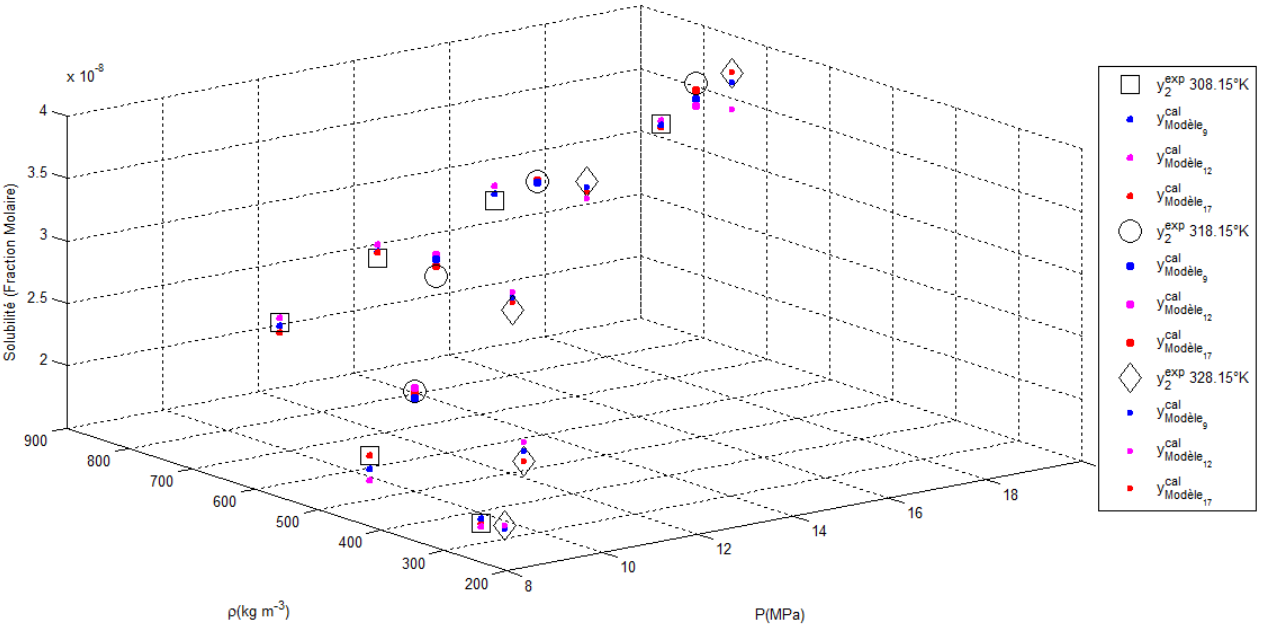


Figure III.9. Solubilité expérimentale de la curcumine in $scCO_2$ et prédire par les modèles en fonction de la densité du fluide supercritique et la pression (graphe 3D)

III.2.7. Conclusion

L'objectif de cette partie est de développer une méthodologie de calcul de la solubilité de 107 composés solides dans les fluides supercritiques. La majorité de ces composées sont des médicaments utilisés pour des différentes thérapies (antibiotique, antifongiques, anti-inflammatoires, antiépileptiques, anticholestérols, antioxydants ...etc.).

Trois séries des modèles empiriques et semi-empiriques sont :

1. Modèles basés sur la densité du solvant.
2. Modèles contient des combinaisons entre les variables intensives (T, P, ρ_1) et la solubilité en fraction molaire (y_2).
3. Modèles contient des modifications par rapport aux modèles trouvés dans la littérature.

A partir de chaque série un meilleur modèle a été sélectionné, en suite une comparaison entre ces trois modèles a été faite afin de choisir le modèle optimal en basant sur l'AARD moyenne minimale, le coefficient de détermination et le nombre des systèmes corrélé par chaque modèle. En fin une phase de validation a été réalisée en utilisant un autre système qui n'est pas utilisé dans la construction des modèles. Les résultats montrent que le modèle 17 suit parfaitement la l'ure expérimentale, ce qui confirme l'efficacité irréfutable de ce dernier lors de la prédiction de la solubilité en fraction molaire en cas d'interpolation.

CONTRIBUTION 02 :
Modélisation neuronale de la solubilité des solutés solides dans le CO2 supercritique.

III.2.1. Introduction

Les propriétés thermodynamiques du CO_2 dans les conditions supercritiques et la facilité de récupération du CO_2 sont responsables de l'utilisation répandue du solvant CO_2 pour séparer différents types de substances (Asiabi et al., 2015; Henley et al., 2011).

Étant donné que les corrélations basées sur la densité disponibles et les équations d'états ne sont pas suffisamment précis pour prédire les données de solubilité de différents solutés dans le CO_2 supercritique, dans cette contribution nous avons construire des modèles basés sur le réseau de neurone artificiel (RNA) pour estimer les valeurs de solubilités de 107 solutés solides dans le dioxyde de carbone supercritique sur de larges fissures de température et de pression.

Dans cette contribution, nous comparons l'effet de l'utilisation de descripteurs moléculaires (la méthode QSAR) comme entrés pour prédire la solubilité des médicaments solides dans le dioxyde de carbone supercritique. Nous construisons 4 architecture de réseau neuronal :

1. Un RNA simple, en utilisant les variables intensives, la masse moléculaire et le numéro de système comme entrée et la solubilité en tant que sortie.
2. Un RNA-2D-2D, en utilisant les variables intensives plus des descripteurs sélectionnés à partir d'une liste qui contiennent des structures de molécules 2D et la base de données contient uniquement les descripteurs 2D comme entrées et la solubilité en tant que sortie.
3. Un RNA-2D-All, en utilisant les variables intensives plus des descripteurs sélectionnés à partir d'une liste contiennent des structures de molécules 2D et la base de données contient des descripteurs 2D et 3D comme entrées et la solubilité en tant que sortie.
4. Un RNA-3D-3D, en utilisant les variables intensives plus des descripteurs sélectionnés à partir d'une liste contiennent des structures de molécules 3D et la base de données contient des descripteurs 3D comme entrées et la solubilité en tant que sortie.

III.2.2. La méthode QSAR

III.2.2.1. Introduction

Les produits chimiques sont des éléments essentiels de la civilisation humaine. Les applications des produits chimiques à l'ère moderne couvrent une large gamme, des environnements industriels aux ménages. L'objectif estimé d'un chimiste réside dans le développement de produits chimiques avec le profil désiré de leur meilleure manifestation comportementale, c'est-à-dire activité / propriété. En chimie, le chimiste veut améliorer l'efficacité de la molécule tout en réduisant les effets toxiques exercés. Maintenant, suivant l'axiome général, le développement d'un nouveau produit chimique peut être réalisé soit en concevant une nouvelle entité complète, soit en modifiant une

ancienne. Dans les deux cas, les chimistes doivent posséder suffisamment de connaissances sur le produit chimique et son potentiel d'interaction avec le système biologique. Il est évident que l'activité biologique (y compris la toxicité) et les propriétés de tout produit chimique (p. Ex., Médicaments, produits pharmaceutiques et cancérogènes) dépendent de son interaction avec le système biologique concerné. Par conséquent, l'objectif principal d'un concepteur chimique réside dans l'établissement d'une explication rationnelle du mécanisme d'action du produit chimique, ce qui peut conduire à une dérivation théorique et donc à l'adaptation de sa structure. Parfois, le manque d'action appropriée pour la drogue limite la dérivation hypothétique. Afin d'établir une base théorique appropriée pour l'action des produits chimiques, il faudrait avoir un aperçu profond de la chimie fondamentale du système qui contrôle les comportements physico-chimiques et structurels (Hansch et al., 1995).

III.2.2.2. Relations activité - structure

III.2.2.2.1. Idéologie

La chimie a été une science centrale essentielle qui corrèle et explore différents problèmes de discipline scientifique (Balaban and Klein, 2006). Compte tenu du nombre important de produits chimiques qui influent sur notre vie du point de vue de l'application industrielle, de la recherche en laboratoire, et la consommation des ménages, il sera intéressant de développer une base qui donne une raison appropriée pour expliquer le comportement des produits chimiques. Une fois que cette base est développée, elle nous permettra également de modifier le comportement de ces produits chimiques en incorporant des changements structurels raisonnables. La notion relation activité-structure (SAR) tente d'établir une relation entre les différentes caractéristiques comportementales des composés chimiques (à savoir, activité / propriété / toxicité et informations chimiques dérivables des structures chimiques). Un tel processus implique une connaissance significative de la chimie et de la physique (définition des composés), des mathématiques (permettant une analyse de modélisation) et de la biologie (donner des raisons pour le profil d'activité biologique des produits chimiques) pour développer une base rationnelle. En d'autres termes, l'étude SAR offre une option pour construire une équation mathématique pour un ensemble de produits chimiques pour leur comportement spécifique d'activité / propriété / toxicité en utilisant des informations sur leur structure chimique. On peut noter que, étant donné que nous parlons d'équations mathématiques, cela implique des nombres ou des quantités. Si l'information comportementale des produits chimiques est exprimée en termes qualitatifs, l'étude est appelée SAR qualitative, alors que la présence de nombres quantitatifs représentant l'activité des produits chimiques définit le processus comme une étude

QSAR (Todeschini et al., 2009). Cependant, les études qualitatives et quantitatives sont appelées collectivement QSAR.

La dénomination est également influencée par le point final modélisé, et le processus peut être défini en conséquence comme QSAR / QSPR / QSTR considérant que la réponse est l'activité, la propriété ou la toxicité.

L'axiome central de la modélisation QSAR réside dans la présentation de la réponse chimique en termes d'attributs moléculaires. Ces caractéristiques chimiques peuvent représenter des paramètres théoriquement déterminés ainsi que des variables expérimentales (à savoir différents types de propriétés physico-chimiques). Par conséquent, on peut observer que toute caractéristique qui contient des informations chimiques importantes et fondamentales peut être utilisée comme variable prédictive ou des descripteurs. Mathématiquement, la réponse d'un produit chimique peut être représentée comme fonction des attributs structurels et de la propriété :

$$Réponse = f(caractéristiques\ chimiques, propriété) \qquad \text{III.3}$$

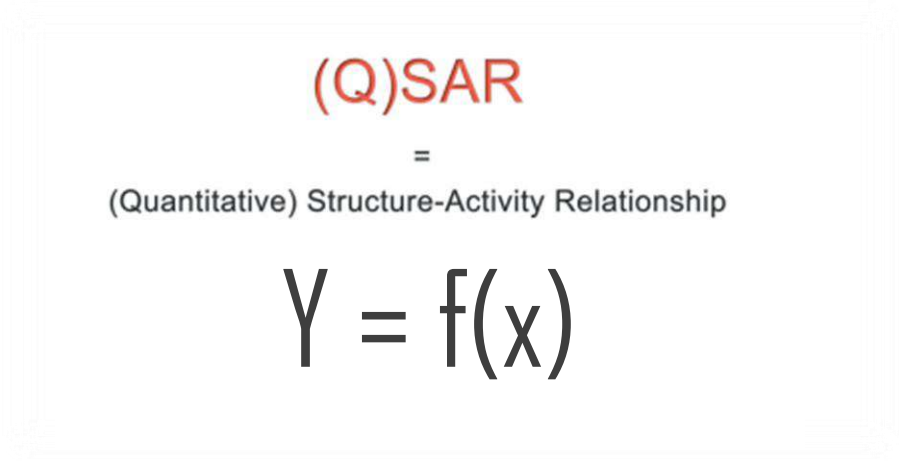


Figure III.10. La méthode QSAR visent à trouver la corrélation entre les propriétés structurelles des produits chimiques et leur activité.

III.2.2.2.2. Objectifs du développement du modèle QSAR

Maintenant, une interrogation pourrait être soulevée concernant les possibilités et les opportunités réalisables à l'aide de ces modèles de corrélation. L'étude QSAR offre une option pour le développement d'une stratégie rationnelle vers la conception d'une nouvelle molécule possédant les caractéristiques souhaitées. La manifestation comportementale des produits chimiques dépend en grande partie de divers paramètres structurels et physico-chimiques. Même des modifications subtiles de la structure peuvent entraîner un changement significatif de son activité pharmacologique /

toxicologique. Par conséquent, il est évident que la chimie des composés est cruciale pour déterminer leur comportement. L'utilisation de descripteurs appropriés ayant une fonction de diagnostic appropriée peut permettre à un chimiste de construire des informations convaincantes sur le mode d'action des produits chimiques, puis de les affiner.

QSAR est une étude théorique pour construire des conclusions rationnelles et réduit le fardeau d'expérimentation pratique ainsi que le temps. Il utilise comparativement moins de composés pour établir une relation mathématique qui peut être employée pour caractériser un grand nombre de produits chimiques appartenant à une classe identique ou différente selon la nature du modèle (local ou global). Cette caractéristique est particulièrement importante pour évaluer l'activité biologique et / ou la toxicité des produits chimiques qui impliquent aux animaux de laboratoire. La technique QSAR peut conduire à l'utilisation d'une expérimentation moins animale (puisqu'elle utilise comparativement moins de grandeur de données expérimentales et peut donner la priorité aux composés pour une autre expérimentation), ce qui constitue une alternative précieuse aux études sur les animaux. Dans la section suivante, nous expliquerons brièvement les principaux objectifs du développement de modèles prédictifs quantitatifs (Todeschini et al., 2009; Tute MS, 1990).

III.2.2.2.3. Prédiction de l'activité / propriété / toxicité

L'objectif principal d'un modèle QSAR est de prévoir la valeur de la réponse (activité / propriété / toxicité) des nouveaux produits chimiques. La prédiction est faite pour les composés qui ne sont pas utilisés pour développer le modèle (test ou validation). Maintenant, la prédiction pour les composés inconnus peut être obtenue par une extrapolation des données, et il est également évident que, dans cette méthode, il existe un certain degré d'incertitude. Dans l'analyse QSAR, les produits chimiques de formation sont généralement utilisés pour définir un domaine chimique et tout composé nouveau, non testé ou hypothétique situé dans ce domaine peut être prédit par une équation QSAR donnée. Ainsi, l'étude QSAR fournit un bon soutien raisonnable à un chimiste dans le choix de molécules avec les propriétés souhaitées avant leur synthèse.

III.2.2.3. Développement historique de QSAR

Le développement de la méthodologie QSAR n'a pas été une découverte soudaine. Plutôt, il a été une évolution progressive d'une idéologie. L'idéologie est venue comme une combinaison résultante de concepts fondamentaux de la chimie et des mathématiques, qui a ensuite été extrapolé à d'autres branches interdisciplinaires des sciences naturelles. Dans le passé lointain, la notion est

venue de l'approche de la corrélation du comportement des composés chimiques avec leurs la nature, et plus tard, la nature de la chimie a été explorée en termes d'information chimique.

Bien que l'expérimentation QSAR d'aujourd'hui implique de nombreuses étapes de prétraitement, ainsi que des critères de développement, y compris des spécifications de fiabilité statistique, l'objectif principal reste axé sur sa capacité prédictive inhérente et sa réceptivité à l'explication du mécanisme impliqué. Considérons le travail du grand philosophe Galileo Galilei, qui a pensé à une relation quantitative entre les êtres humains et leur environnement au seizième siècle. Bien que nous parlions ici de chimie quantitative, il est surprenant que le concept de corrélation chimique ait été conçu avant toute notion de structure moléculaire.

Le développement des fondamentaux des structures moléculaires peut être attribué à la contribution décisive de Kekule, Couper et Crum-Brown pendant la période de 1858-1870. Cependant, c'était Blake qui a noté une action similaire provoquée par des sels de bases isomorphes en 1841 (Charton, 2008). Dix ans plus tard, Horsford et Baird (Horsford and Baird, 1851) ont observé une relation entre le goût des composés et leur composition et quelques années après, en 1854, Pelikan (Pelikan EV, 1855) a noté que la composition chimique influe sur l'effet toxique des composés. En 1858, Borodin est venu à la conclusion qu'une propriété toxicologique est étroitement liée à la composition chimique des composés (Borodin A, 1858). Un type similaire de comportement sur les organismes a été observé par des produits chimiques possédant les mêmes éléments ou participant à une réaction chimique similaire. En 1863, Cros (Cros 1863) a observé une relation entre la solubilité dans l'eau et la toxicité des alcools primaires. Par conséquent, on constate que la notion précoce de modélisation QSAR est associée au domaine de la toxicologie. À cette époque, Mendeleev a observé une relation entre les éléments en utilisant leur poids atomique et a développé une table périodique d'éléments (Tute MS, 1990). L'utilisation du poids atomique pour développer la « règle de huit » par Mendeleev peut être visualisée comme l'une des plus anciennes méthodes d'utilisation d'un «paramètre» dans une étude relationnelle impliquant la chimie. La première proposition pour l'existence d'une relation mathématique entre la structure chimique et l'activité a été menée inévitablement par Cru-Brown et Fraser en 1868 (Brown and Fraser, 1868) en montrant l'activité physiologique (φ) comme une fonction mathématique de la constitution chimique (C), équation (III.2). On peut remarquer que le terme *constitution chimique* ne représente que la composition élémentaire à cette période puisque les concepts de structure moléculaire n'ont pas été établis. Ils ont montré qu'une série de dérivés de la strychnine possédant une activité paralytique musculaire similaire à Curare peut être préparée en faisant varier le substituant quaternaire.

$$\varphi = f(C) \qquad \text{III.4}$$

En 1869, Richardson (Richardson, 1869) a observé le poids moléculaire de l'alcool primaire contrôlant proportionnellement son effet narcotique. Reynolds en 1877 a effectué une autre étude préliminaire et a rapporté un effet de constitution chimique sur l'activité physiologique des produits chimiques. Quelques années plus tard, en 1893, Richet (Richet, 1893) a donné plus de confiance à la proposition de Crum-Brown et Fraser en montrant que la toxicité de l'éther, de l'alcool et des cétones est inversement affectée par leur mesure de solubilité.

Au début du siècle, Meyer (Meyer, 1899), Overton (Overton, 1899) et Baum (Baum, 1899) ont ajouté de la lumière à l'équation (III.2) en employant le coefficient de partage huile d'olive / eau pour corréler les puissances des substances narcotiques. Overton a effectué une recherche approfondie dans cette perspective et a observé une augmentation systémique de la puissance des stupéfiants dans les têtards lorsque la longueur de la chaîne en groupe a été augmentée et a conclu le partage des composés dans les cellules lipidiques pour être responsable de cette action. De plus, tout en expérimentant la morphine, il a observé des effets toxiques variables sur les humains et les têtards et a supposé une différence dans la structure des protéines dans le corps responsable de la formation du complexe. Une autre essence de la recherche a été ajoutée par Traube en 1904 (Traube, 1904), qui a observé que la narcose était linéairement liée à la tension superficielle des produits chimiques et plus tard, en 1912, Seidell (Seidell, 1912) a considéré à la fois le coefficient de partage et la mesure de solubilité pour établir corrélation. Une impulsion supplémentaire à cette ligne de pensée a été instituée en 1939 par Ferguson (Ferguson, 1939), qui a fusionné le concept de thermodynamique avec la narcose et le coefficient de partage et a déclaré que la saturation relative de la substance dans la phase appliquée est liée à l'activité narcotique. Au début des années 1940, Albert et al. (Albert et al., 1941) et Bell et Roblin (Bell and Roblin, 1942) ont exploré l'influence de l'ionisation des bases et des acides faibles vers leur activité bactériostatique. On notera que, à ce moment-là, des recherches pionnières se déroulaient dans le domaine de la chimie organique physique pour dériver l'impact des substituants de composés organiques sur le comportement des produits chimiques. L'étude décisive de Hammett (Hammett, 1935) sur le taux de réaction relatif des dérivés d'acide benzoïque méta et par-substitué a conduit au développement du célèbre substituant électronique constant Hammett sigma (σ), comme le montrent les équations (III.3) et (III.4).

$$\log(k_X/k_H) = \rho\sigma_X \quad \text{III.5}$$

Or,

$$\log(K_X/K_H) = \rho\sigma_X \quad \text{III.6}$$

Où k_H et $k_X$sont les termes de constante de taux respectifs, K_H et K_X sont respectifs les termes constants d'équilibre pour les dérivés d'acide benzoïque non substitués et substitués, ρ est la constante de réaction, et σ_X est la constante électronique Hammett du substituant X. Il s'agissait du premier modèle linéaire d'énergie libre (LFER) établi dans le domaine de QSAR depuis les termes d'ionisation constante utilisés pour décrire σ peuvent être liés à l'équation d'énergie libre (III.5), où G est le changement d'énergie libre de Gibbs, R est la constante de gaz idéale, et T est la température (en Kelvin) :

$$\Delta G^0 = -RTlnK \quad \text{III.7}$$

Une recherche approfondie sur cet aspect a permis d'élaborer un formalisme "sigma-rho" à cette époque.

Plus tard, Taft (Taft, 1952) a introduit le premier paramètre stérique E_s et en utilisant l'analyse LFER sur les taux d'hydrolyse catalysée par des bases et des esters aliphatiques et des antécédents séparés de QSAR et des développements historiques 33 des effets polaires, stériques et de résonance. Au début des années 1960, Corwin Hansch, également connu sous le nom de « Père du QSAR moderne », a donné une nouvelle impulsion à la recherche QSAR en utilisant les constantes de Hammett et les paramètres d'hydrophobicité pour développer des modèles de corrélation sur les régulateurs de croissance des plantes (Hansch et al., 1962). Une mesure relative d'hydrophobicité du paramètre substituant (π) a été introduite. Plus tard, la célèbre équation linéaire de Hansch a été développée par Hansch et Fujita (Hansch and Fujita, 1964) en combinant des termes constants hydrophobes avec le sigma Hammett (σ). L'équation principale est présentée dans l'équation (III.7), alors que l'on peut noter que cette forme de l'équation a été modifiée et prolongée plusieurs fois, y compris le développement de sa forme parabolique :

$$\log {}^{1}/_{C} = k_1\pi + k_2\sigma + k_3E_s + k_4 \quad \text{III.8}$$

Où k_1, k_2 et k_3 sont les termes du coefficient, tandis que k_4 présente une constante. Parallèlement, une autre approche significative de la modélisation QSAR a été instituée par Free et Wilson (Free and Wilson, 1964) sur une série congénère de composés. Dans cette approche, l'activité biologique est exprimée comme la contribution additionnée de la fraction parentale et de chaque caractéristique structurelle présente. L'approche Free-Wilson peut être définie dans l'équation (III.9),

où μ est la contribution de la fraction parent et a_i est la contribution des caractéristiques structurelles individuelles, x_i est une variable qui représente la présence ($x_i = 1$) ou l'absence ($x_i = 0$) d'une caractéristique structurelle spécifique :

$$BA = \sum a_i x_i + \mu \qquad \text{III.10}$$

Les limites de cette équation ont été abordées par Fujita et Ban (Fujita and Ban, 1971), qui ont développé une équation similaire simplement en utilisant l'activité logarithmique pour porter le terme d'activité au même niveau que d'autres terminologies énergétiques gratuites.

Par conséquent, il est évident que le parcours du QSAR a commencé avec le concept d'obtenir une corrélation entre l'activité des produits chimiques et la composition chimique, ce qui a conduit à l'exploration de la chimie des composés. Après quelques hypothèses de base au stade initial, les chercheurs ont eu des équations mathématiques portant des relations thermodynamiques qui peuvent être considérées comme des propriétés chimiques. Au fur et à mesure que le temps progresse, des paramètres indépendants dans l'équation QSAR ont également été modifiés et des mesures structurelles (à savoir constantes de substitution, occurrence de substituant, etc.) sont apparues. Maintenant, il convient de noter que, avec différentes approches et algorithmes, la voie de QSAR a été pavée par la découverte et le développement de diverses variables indépendantes connues sous le nom de descripteurs. Au vingtième siècle, la recherche et la connaissance de la structure chimique augmentaient, et les scientifiques essayaient de trouver un moyen de créer une représentation numérique des structures. La théorie des graphes chimiques, bien qu'un ancien concept, est devenue un outil inévitable pour cette quantification des structures chimiques, et en 1947, les premiers descripteurs théoriquement dérivés (l'indice Wiener (Wiener, 1947) et l'indice Platt (Platt, 1947)) ont été signalés. L'utilisation de descripteurs théoriques dérivés de la structure a acheté une nouvelle compréhension des bases du QSAR pour les applications dans les sciences pharmaceutiques et l'évaluation des risques en recherche dans le paradigme QSAR. Le concept de la théorie des graphes et de la topologie chimique a été considérablement exploité dans les années 1970 et 1980 pour résoudre différents problèmes de QSAR. La méthode de différence topologique minimale (MTD) de Simon (Simon, 1974) et la connectivité moléculaire de Hall et Kier (Hall and Kier, 1977) peuvent être citées à cet égard, bien que ce chapitre ne mentionne pas toutes les contributions remarquables de différents chercheurs à ce moment-là dans le monde. Tout au long de ce temps, les études QSAR étaient principalement unidimensionnelles (1D) et bidimensionnelles (2D). À la fin des années 1980, les caractéristiques moléculaires ont été caractérisées en fonction de leur alignement dans l'espace.

L'énergie d'interaction entre une molécule et des sondes dans l'espace 3D connu sous le nom de champ d'interaction moléculaire (MIF) a été utilisée pour déterminer la caractéristique chimique des molécules. Ce domaine d'études a été mis au point par Goodford, avec sa méthode GRID (Goodford, 1985), et Cramer et al. (Cramer et al., 1988) avec ses formalismes comparatifs d'analyse de champ moléculaire (CoMFA). Certaines autres méthodes et descripteurs basés sur le FIM incluent l'analyse comparée des indices de similarité moléculaire (CoMSIA), les descripteurs G-WHIM, la méthode Compass, l'analyse de terrain Voronoi, les descripteurs GRIND et l'approche VolSurf. En suivant les formalismes tridimensionnels (3D) -QSAR, d'autres approches dimensionnelles supérieures impliquant l'interaction du récepteur-ligand, ainsi que certaines techniques plus récentes telles que le QSAR basé sur l'hologramme (HQSAR), le QSAR binaire et le QSAR inverse ont évolué. Il n'est pas surprenant de dire que de nouveaux concepts dans le paradigme QSAR ainsi que la modification de ceux existants sont encore en cours.

III.2.2.4. Applications de QSAR, (Roy et al., 2015)

Différents types de produits chimiques influent sur une grande partie de l'activité humaine allant des expériences de laboratoire aux processus industriels, y compris les applications domestiques. Les attributs chimiques facilitent la conception de meilleurs analogues par l'optimisation de la réponse souhaitée et / ou la réduction des résultats dangereux, le cas échéant. Le QSAR présente une idéologie pour développer une corrélation mathématique entre la réponse et l'information des objets employés (c'est-à-dire des produits chimiques) ici. Par conséquent, toute réponse appropriée suscitée par des produits chimiques peut être modélisée pour développer une base théorique pour ce processus. Compte tenu d'une vaste zone de produits chimiques fonctionnels (p. Ex., Produits pharmaceutiques, produits agrochimiques, arômes, parfums, réactifs analytiques, solvants et produits chimiques ménagers), le domaine d'application des techniques QSAR sera également significativement important et, dans la pratique, il sera difficile de mentionner Tous les processus où QSAR peut être d'une grande importance. Nous savons déjà que, du point de vue global de la nature de la modélisation des données de réponse, le domaine d'application de l'analyse QSAR peut être visualisé en termes de modélisation d'une activité biologique, d'une propriété et de points de toxicité.

III.2.3. Logiciels et interfaces utilisés

III.2.3.1. DRAGON 7.0 (Todeschini et al., 2009)

Dragon est l'application la plus utilisée au monde pour le calcul des descripteurs moléculaires. Sa nouvelle version, Dragon 7.0, offre une interface utilisateur améliorée, de nouveaux descripteurs et des fonctionnalités supplémentaires telles que le calcul des empreintes digitales et le support des structures déconnectées.

Descripteurs moléculaires : Dragon calcule 5270 descripteurs moléculaires couvrant la plupart des approches théoriques. La liste des descripteurs comprend les types d'atomes les plus simples, les groupes fonctionnels et le nombre de fragments, les descripteurs topologiques et géométriques, les descripteurs tridimensionnels, mais aussi plusieurs estimations de propriétés (telles que logP) et des alertes semblables à des médicaments (comme le Lipinski alerte). Le large éventail d'approches et de théories pour le calcul des descripteurs ainsi que l'exactitude et la précision de leur mise en œuvre sont assurés par la supervision scientifique du groupe de recherche Milano Chemometrics et QSAR du prof. Roberto Todeschini, auteur avec dr. Viviana Consonni de Molecular Descriptors for Chemoinformatics (la référence la plus complète pour la théorie des descripteurs).

Interface utilisateur et calcul de lot : Dragon offre une interface utilisateur graphique conviviale et intuitive (sur les plates-formes Windows et Linux) et une interface de ligne de commande, utile pour le traitement par lots de grands ensembles de données.

Les empreintes digitales et les fragments moléculaires : Dragon fournit maintenant également le calcul d'empreintes moléculaires hachées, entièrement personnalisable avec plusieurs paramètres, ainsi que la génération de tous les fragments moléculaires utilisés dans la procédure d'empreintes digitales.

Outils d'analyse : l'interface utilisateur graphique fournit des outils avancés pour analyser les descripteurs calculés (statistiques uni-variées étendues, corrélation par paire, analyse des composantes principales) et la possibilité d'importer des variables définies par l'utilisateur (comme les valeurs expérimentales disponibles) et effectuer une analyse sur l'ensemble fusionné des descripteurs calculés et des variables de l'utilisateur.

Structures déconnectées : à partir de la version 7.0, Dragon permet le calcul de descripteurs aussi sur des molécules à structure déconnectée (par exemple sels, liquides ioniques), fournissant des approches théoriques différentes pour étendre des algorithmes de descripteurs sur de telles structures.

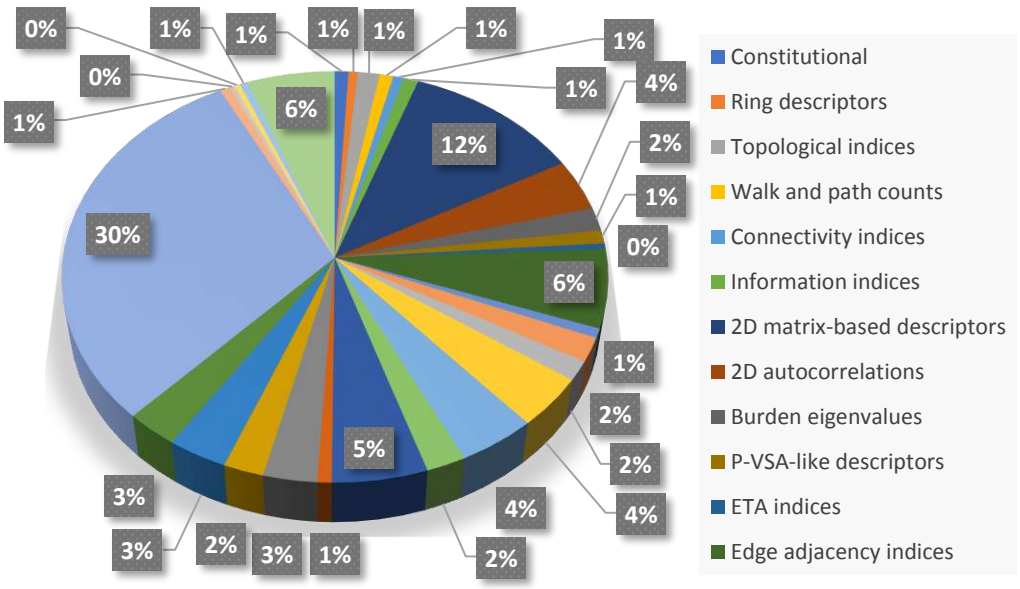


Figure III.11. Un cercle représente le pourcentage de chaque type des descripteurs dans la base de données de logiciel DRAGON 7.0

III.2.3.2. Algorithme génétique

Un algorithme génétique (GA) est une méthode heuristique de recherche qui imite le processus de la sélection naturelle. Lorsque la recherche exhaustive n'est pas réalisable, des méthodes heuristiques sont utilisées pour accélérer le processus de recherche d'une solution satisfaisante. Les algorithmes génétiques appartiennent à la plus grande classe d'algorithmes évolutifs (EA), qui génèrent des solutions aux problèmes d'optimisation en utilisant des techniques inspirées de l'évolution naturelle, telles que l'héritage, le croisement, la mutation et la sélection.

L'évolution généralement démarre d'une population d'individus générés au hasard, et est un processus itératif, avec la population dans chaque itération est connu comme une génération. Dans chaque génération, l'aptitude de chaque individu dans la population est évaluée, La condition

physique est habituellement la valeur de la fonction objective dans le problème d'optimisation étant résolu. Les individus les plus aptes sont stochastiquement sélectionnés parmi la population actuelle, et le génome de chaque individu est modifié (recombiné et éventuellement muté au hasard) pour former une nouvelle génération. La nouvelle génération de solutions candidates est ensuite utilisée dans la prochaine itération de l'algorithme. Généralement, l'algorithme se termine lorsque soit un nombre maximum de générations a été produit, soit un niveau de conditionnement physique satisfaisant a été atteint pour la population.

Initialisation de l'algorithme génétique Au départ, des nombreuses solutions individuelles sont généralement générées de façon aléatoire pour former une population initiale, ce qui permet toute la gamme de solutions possibles. Parfois, les solutions peuvent être «ensemencées» dans les zones où des solutions optimales sont susceptibles d'être trouvées.

Sélection, au cours de chaque génération successive, une partie de la population existante est sélectionnée pour élever une nouvelle génération. Les solutions individuelles sont sélectionnées au moyen d'une fonction de conditionnement physique, qui évalue chaque individu et en fonction de cette fonction de conditionnement physique les meilleurs individus sont sélectionnés.

Dans l'outil algorithme génétique v4.0, la fonction de remise en forme est basée sur l'erreur absolue moyenne (MAE) et $MAE + 3\times \sigma_{AE}$ qui ont récemment été proposés dans la littérature, (Roy et al., 2015).Ici, 'σ_{AE}' représente l'écart-type des valeurs d'erreur absolue (AE). La fonction de condition physique (Ambure et al., 2016) (Eq. III.11) est définie comme suit :

$$F = \sum_{i=0}^{i=k} \frac{\hat{P}_{i,Threshold} - \hat{P}_i}{\hat{P}_{i,Threshold} - \hat{P}_{i,Ideal}} \qquad \text{III.11}$$

Ou: F : Score de forme physique,
k : Nombre de paramètre,
$\hat{P}$: Valeur du paramètre respectif pour une équation,
$\hat{P}_{i,Ideal}$: Valeur idéale de ce paramètre,
$\hat{P}_{i,Threshold}$: Seuil de ce paramètre.

Les mesures impliquées dans cette fonction de conditionnement physique ainsi que leur valeur idéale et seuil sont présentées dans le tableau III.11. Ici, un score de condition physique plus élevé signifie une meilleure adaptation du modèle QSAR en termes de qualité de prédiction réelle.

Tableau III.11. Les paramètres ainsi que leur valeur idéale et seuil comme employé dans la fonction de remise en forme de l'outil d'algorithme génétique interne.

Paramètres	Valeur idéale	Valeur de seuil
MAE	0.0	$0.15\times TrainYObsRange*$
$MAE + 3\times \sigma_{AE}$	0.0	$0.25\times TrainYObsRange*$

*TrainYObsRange = Gamme de valeurs de réponses observées

III.2.4. La méthodologie de l'analyse QSAR

III.2.4.1. Les descripteurs moléculaires

Le descripteur moléculaire est le résultat final d'une procédure logique et mathématique qui transforme l'information chimique chiffrée dans une représentation symbolique d'une molécule dans un nombre utile ou le résultat de quelque expérience standard.

III.2.4.2. La base de données

Le calcul des descripteurs moléculaires par le logiciel DRAGON nécessite la collection des notation SMILES (The simplified molecular-input line-entry system), qui définit comme une spécification sous forme d'une notation de ligne pour décrire la structure des espèces chimiques en utilisant des chaînes ASCII courtes. Les chaînes SMILES peuvent être importées par la plupart des éditeurs de molécules pour la conversion dans des dessins bidimensionnels ou des modèles tridimensionnels des molécules. Nous avons utilisé le site PubChem (https://pubchem.ncbi.nlm.nih.gov/) pour la collection des notations SMILES.

Le tableau III.12 représente la liste des types des descripteurs calculés par le logiciel DRAGON 7.0.

Tableau III.12. Liste des classes des descripteurs introduit dans le logiciel DRAGON 7.0

Classe	N° des descripteurs	Classe	N° des descripteurs
Constitutional	47	RDF descriptors	210
Ring descriptors	32	3D-MoRSE descriptors	224
Topological indices	75	WHIM descriptors	114
Walk and path counts	46	GETAWAY descriptors	273
Connectivity indices	37	Randic molecular profiles	41
Information indices	50	Functional groups count	154
2D matrix-based descriptors	607	Atom-centred fragments	115
2D autocorrelations	213	Atom-type E-state indices	172
Burden eigenvalues	96	CATS 2D	150
P-VSA-like descriptors	55	2D Atom Pairs	1596
ETA indices	23	3D Atom Pairs	36
Edge adjacency indices	324	Charge descriptors	15
Geometrical descriptors	38	Molecular properties	20
3D matrix-based descriptors	99	Drug-like indices	28
3D autocorrelations	80	CATS 3D	300
Totale		5270	

III.2.4.3. Réduction des variables

Pour affiner notre base de données nous avons utilisé deux méthodes pour la réduction du nombre du descripteurs :

1. La première a été fait à l'aide de logiciel DRAGON 7.0 qui été centré dans l'analyse de la base total.
2. La sélection des descripteurs les plus influencer sur la molécule par l'algorithme génétique.

A. Réduction par DRAGON 7.0

Dans cette première étape d'analyse et réduction nous avons utilisés les options introduites dans le logiciel :

- ❖ Exclure des descripteurs à valeur constante.
- ❖ Exclure des descripteurs à valeur constante et presque constante.
- ❖ Exclure les descripteurs avec un écart type inférieur à 0,05.
- ❖ Exclure des descripteurs avec au moins une valeur manquante.
- ❖ Exclure des descripteurs avec une corrélation supérieure ou égale à 0,75.

Tableau III.13. Représentation des résultats de réduction par le logiciel DRAGON 7.0.

	Avant la réduction	Après la réduction
Structure 2D – Descripteurs 2D	2566	207
Structure 2D – toutes les Descripteurs	5270	207
Structure 3D – Descripteurs 3D	739	124

B. Réduction par l'algorithme génétique

Dans cette deuxième étape de réduction nous avons utilisés les options suivantes :

- ❖ Couchez l'option de prétraitement de la base.
- ❖ Coupure de variance : 0.001.
- ❖ Coupure inter corrélation : 0.9.
- ❖ Nombre totale d'itérations / générations : 100.
- ❖ Longueur de l'équation : 3.
- ❖ Nombre d'équations sélectionnées dans chaque génération : 30.
- ❖ Couchez l'option de la validation du processus avec 10 modèles aléatoires.

Tableau III.14. Représentation des résultats de réduction par l'algorithme génétique.

	Avant la réduction	Après la réduction
Structure 2D – Descripteurs 2D	207	4
Structure 2D – toutes les Descripteurs	207	3
Structure 3D – Descripteurs 3D	124	3

III.2.5. Modélisation par réseau de neurone artificiel

Les réseaux neuronaux artificiels sont bien connus pour leurs structures dans lesquelles une grande quantité d'informations traitées via un réseau interconnecté comprenant plusieurs nœuds dans des couches successives, (Graupe, 2007). Le flux d'informations dépend de l'architecture du réseau qui imite celle présente dans les neurones des corps humains.

Sur la base de la structure biologiquement inspirée, la classification des motifs dans plusieurs domaines, y compris la reconnaissance, l'optimisation, le contrôle et la prédiction, est possible une fois que le modèle de réseau a été formé et testé bien pour différents ensembles de données. La sélection des facteurs d'influence les plus importants, la structure du réseau, le processus d'apprentissage et le problème des minimums locaux sont responsables du processus de développement du modèle. Après avoir sélectionné les variables d'entrée par voie théorique ou semi-empirique, le processus d'apprentissage accompagné de phases de formation et de test peut être effectué après la normalisation et l'élimination des valeurs aberrantes des variables d'entrée, (Passino, 2005). L'organigramme de la figure III.5 illustre le principe de calcul de la solubilité en utilisant les RNA.

III.2.5.1. Structure de la base de données

Afin de construire les modèles de réseaux neuronaux, les données expérimentales sont obtenues à partir de références pour 107 solutés solides. Les types de solutés et leurs gammes expérimentales sont fournis dans le tableau III.1. Les variables d'entrée sont la température $T(°K)$, la pression $P(Bar)$, la densité du CO2 $\rho_1(kg/m^3)$, les descripteurs retenus alors que la sortie est la solubilité du soluté solide dans le CO2 supercritique y_2(en fraction molaire). Le nombre total de 3010 points de données utilisés pour développer les réseaux de neurones diviser en 68%, 16% et 16% des données associées aux ensembles d'apprentissage, de validation et de test, respectivement. Par souci de comparaison, on utilise le AARD (Eq. III.2) et R^2 (Eq. III.3) pour détermine les performances des réseaux.

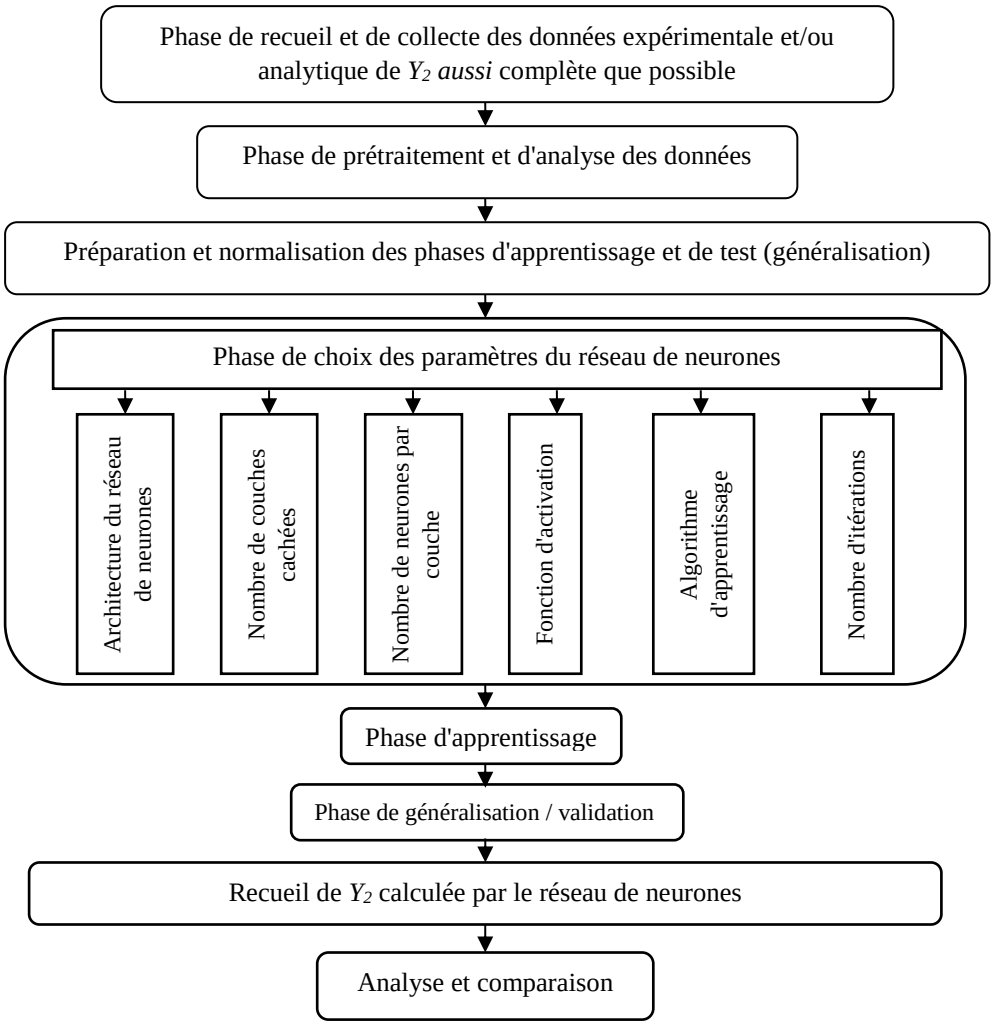


Figure III.12. Méthodologie de calcul de la solubilité par les réseaux de neurones.

Dans ce travail, quatre architectures de réseaux neuronaux multicouches sont utilisées afin d'étudier la performance de chaque modèle pour une estimation précise les solubilités.

III.2.5.2. Conception de la topologie de RN

Dans ce travail nous avons choisi de réaliser quatre topologies de RNA afin de comparer le choix des variables d'entrées. Le tableau III.15 résume les quatre topologies utiliser.

Tableau III.15. Topologies des RNAs utilisés

RNA	Variables d'entrées	Variables de sortie	N° de couches cachées
RNA simple	T, P, ρ_1, M_m, le N° du système	y_2	2
RNA 2D-2D	T, P, ρ_1, SpMax4_Bh(s), SM06_AEA(dm), nN, LLS_01	y_2	2
RNA 2D-All	T, P, ρ_1, SpMax4_Bh(s), SM06_AEA(dm), LLS_01	y_2	2
RNA 3D-3D	T, P, ρ_1, Mor14u, Mor18u, CATS3D_02_LL	y_2	2

III.2.5.2.1. Choix des entrées et des sorties

Afin de rassembler les 107 systèmes {solutés solides-scCO_2}, on a utilisé quatre réseaux **:**

- 1er RNA : le RNA simple
  - Entrés : T, P, ρ_1 ,la masse moléculaire et le numéro de système. Ce qui donne une matrice (3010×5).
  - Sortie : y_2. Ce qui donne un vecteur (3010×1).
- 2ème RNA : le RNA 2D-2D
  - Entrés : T, P, ρ_1 et les descripteurs sélectionnés (SpMax4_Bh(s), SM06_AEA(dm), nN, LLS_01). Ce qui donne une matrice (3010×7)
  - Sortie : y_2. Ce qui donne un vecteur (3010×1)
- 3ème RNA : le RNA 2D-All
  - Entrés : T, P, ρ_1 et les descripteurs sélectionnés (SpMax4_Bh(s), SM06_AEA(dm), LLS_01). Ce qui donne une matrice (3010×6)
  - Sortie : y_2. Ce qui donne un vecteur (3010×1)
- 4ème RNA : le RNA 3D-3D
  - Entrés : T, P, ρ_1 et les descripteurs sélectionnés (Mor14u, Mor18u, CATS3D_02_LL). Ce qui donne une matrice (3010×6)
  - Sortie : y_2. Ce qui donne un vecteur (3010×1)

III.2.5.2.2. Préparation de la base de données

La base de données d'entrée est réalisée par une matrice rectangulaire contient 3010 lignes et 15 colonnes répartie de la façon suivante :

- Les colonnes (11, 12, 13,14, 15 et 16) pour le RNA simple.
- Les colonnes (4, 5, 6, 7, 11, 12, 13 et 16) pour le RNA 2D-2D.
- Les colonnes (1, 2, 3, 11, 12,13 et 16) pour le RNA 2D-All.
- Les colonnes (8, 9, 10, 11, 12, 13 et 16) pour le RNA 3D-3D.

La base de données ainsi constituée est divisée en trois sous - bases :

- Une base d'apprentissage qui permet l'optimisation de l'architecture de RN
- Une base de test utilisée pour déterminer les performances prédictives du RN.
- Une base de validation pour valider le modèle de RN.

III.2.5.2.3. Normalisation des données

Les données sont représentées en quinze lignes, correspondant respectivement aux : {T, P, ρ_1, SpMax4_Bh(s), SM06_AEA(dm), nN, LLS_01, Mor14u, Mor18u, CATS3D_02_LL, Masse moléculaire, le N° de système et y_2}.

Les valeurs numériques ont été normalisées afin d'améliorer l'optimisation. La normalisation se fait par colonne, chaque paramètre des entrées et de sorties (x_i) est normalisé (x_{in}) par l'expression (Premnmx sur MATLAB) suivante :

$$x_{in} = 2\frac{\{x_i - \min(x_i)\}}{\{\max(x_i) - \min(x_i)\}} - 1 \qquad \text{III.12}$$

La dénormalisation des données pour la sortie (x_{in}) a été fait par l'expression (Postmnmx sur MATLAB) suivante :

$$x_i = \frac{x_{in}\{\max(x_i) - \min(x_i)\} + \{\max(x_i) + \min(x_i)\}}{2} \qquad \text{III.13}$$

III.2.5.3. Résultats et discussion

Nous avons essayé avec une ensuite deux couches cachées, d'après les résultats un RN avec deux couches cachées semble meilleur, raison pour laquelle on donne l'organigramme pour deux couches cachées, le nombre de neurones dans les couches cachées a été sélectionné aléatoirement.

Il n'y a pas de formule magique pour sélectionner le nombre optimal de neurones dans les couches cachées. Nous avons donc suivi un moyen de trouver le meilleur nombre de neurones dans les couches cachées en utilisant un algorithme basé sur le langage MATLAB. Nous avons atteint le nombre de neurones dans les couches cachées comme indiqué dans le tableau III.6.

Tableau III.16. Architecture des RNAs utilisés

	Type du réseau	Réseau de neurones statique	
RNA simple	Couche	Nb. Des neurones	Fonction d'activation
	Couche d'entrée	5	
	Première Couche cachée	10	Log-sigmoïde LOGSIG
	Deuxième Couche cachée	20	Log-sigmoïde LOGSIG
	Couche de sortie	1	Identité purline
	Algorithme d'apprentissage	Levenberg -Marquardt	
RNA 2D-2D	Couche	Nb. Des neurones	Fonction d'activation
	Couche d'entrée	7	
	Première Couche cachée	25	Log-sigmoïde LOGSIG
	Deuxième Couche cachée	10	Log-sigmoïde LOGSIG
	Couche de sortie	1	Identité purline
	Algorithme d'apprentissage	Levenberg -Marquardt	

RNA 2D-All	Couche	Nb. Des neurones	Fonction d'activation
	Couche d'entrée	6	
	Première Couche cachée	10	Tangente hyperbolique TANSIG
	Deuxième Couche cachée	20	Tangente hyperbolique TANSIG
	Couche de sortie	1	Identité purline
	Algorithme d'apprentissage	Levenberg -Marquardt	
RNA 3D-3D	Couche	Nb. Des neurones	Fonction d'activation
	Couche d'entrée	6	
	Première Couche cachée	10	Log-sigmoïde LOGSIG
	Deuxième Couche cachée	20	Log-sigmoïde LOGSIG
	Couche de sortie	1	Identité purline
	Algorithme d'apprentissage	Levenberg -Marquardt	

Dans notre travail on exploite la technique de RNA, en utilisant le logiciel MATLAB R2009b afin de prédire la solubilité.

L'optimisation de l'architecture du RN consiste à la détermination :

- Du nombre des neurones dans la couche cachée,
- De la fonction d'activation de la couche de sortie,
- De l'algorithme d'apprentissage.

Pour comparer la précision de chaque RNA on à calculer deux constantes : l'AARD (Eq. III.2) et R^2 (Eq. III.3). Le tableau III.17 illustre les résultats trouvés.

Tableau III.17. Résultats des AARDs et R^2 pour les 107 système (solutés-scCO_2)

N°	RNAS	RNA2D	RNAAll	RNA3D
1	11,02	**9,62**	9,81	10,06
2	**6,99**	8,92	10,22	19,25
3	**35,28**	35,91	39,84	36,38
4	15,61	16,83	**15,21**	15,65
5	18,13	**5,64**	8,36	16,10
6	**11,70**	15,82	18,33	35,48
7	**6,87**	8,59	11,83	11,27
8	14,77	**6,54**	6,82	12,66
9	9,35	**6,46**	7,93	11,00
10	19,98	19,71	20,79	**19,61**
11	5,22	4,34	**3,83**	8,29
12	25,41	**22,27**	28,66	22,32
13	11,51	**10,32**	12,24	13,70
14	11,40	**7,37**	9,16	11,52
15	9,31	**8,03**	13,48	15,50
16	13,44	**7,62**	7,73	16,46
17	5,47	7,68	**5,36**	16,36
18	13,77	**12,04**	12,31	28,15
19	21,57	**9,98**	20,82	17,93
20	48,68	**10,84**	16,56	41,76
21	32,47	11,24	**8,60**	17,00
22	43,13	12,17	**8,64**	41,86
23	32,62	**5,99**	7,03	46,33

24	24,03	**9,48**	18,22	26,70
25	17,31	**10,37**	12,46	13,35
26	14,84	17,80	**10,71**	19,86
27	16,60	14,10	**6,71**	14,87
28	21,88	**14,28**	14,30	17,12
29	14,26	11,06	**9,19**	13,91
30	31,14	**13,10**	22,21	17,54
31	62,76	16,20	**11,37**	13,26
32	17,05	**7,15**	12,64	10,81
33	8,59	11,45	**7,12**	10,68
34	16,79	**8,77**	14,55	18,75
35	**7,35**	16,81	12,15	34,78
36	13,99	13,37	17,88	**13,30**
37	10,86	**9,35**	10,15	12,29
38	15,26	**6,45**	12,23	14,68
39	21,22	**20,83**	23,57	27,58
40	22,59	**17,99**	22,15	24,86
41	22,80	**16,28**	17,74	21,26
42	9,43	**7,83**	8,76	15,44
43	16,18	14,70	**13,84**	17,56
44	12,14	**6,73**	11,30	11,34
45	**10,61**	11,74	16,19	11,99
46	25,62	**21,54**	22,14	29,13
47	13,85	12,45	**7,72**	22,99
48	**3,89**	11,60	9,73	8,91
49	47,26	**8,42**	9,55	49,62
50	14,13	15,95	**13,52**	18,52
51	15,19	**11,07**	12,99	15,45
52	10,33	7,24	**6,44**	17,01
53	18,34	**17,47**	19,08	29,65
54	14,73	**11,29**	14,81	13,51
55	12,23	**4,91**	23,90	7,69
56	**11,36**	12,96	12,12	19,16
57	27,53	39,66	**25,04**	30,68
58	7,07	**6,34**	7,39	22,53
59	**23,37**	25,64	134,75	24,38
60	9,61	**8,78**	12,96	36,97
61	22,13	**21,49**	29,87	22,17
62	**10,68**	11,28	14,53	15,42
63	**5,12**	5,31	17,33	35,40
64	6,73	6,82	**4,52**	8,74
65	17,99	16,91	**15,23**	19,35
66	17,21	**6,86**	14,49	19,24
67	11,84	**6,51**	10,84	6,51
68	18,05	**17,95**	23,06	18,70
69	15,00	14,99	**12,98**	15,06
70	6,95	**3,87**	8,34	21,35
71	11,05	**8,03**	19,77	14,42
72	10,29	**9,36**	9,90	17,35
73	28,22	**24,80**	30,16	75,56
74	16,79	**13,35**	15,97	35,54
75	**4,94**	8,07	18,89	27,72
76	**7,35**	7,54	9,70	19,91
77	17,64	19,14	23,91	**15,63**
78	**21,59**	25,48	27,27	27,72
79	11,19	12,90	11,01	**9,75**
80	14,27	**10,08**	12,52	26,05
81	9,76	**8,08**	11,25	9,34
82	12,25	**9,56**	9,79	23,70
83	**3,60**	5,61	5,55	12,10
84	17,25	**15,71**	17,30	28,98
85	**23,71**	30,49	43,57	32,54
86	**17,44**	25,16	29,79	18,83
87	**8,21**	12,11	71,91	19,70
88	18,45	19,44	17,19	**16,51**
89	37,64	26,90	**22,69**	29,16
90	12,54	14,94	**12,54**	20,38
91	13,09	**7,16**	22,38	26,03
92	10,10	**8,00**	11,25	32,52
93	8,56	**7,59**	13,26	10,97
94	18,03	**11,88**	21,23	30,04
95	8,47	**4,06**	12,62	12,18
96	**9,69**	10,66	23,04	25,03
97	33,75	39,15	**25,35**	114,57
98	17,67	15,94	13,48	**12,81**
99	12,49	9,17	**9,00**	13,72
100	10,84	8,84	**6,42**	41,30
101	10,37	**10,22**	17,48	30,14
102	**6,90**	7,26	8,39	9,09
103	43,86	**26,29**	38,25	32,83
104	6,81	9,40	**6,56**	12,99
105	8,75	8,08	**7,74**	19,74
106	25,75	**17,11**	19,84	56,32
107	8,69	**8,23**	14,54	12,72
Min AARD	**3,60**	3,87	3,83	6,51
Max AARD	62,76	**39,66**	134,75	114,57
Moyenne	16,59	**12,92**	16,67	22,14
R^2	0,9899	**0,9931**	0,9894	0,9860

Le tableau III.17 montre que la gamme des valeurs de AARD est entre 3.60% et 134.75%, mais quand nous prendrons les meilleures des AARD mentionné en gras, nous trouverons une gamme des valeurs entre 3.60% et 35.28% avec une moyenne de 11.38%.

L'AARD minimum (3.60%) est obtenu par le RNA simple pour le système N° 83 ($scCO_2$ - Piracetam), et l'AARD maximum (134.75%) est obtenu par le RNA 2DAll pour le système N° 59 ($scCO_2$ - Megestrol acetate).

La comparaison globale entre la performance des RNAs indique que le RNA 2D2D a donné la meilleure corrélation, ou la gamme de l'ARRD est entre 3.87% et 39.66% avec une moyenne de 12.92%. Le second RNA qui a la meilleure corrélation est le RNA 2DAll ou la gamme de l'AARD est entre 3.83% et 134.75% avec une moyenne de 16.67%. Le troisième avec une meilleure corrélation est le RNA simple ou la gamme de l'AARD est entre 3.60% et 62.76% avec une moyenne de 16.59%. Ces trois réseaux ensemble représente une meilleure corrélation de 76% des 107 composés considéré.

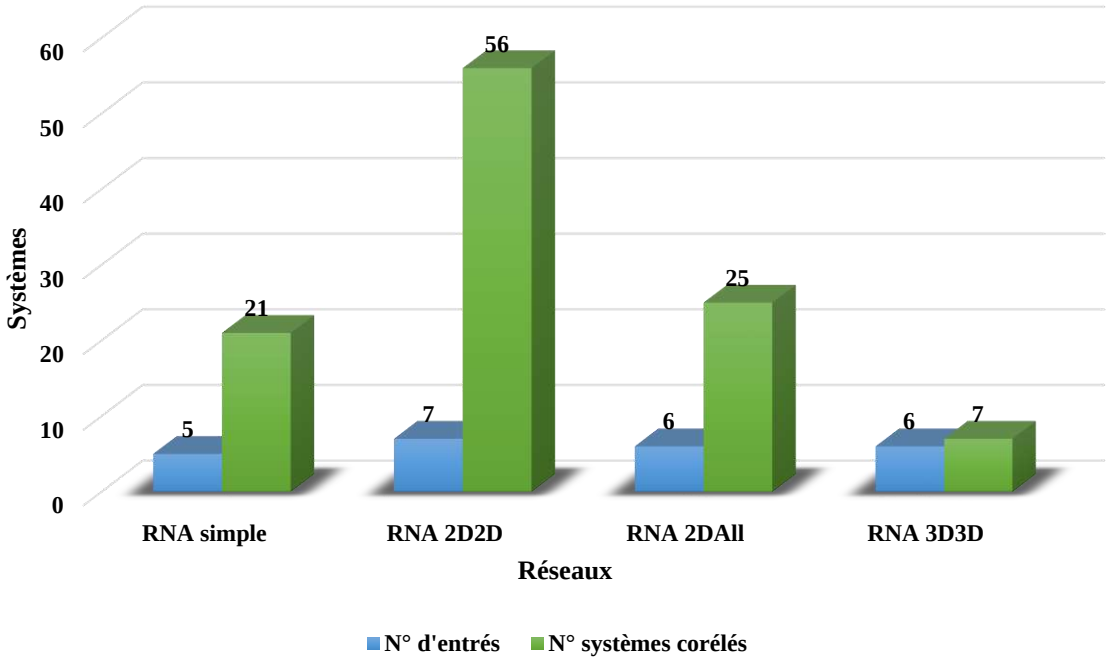


Figure III.13. Le nombre des systèmes corrélés et le nombre d'entrées pour chaque réseau.

III.2.6. Equation du RNA

D'après la théorie du réseau de neurone artificiel on peut tirer une relation entre la sortie et les entrés. L'équation III.14 représente la relation sortie-entrés du RNA 2D2D.

$$y = \sum_{j=1}^{k} w_{j,s} \left(\frac{1}{1+e^{-\left(\sum_{i=1}^{m} w_{i,j}\left(\frac{1}{1+e^{-\left(\sum_{e=1}^{n} w_{e,i} x_e\right)+b_i}}\right)+b_j\right)}} \right) + b_s \qquad \text{III.14}$$

Où : e, s, n, m et k sont l'entré, la sortie, le nombre de neurones dans la couche d'entrée, la couche cachée 1 et la couche caché 2 respectivement.

III.2.7. Validation

Dans cette phase on a choisi le système (Curcumin – $scCO_2$) qui n'été pas sélectionnés parmi les 107 composés introduit dans l'étape de la modélisation pour vérifier la validation de RNA 2D2D sélectionner, tableau III.18 représente les valeurs expérimentales et calculés pour le système sélectionné.

Les graphes de validation sont représentés dans les figures III.14 et III.15 respectivement.

Tableau III.18. Les valeurs de ($T(°K)$, $P(MPa)$, $\rho_1(kg/m^3)$ et y^{exp}) sont mesuré par (Zhan et al., 2017) et les valeurs des descripteurs (SpMax4_Bh(s), SM06_AEA(dm), nN, LLS_01) calculé par le logiciel DRAGON7.0

$T(°K)$	$P(Bars)$	$\rho_1(kg/m^3)$	SpMax4_Bh(s)	SM06_AEA(dm)	nN	LLS_01	$y_2^{exp}.10^8$	$y_2^{cal}(RNA).10^8$
308,15	80	419,09	6.467	0.479	0	0.33	2,07	2,53
308,15	100	712,81	6.467	0.479	0	0.33	2,51	3,25
308,15	130	785,7	6.467	0.479	0	0.33	2,68	2,90
308,15	160	827,17	6.467	0.479	0	0.33	2,85	2,94
308,15	200	867,72	6.467	0.479	0	0.33	3,11	3,16
308,15	80	241,05	6.467	0.479	0	0.33	1,82	1,23
318,15	100	498,25	6.467	0.479	0	0.33	2,31	1,06
318,15	130	693,65	6.467	0.479	0	0.33	2,69	3,77
318,15	160	759,98	6.467	0.479	0	0.33	3,12	3,42
318,15	200	812,69	6.467	0.479	0	0.33	3,52	3,51
328,15	80	203,64	6.467	0.479	0	0.33	1,86	5,06
328,15	100	325,07	6.467	0.479	0	0.33	2,03	1,12
328,15	130	571,33	6.467	0.479	0	0.33	2,62	6,95
328,15	160	681,12	6.467	0.479	0	0.33	3,25	4,62
328,15	200	754,61	6.467	0.479	0	0.33	3,70	4,32

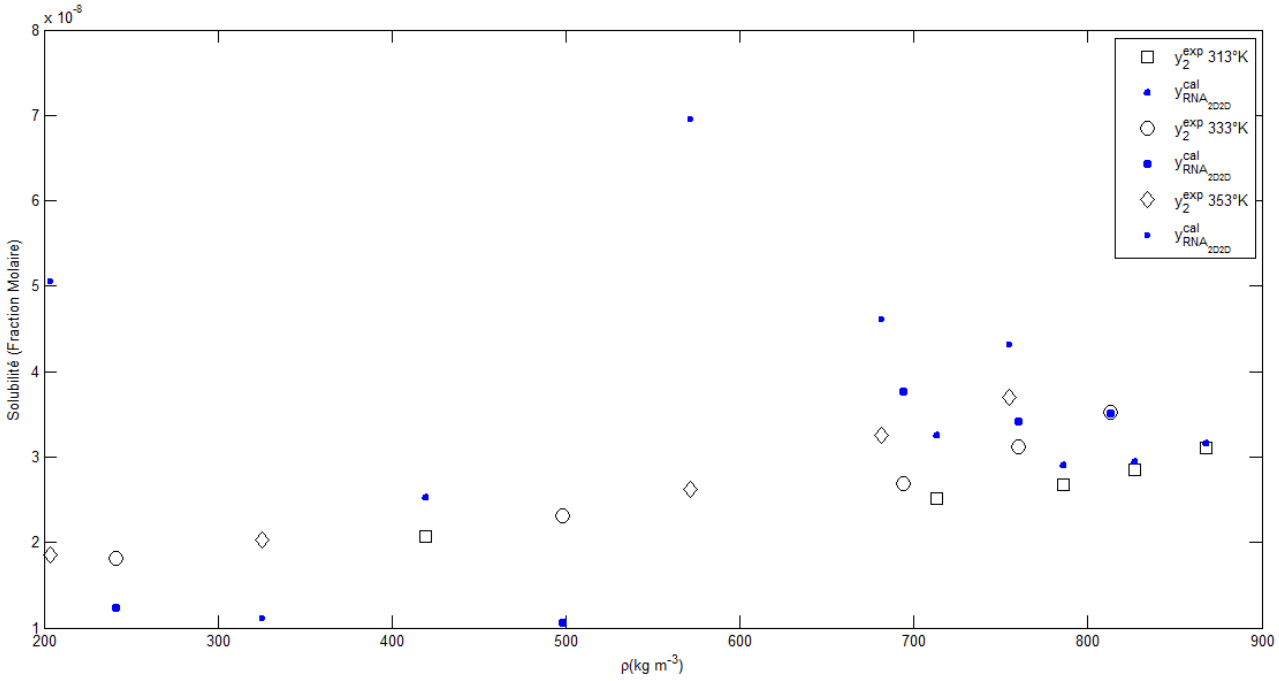


Figure III.14. Solubilité expérimentale de la curcumine in $scCO_2$ et prédire par le RNA 2D2D en fonction de la densité du fluide supercritique (graphe 2D)

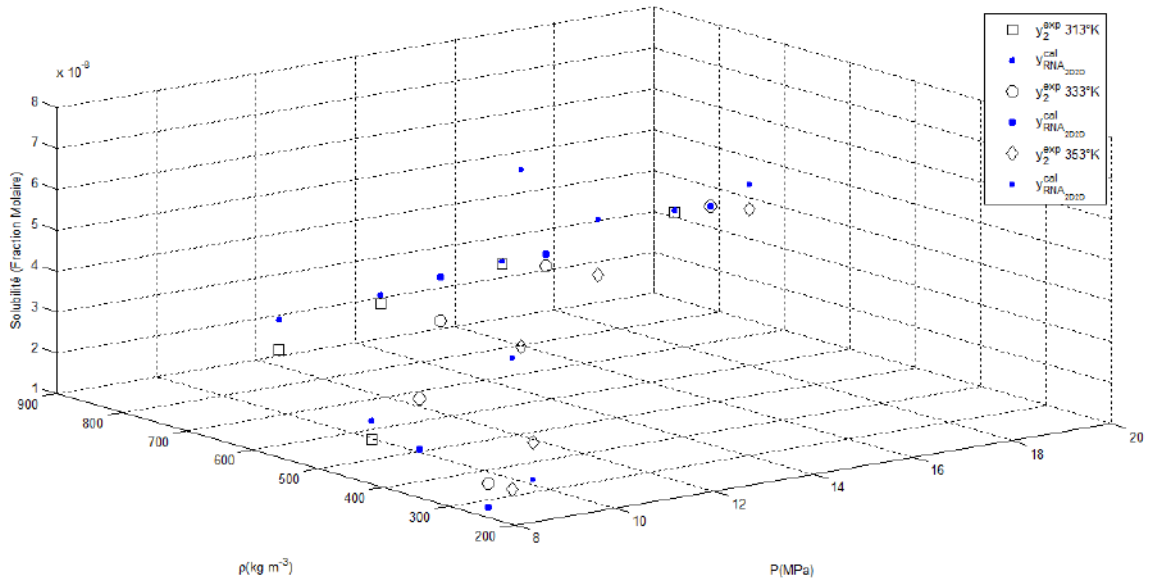


Figure III.15. Solubilité expérimentale de la curcumine in $scCO_2$ et prédire par le RNA 2D2D en fonction de la densité du fluide supercritique et la pression (graphe 3D)

III.2.8. Conclusion

La modélisation neuronale de la solubilité du 107 composés solides dans le CO_2 supercritique choisis les descripteurs moléculaires et les variables intensive comme des entrés a fait sortir les remarques suivantes :

- Les résultats de la modélisation neuronale montrent que les RNA, comme il était montré dans plusieurs études passés peut-être utilisé avec fiabilité pour représenter les données expérimentales de la solubilité des solutés solides dans le CO_2 supercritique.
- La modélisation neuronale par les descripteurs choisi à partir des structures 2D donne un meilleur résultat par rapport aux autres.

Les résultats trouvés montrent que les descripteurs moléculaires peuvent influencer sur les résultats de prédiction de la solubilité mais reste que la modélisation semi-empirique basé sur les variables intensives donne un meilleur résultat pour la prédiction de la solubilité.

Conclusion générale

Les équations empiriques, semi-empirique et le réseau de neurones ont été comparés pour la prédiction de la solubilité de 107 composés différents dans le dioxyde de carbone supercritique. Les équations semi-empiriques trouvent de larges applications dans la modélisation de la solubilité en raison de leurs facilités d'utilisation et ont permis d'obtenir des précisions plus élevées. L'exactitude de ces équations est généralement jugée sur l'accord avec des données expérimentales sur lesquelles elles sont basées. Le réseau neuronal est un puissant modèle de données informatiques capables de capturer et de représenter des relations d'entrée/sortie complexes. La motivation pour le développement de la technique des réseaux de neurones découle du désir de développer un système artificiel capable de réaliser des tâches «intelligentes» semblables à celles réalisées par le cerveau humain. Cette petite recherche a fait l'objet de prédire la solubilité des solutés solides dans le CO_2 supercritique utilisant deux méthodes de modélisation.

Dans la première contribution nous avons fait la prédiction de la solubilité des solutés solides dans le CO_2 supercritique utilisant les équations semi-empirique. Afin d'évaluer la capacité de ces modèles à corréler avec précision les données de la solubilité, nous avons effectué dans cette contribution une étude comparative sur une base de 107 composés totalisant un ensemble de 3010 points expérimentaux, recueillis à partir de la littérature publiée au cours des 25 dernières années. Pour corréler ces données nous avons utilisé une technique d'optimisation hybride combinant des méthodes d'intelligence artificielle, notamment l'optimisation par l'algorithme génétique, avec les méthodes d'optimisation classique utilisant l'algorithme de Levenberg-Marquardt. Sur la base des résultats obtenus, un nouveau modèle basé sur la densité a été proposé. Le modèle proposé a donné une meilleure corrélation globale avec une erreur relative absolue moyenne (AARD = 7.45 %) et un coefficient de détermination (R^2 = 0.9964). Les résultats du travail effectué dans cette contribution nous ont permis de mieux comprendre quelques aspects de la modélisation empirique et semi-empirique de la solubilité des solutés solides dans le CO_2 supercritique.

- Les résultats de corrélation montrent que les modèles qui donnent la solubilité en fonction de la température et la pression ne donnent pas un bon résultat de corrélation par contre les modèles qui donne la solubilité en fonction de la température et la densité

donnent beaucoup plus une meilleure corrélation ceci a été démontrer par les résultats trouvés et les modifications effectués.

- Le choix de la base de données est primordial lorsqu'il s'agit de comparer les modèles empiriques et semi-empiriques. Plus la base de données est large et complète plus les résultats de comparaison sont significatifs.

- Pour comprendre davantage la capacité de corrélation d'un modèle donnée, il nous semble qu'il faut aller à une analyse approfondie des résultats obtenus. Par exemple, il faut étudier la qualité et la quantité des données expérimentaux des composés pour lesquels le modèle considéré a donné de meilleures ou mauvaises corrélations.

Dans la deuxième contribution nous avons fait la prédiction de la solubilité des solutés solides dans le CO_2 supercritique utilisant le réseau de neurone artificiel. Dans cette contribution nous avons collecté les notation SMILES d'après le site « PubChem » pour les 107 composés utilisés dans la première contribution, ces notations sont utilisées pour évaluer les descripteurs moléculaires. D'abord nous avons utilisé le logiciel DRAGON7.0 pour évaluer les valeurs des descripteurs moléculaire à l'aide des notation SMILES, ensuite ces valeurs sont traitées par le logiciel DRAGON7.0 soit même et par l'algorithme génétique pour déterminer les valeurs des descripteurs les plus influençant sur chaque composé. Pour évaluer la performance des RNA, Nous avons utilisé l'algorithme l'optimisation classique Levenberg-Marquardt (la fonction ***trainlm*** de MATLAB) comme un algorithme d'apprentissage et les fonctions ***logsig*** et ***tansig*** de MATLAB comme des fonction d'activation. Le réseau RNA 2D2D a donné une meilleure corrélation globale avec une erreur relative absolue moyenne (AARD = 12.92%) et un coefficient de corrélation (R^2 = 0.9931). Les résultats du travail effectué dans cette contribution nous ont permis de mieux comprendre quelques aspects de la modélisation neuronale de la solubilité des solutés solides dans le CO_2 supercritique.

- Les résultats trouvés montrent que les descripteurs 2D peut améliorer la prédiction de la solubilité des solutés solides dans le CO_2 supercritique.

- Le nombre de neurones dans les couches doit être choisi avec précaution.

- Les descripteurs 3D donnent un mauvais résultat de prédiction de la solubilité.

Dans le but d'améliorer notre travail, nous allons viser les points suivants :

- Modélisation empirique / semi empirique :
  - Développement des autres modèles empiriques et/ou semi-empiriques avec un minimum de paramètres,
  - Utilisation d'autre méthodes d'optimisation pour l'optimisation des coefficients des modèles tel que : PSO, ABC, ACO, …. etc.
- Modélisation par intelligence artificielle :
  - Utilisation des modèles basé sur la méthode des MACHINES À VECTEURS DE SUPPORT (SVM),
  - Utilisation des modèles basé sur la méthode des LOGIQUES FLOUES,
  - Couplage entre le RNA et la LOGIQUE FLOUE,
  - Optimisation de la convergence du RNA par les méthodes d'optimisation tel que : PSO, ABC, …. etc.

Références bibliographiques

Abdallah el hadj, A. (2009) Prédiction de la solubilité dans les phases supercritiques à partir des équations basées sur la théorie associative. Université Yahia Farès de Médéa.

Adachi, Y., Lu, B.C.-Y. (1984) Simplest equation of state for vapor-liquid equilibrium calculation: A modification of the van der waals equation. AIChE Journal. 30(6), 991–993.

Adachi, Y., Lu, B.C.-Y., 1984. Simplest equation of state for vapor-liquid equilibrium calculation: A modification of the van der waals equation. AIChE J. 30, 991–993. doi:10.1002/aic.690300619

Adachi, Y., Sugie, H. (1986) A new mixing rule-modified conventional mixing rule. Fluid Phase Equilibria. 28(2), 103–118.

Albert, A., Rubbo, S., Goldacre, R., 1941. Correlation of basicity and antiseptic action in an acridine series. Nature 147, 332–333. doi:10.1038/147332a0

Ambure, P., Roy, K., Provenzani, R., Marengo, E., Wal, L. van der, Gramatica, P., Desrayaud, S., Rondeau, J.-M., Mobitz, H., Neumann, U., 2016. Understanding the structural requirements of cyclic sulfone hydroxyethylamines as hBACE1 inhibitors against Aβ plaques in Alzheimer's disease: a predictive QSAR approach. RSC Adv. 6, 28171–28186. doi:10.1039/C6RA04104C

Amooey, A.A., 2014. A simple correlation to predict drug solubility in supercritical carbon dioxide. Fluid Phase Equilib. 375, 332–339. doi:10.1016/j.fluid.2014.05.025

Asgarpour Khansary, M., Amiri, F., Hosseini, A., Hallaji Sani, A., Shahbeig, H., 2015. Representing solute solubility in supercritical carbon dioxide: A novel empirical model. Chem. Eng. Res. Des. 93, 355–365. doi:10.1016/j.cherd.2014.05.004

Asghari-Khiavi, M., Yamini, Y., 2003. Solubility of the Drugs Bisacodyl, Methimazole, Methylparaben, and Iodoquinol in Supercritical Carbon Dioxide. J. Chem. Eng. Data 1, 61–65. doi:10.1021/JE020080H

Asghari-Khiavi, M., Yamini, Y., Farajzadeh, M.A., 2004. Solubilities of two steroid drugs and their mixtures in supercritical carbon dioxide. J. Supercrit. Fluids 30, 111–117. doi:10.1016/j.supflu.2003.07.002

Ashour, I., Almehaideb, R., Fateen, S.-E., Aly, G. (2000) Representation of solid-supercritical fluid phase equilibria using cubic equations of state. Fluid Phase Equilibria. 167(1), 41–61.

Asiabi, H., Yamini, Y., Latifeh, F., Vatanara, A., 2015. Solubilities of four macrolide antibiotics in supercritical carbon dioxide and their correlations using semi-empirical models. J. Supercrit. Fluids 104, 62–69. doi:10.1016/j.supflu.2015.05.018

Asiabi, H., Yamini, Y., Tayyebi, M., Moradi, M., Vatanara, A., Keshmiri, K., 2013. Measurement and correlation of the solubility of two steroid drugs in supercritical carbon dioxide using semi empirical models. J. Supercrit. Fluids 78, 28–33. doi:10.1016/j.supflu.2013.03.018

Balaban, A.T., Klein, D.J., 2006. Is chemistry "The Central Science"? How are different sciences related? Co-citations, reductionism, emergence, and posets. Scientometrics 69, 615–637. doi:10.1007/s11192-006-0173-2

Bartle, K.D., Clifford, A.A., Jafar, S.A., Shilstone, G.F. (1991) Solubilities of Solids and Liquids of Low Volatility in Supercritical Carbon Dioxide. Journal of Physical and Chemical Reference Data. 20(4), 713–756.

Bartle, K.D., Clifford, A.A., Jafar, S.A., Shilstone, G.F., 1991. Solubilities of Solids and Liquids of Low Volatility in Supercritical Carbon Dioxide. J. Phys. Chem. Ref. Data 20, 713–756. doi:10.1063/1.555893

Baum, F., 1899. Zur Theorie der Alkoholnarkose. Arch. für Exp. Pathol. und Pharmakologie 42, 119–137. doi:10.1007/BF01834480

Bazua, E, R. (1983) Cubic equation of state for mixtures Containing Polar copounds. Chem. Ing. Therm, by Newman.

Bell, P.H., Roblin, R.O., 1942. Studies in Chemotherapy. VII. A Theory of the Relation of Structure to Activity of Sulfanilamide Type Compounds. J. Am. Chem. Soc. 64, 2905–2917. doi:10.1021/ja01264a055

Bellazzi, R., Abu-Hanna, A. (2009) Artificial intelligence in medicine AIME'07. Artificial Intelligence in Medicine. 46(1), 1–3.

Bian, X., Du, Z., Tang, Y., 2011. An improved density-based model for the solubility of some compounds in supercritical carbon dioxide. Thermochim. Acta 519, 16–21. doi:10.1016/j.tca.2011.02.023

Bian, X.-Q., Zhang, Q., Du, Z.-M., Chen, J., Jaubert, J.-N., 2016. A five-parameter empirical model for correlating the solubility of solid compounds in supercritical carbon dioxide. Fluid Phase Equilib. 411, 74–80. doi:10.1016/j.fluid.2015.12.017

Borodin A, 1858. Ob analogii myshiakovoi kislotys fosforoyu v khimischeskom i toksicheskov otnosheniyakh (Concerning the analogy between arsenic and phosphoric acids in chemical and toxicological respects). St. Petersburg.

Boston, J.F., Mathias, P.M. (1980) Phase Equilibria in a Third-Generation Process Simulator. Proceedings of the 2nd International Conference on Phase Equilibria and Fluid Properties in the Chemical Process Industries. (August), 823–849.

Brown, A.C., Fraser, T.R., 1868. On the Connection between Chemical Constitution and Physiological Action; with special reference to the Physiological Action of the Salts of the Ammonium Bases derived from Strychnia, Brucia, Thebaia, Codeia, Morphia, and Nicotia. J. Anat. Physiol. 2, 224–42.

Burgos-Solórzano, G.I., Brennecke, J.F., Stadtherr, M.A., 2004. Solubility measurements and modeling of molecules of biological and pharmaceutical interest with supercritical CO2. Fluid Phase Equilib. 220, 57–69. doi:10.1016/j.fluid.2004.01.036

Ch, R., Madras, G., 2010. An association model for the solubilities of pharmaceuticals in supercritical carbon dioxide. Thermochim. Acta 507, 99–105. doi:10.1016/j.tca.2010.05.006

Chapman, W.G., Gubbins, K.E., Jackson, G., Radosz, M. (1989) SAFT: Equation-of-state solution model for associating fluids. Fluid Phase Equilibria. 52, 31–38.

Charton, M., 2008. Philip S. Magee: a life in QSAR. J. Comput. Aided. Mol. Des. 22, 335–337. doi:10.1007/s10822-007-9153-8

Chen, Y.M., Chen, Y.P., 2009. Measurements for the solid solubilities of antipyrine, 4-aminoantipyrine and 4-dimethylaminoantipyrine in supercritical carbon dioxide. Fluid Phase Equilib. 282, 82–87. doi:10.1016/j.fluid.2009.04.019

Chen, Y.M., Lin, P.C., Tang, M., Chen, Y.P., 2010. Solid solubility of antilipemic agents and micronization of gemfibrozil in supercritical carbon dioxide. J. Supercrit. Fluids 52, 175–182. doi:10.1016/j.supflu.2009.12.012

Cheng, K.W., Tang, M., Chen, Y.P., 2002. Solubilities of benzoin, propyl 4-hydroxybenzoate and mandelic acid in supercritical carbon dioxide. Fluid Phase Equilib. 201, 79–96. doi:10.1016/S0378-3812(02)00070-5

Chim, R., Marceneiro, S., Braga, M.E.M., Dias, A.M.A., de Sousa, H.C., 2012. Solubility of norfloxacin and ofloxacin in supercritical carbon dioxide. Fluid Phase Equilib. 331, 6–11. doi:10.1016/j.fluid.2012.06.023

Chim, R.B., de Matos, M.B.C., Braga, M.E.M., Dias, A.M.A., de Sousa, H.C., 2012. Solubility of Dexamethasone in Supercritical Carbon Dioxide. J. Chem. Eng. Data 57, 3756–3760. doi:10.1021/je301065f

Chrastil, J. (1982) Solubility of solids and liquids in supercritical gases. The Journal of Physical Chemistry. 86(15), 3016–3021.

Chrastil, J., 1982. Solubility of solids and liquids in supercritical gases. J. Phys. Chem. 86, 3016–3021. doi:10.1021/j100212a041

Coimbra, P., Duarte, C.M.M., De Sousa, H.C., 2006. Cubic equation-of-state correlation of the solubility of some anti-inflammatory drugs in supercritical carbon dioxide. Fluid Phase Equilib. 239, 188–199. doi:10.1016/j.fluid.2005.11.028

Coimbra, P., Fernandes, D., Ferreira, P., Gil, M.H., de Sousa, H.C., 2008. Solubility of Irgacure?? 2959 photoinitiator in supercritical carbon dioxide: Experimental determination and correlation. J. Supercrit. Fluids 45, 272–281. doi:10.1016/j.supflu.2008.01.014

Coimbra, P., Fernandes, D., Gil, M.H., Sousa, H.C. de, 2008. Solubility of Diflunisal in Supercritical Carbon Dioxide. J. Chem. Eng. Data 53, 1990–1995. doi:10.1021/je800384h

Coley, D.A. (1999) An Introduction to Genetic Algorithms for Scientists and Engineers. , 125–126.

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. J. Am. Chem. Soc. 110, 5959–5967. doi:10.1021/ja00226a005

Cros AFA, 1863. Action de l'alcool amylique sur l'organisme. Université de Strasbourg (1538-1969). Faculté de médecine.

Dashtizadeh, A., Pazuki, G.R., Taghikhani, V., Ghotbi, C. (2006) A new two-parameter cubic equation of state for predicting phase behavior of pure compounds and mixtures. Fluid Phase Equilibria. 242(1), 19–28.

Dean, J.R., Kane, M., Khundker, S., Dowle, C., Tranter, R.L., Jones, P., 1995. Estimation and determination of steroid solubility in supercritical carbon dioxide. Analyst 120, 2153. doi:10.1039/an9952002153

Del Valle, J.M., Aguilera, J.M., 1988. An improved equation for predicting the solubility of vegetable oils in supercritical carbon dioxide. Ind. Eng. Chem. Res. 27, 1551–1553. doi:10.1021/ie00080a036

Dreyfuss, G., Kim, V.N., Kataoka, N. (2002) MESSENGER-RNA-BINDING PROTEINS AND THE MESSAGES THEY CARRY. Nature Reviews Molecular Cell Biology. 3(3), 195–205.

Du, Y., Guo, T. (1988) The prediction of gas Hydrates Formation Condition. Acadimic Periodical Press: Beijing, China, 486.

Ferguson, J., 1939. The use of chemical potentials as indices of toxicity. Proc. R. Soc. London. Ser. B, 127, 387–404.

Free, S.M., Wilson, J.W., 1964. A Mathematical Contribution to Structure-Activity Studies. J. Med. Chem. 7, 395–399. doi:10.1021/jm00334a001

Fujita, T., Ban, T., 1971. Structure-activity relation. 3. Structure-activity study of phenethylamines as substrates of biosynthetic enzymes of sympathetic transmitters. J. Med. Chem. 14, 148–152. doi:10.1021/jm00284a016

Gallinari, P., Gascuel, O. (1996) Statistique, apprentissage et généralisation. Application aux réseaux de neurones. Revue d'intelligence artificielle.

Garlapati, C., Madras, G., 2009. Solubilities of solids in supercritical fluids using dimensionally consistent modified solvate complex models, Fluid Phase Equilibria. doi:10.1016/j.fluid.2009.05.013

Garlapati, C., Madras, G., 2010. New empirical expressions to correlate solubilities of solids in supercritical carbon dioxide, Thermochimica Acta. doi:10.1016/j.tca.2009.12.004

Garmroodi, A., Hassan, J., Yamini, Y., 2004. Solubilities of the Drugs Benzocaine, Metronidazole Benzoate, and Naproxen in Supercritical Carbon Dioxide. J. Chem. Eng. Data 49, 709–712. doi:10.1021/je020218w

Gasem, K.A.., Gao, W., Pan, Z., Robinson, R.. (2001) A modified temperature dependence for the Peng–Robinson equation of state. Fluid Phase Equilibria. 181(1–2), 113–125.

Gibbons, R.M., Laughton, A.P. (1984) An equation of state for polar and non-polar substances and mixtures. Journal of the Chemical Society, Faraday Transactions 2. 80(9), 1019.

Goodford, P.J., 1985. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. J. Med. Chem. 28, 849–857. doi:10.1021/jm00145a002

Gordillo, M.., Blanco, M.., Molero, A., Martinez de la Ossa, E., 1999. Solubility of the antibiotic Penicillin G in supercritical carbon dioxide. J. Supercrit. Fluids 15, 183–190. doi:10.1016/S0896-8446(99)00008-X

Graupe, D., 2007. Principles of Artificial Neural Networks, Advanced Series in Circuits and Systems. WORLD SCIENTIFIC. doi:10.1142/8868

Gross, J., Sadowski, G. (2000) Application of perturbation theory to a hard-chain reference fluid: an equation of state for square-well chains. Fluid Phase Equilibria. 168(2), 183–199.

Hadj-Kali, M. (2004) Application de la simulation moléculaire pour le calcul des équilibres liquide - vapeur des nitriles et pour la prédiction des azéotropes.

Hagan, M.T., Demuth, H.B., Beale, M.H. (1995) Neural Network Design. Boston Massachusetts PWS. 2, 734.

Haghbakhsh, R., Hayer, H., Saidi, M., Keshtkari, S., Esmaeilzadeh, F., 2013. Density estimation of pure carbon dioxide at supercritical region and estimation solubility of solid compounds in supercritical carbon dioxide: Correlation approach based on sensitivity analysis. Fluid Phase Equilib. 342, 31–41. doi:10.1016/j.fluid.2012.12.029

Hall, L.H., Kier, L.B., 1977. Structure-activity studies using valence molecular connectivity. J. Pharm. Sci. 66, 642–4.

Hammett, L.P., 1935. Some Relations between Reaction Rates and Equilibrium Constants. Chem. Rev. 17, 125–136. doi:10.1021/cr60056a010

Hansch, C., Fujita, T., 1964. p-σ-π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. J. Am. Chem. Soc. 86, 1616–1626. doi:10.1021/ja01062a035

Hansch, C., Leo, A., Hoekman, D.H., 1995. Exploring QSAR: structure activity relations in chemistry and biology. American Chemical Society, Washington.

Hansch, C., Maloney, P.P., Fujita, T., Muir, R.M., 1962. Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients. Nature 194, 178–180. doi:10.1038/194178b0

Henley, E.J., Seader, J.D., Roper, D.K., 2011. Separation process principles. Wiley.

Heryanto, R., Abdullah, E.C., Hasan, M., 2010. Solubility of Isoniazid in Supercritical Carbon Dioxide. J. Chem. Eng. Data 55, 2306–2309. doi:10.1021/je900622h

Heyen, G. (1983) A New Cubic equation of stae with Extended Range of application. In chemical Eng. Therm. By Newmam,.

Hezave, A.Z., Aftab, S., Esmaeilzadeh, F., 2012. Solubility of sulindac in the supercritical carbon dioxide: Experimental and modeling approach. J. Supercrit. Fluids 68, 39–44. doi:10.1016/j.supflu.2012.04.006

HEZAVE, A.Z., LASHKARBOLOOKI, M., 2013. A NEW SIMPLE CORRELATION FOR CALCULATING SOLUBILITY OF DRUGS IN SUPERCRITICAL CARBON DIOXIDE. J. Theor. Comput. Chem. 12, 1350062. doi:10.1142/S0219633613500624

Hezave, A.Z., Mowla, A., Esmaeilzadeh, F., 2011. Cetirizine solubility in supercritical CO2 at different pressures and temperatures. J. Supercrit. Fluids 58, 198–203. doi:10.1016/j.supflu.2011.05.017

Hezave, A.Z., Shahnazar, S., Rajaei, H., Lashkarbolooki, M., Esmaeilzadeh, F., 2013. Solubility of spironolactone in supercritical carbon dioxide: Experimental and modeling approaches. Fluid Phase Equilib. 355, 130–134. doi:10.1016/j.fluid.2013.07.003

Hojjati, M., Vatanara, A., Yamini, Y., Moradi, M., Najafabadi, A.R., 2009. Supercritical CO2 and highly selective aromatase inhibitors: Experimental solubility and empirical data correlation. J. Supercrit. Fluids 50, 203–209. doi:10.1016/j.supflu.2009.06.015

Hojjati, M., Yamini, Y., Khajeh, M., Vatanara, A., 2007. Solubility of some statin drugs in supercritical carbon dioxide and representing the solute solubility data with several density-based correlations. J. Supercrit. Fluids 41, 187–194. doi:10.1016/j.supflu.2006.10.006

Horsford, E., Baird, S., 1851. Proceedings of the American Assocciation for the Advancement of Science 4th Meeting. American Association for the Ad Science, New Haven, CT. Washington.

Hosseini, M.H., Alizadeh, N., Khanchi, A.R., 2010. Solubility analysis of clozapine and lamotrigine in supercritical carbon dioxide using static system. J. Supercrit. Fluids 52, 30–35. doi:10.1016/j.supflu.2009.11.006

Hozhabr, S.B., Mazloumi, S.H., Sargolzaei, J., 2014. Correlation of solute solubility in supercritical carbon dioxide using a new empirical equation. Chem. Eng. Res. Des. 92, 2734–2739. doi:10.1016/j.cherd.2014.01.026

Huang, Z., Guo, Y. hua, Miao, H., Teng, L. jun, 2014. Solubility of progesterone in supercritical carbon dioxide and its micronization through RESS. Powder Technol. 258, 66–77. doi:10.1016/j.powtec.2014.03.009

Ignacio Medina, Bueno, J.L., 2001. Solubilities of Zopiclone and Nimodipine in Supercritical Carbon Dioxide. doi:10.1021/JE010025T

Ishikawa, T., Chung, W.K., Lu, B.C.-Y. (1980) A cubic perturbed, hard sphere equation of state for thermodynamic properties and vapor-liquid equilibrium calculations. AIChE Journal. 26(3), 372–378.

Ismadji, S., Ju, Y., Soetaredjo, F. E. and Ayucitra, A. (2011) 'Solubility of Azadirachtin in Supercritical Carbon Dioxide at Several Temperatures', pp. 4396–4399.

Ismadji, S., Ju, Y., Soetaredjo, F.E., Ayucitra, A., 2011. Solubility of Azadirachtin in Supercritical Carbon Dioxide at Several Temperatures 4396–4399.

Jafari Nejad, S., Abolghasemi, H., Moosavian, M.A., Maragheh, M.G., 2010. Prediction of solute solubility in supercritical carbon dioxide: A novel semi-empirical model. Chem. Eng. Res. Des. 88, 893–898. doi:10.1016/j.cherd.2009.12.006

Jin, J., Fan, X., Zhang, H., Wang, Y., Zhang, Z., 2014. Solubility of 4-Aminosalicylic Acid in Supercritical Carbon Dioxide and Subcritical 1,1,1,2-Tetrafluoroethane. J. Chem. Eng. Data 59, 2095–2100. doi:10.1021/je5002736

Jodoin, J.F. (1994) Les Réseaux de neurones. Principes et définitions. Persée - Portail des revues scientifiques en SHS.

Johannsen, M., Brunner, G., 1994. Solubilities of the xanthines caffeine, theophylline and theobromine in supercritical carbon dioxide. Fluid Phase Equilib. 95, 215–226. doi:10.1016/0378-3812(94)80070-7

Jouyban, A. (2010) Handbook of Solubility Data for Pharmaceuticals. doi: 10.1017/CBO9781107415324.004.

Jouyban, A., Chan, H.-K., Foster, N.R., 2002. Mathematical representation of solute solubility in supercritical carbon dioxide using empirical expressions. J. Supercrit. Fluids 24, 19–35. doi:10.1016/S0896-8446(02)00015-3

Julián García-González, María J. Molina, Francisco Rodríguez, * and, Mirada, F., 2001. Solubilities of Phenol and Pyrocatechol in Supercritical Carbon Dioxide. doi:10.1021/JE0003795

Kabadi, V.N., Danner, R.P. (1985) A modified Soave-Redlich-Kwong equation of state for water-hydrocarbon phase equilibria. Industrial & Engineering Chemistry Process Design and Development. 24(3), 537–541.

Karelson, M. (2000) Molecular descriptors in QSAR/QSPR. Wiley-Interscience.

Ke, J., Mao, C., Zhong, M., Han, B., Yan, H., 1996. Solubilities of salicylic acid in supercritical carbon dioxide with ethanol cosolvent. J. Supercrit. Fluids 9, 82–87. doi:10.1016/S0896-8446(96)90002-9

Keshmiri, K., Vatanara, A., Yamini, Y., 2014. Development and evaluation of a new semi-empirical model for correlation of drug solubility in supercritical CO2. Fluid Phase Equilib. 363, 18–26. doi:10.1016/j.fluid.2013.11.013

Kevin M. Passino, 2005. Biomimicry for Optimization, Control, and Automation. Springer-Verlag, London. doi:10.1007/b138169

Khamda, M., Hosseini, M.H., Rezaee, M., 2012. Measurement and correlation solubility of cefixime trihydrate and oxymetholone in supercritical carbon dioxide (CO 2). J. Supercrit. Fluids 73, 130–137. doi:10.1016/j.supflu.2012.09.006

Khimeche, K., Alessi, P., Kikic, I., Dahmani, A., 2007. Solubility of diamines in supercritical carbon dioxide. Experimental determination and correlation. J. Supercrit. Fluids 41, 10–19. doi:10.1016/j.supflu.2006.09.004

Knez, Z., Skerget, M., Sencar-Bozic, P., Rizner, A., 1995. Solubility of Nifedipine and Nitrendipine in Supercritical CO2. J. Chem. Eng. Data 40, 216–220. doi:10.1021/je00017a045

Ko, M., Shah, V., Bienkowski, P.R., Cochran, H.D., 1991. Solubility of the antibiotic penicillin V in supercritical CO2. J. Supercrit. Fluids 4, 32–39. doi:10.1016/0896-8446(91)90028-5

Kotnik, P., Škerget, M., Knez, Z., 2011. Solubility of Nicotinic Acid and Nicotinamide in Carbon Dioxide at T = (313.15 to 373.15) K and p = (5 to 30) MPa: Experimental Data and Correlation. J. Chem. Eng. Data 56, 338–343. doi:10.1021/je100697a

Kumar, S.K., Johnston, K.P. (1988) Modelling the solubility of solids in supercritical fluids with density as the independent variable. The Journal of Supercritical Fluids. 1(1), 15–22.

Kumar, S.K., Johnston, K.P., 1988. Modelling the solubility of solids in supercritical fluids with density as the independent variable. J. Supercrit. Fluids 1, 15–22. doi:10.1016/0896-8446(88)90005-8

Kurihara, K., Tochigi, K., Kojima, K. (1987) and Residual Excess Free Energy. . 20(3), 227–231.

Kwak, T.Y., Mansoori, G.A. (1986) Van der waals mixing rules for cubic equations of state. Applications for supercritical fluid extraction modelling. Chemical Engineering Science. 41(5), 1303–1309.

Lafferrere, L. (2002) ET TRANSITIONS DE PHASES Par.

Lashkarbolooki M, Hezave A, Rahnama Y, Ozlati R Rajaei H, E.F., 2013. Solubility of cyproheptadine in supercritical carbon dioxide; experimental and modeling approaches. J. Supercrit. Fluids.

Lee, C.-A., Tang, M., Chen, Y.-P., 2014a. Measurement and correlation for the solubilities of cinnarizine, pentoxifylline, and piracetam in supercritical carbon dioxide. Fluid Phase Equilib. 367, 182–187. doi:10.1016/j.fluid.2014.01.041

Lee, C.-A., Tang, M., Ho, S.-L., Chen, Y.-P., 2014b. Solubilities of chlormezanone, metaxalone and methocarbamol in supercritical carbon dioxide. J. Supercrit. Fluids 85, 11–16. doi:10.1016/j.supflu.2013.10.006

Lejewski, C. (1967) A single axiom for the mereological notion of proper part. Notre Dame Journal of Formal Logic. 8(4), 279–285.

Li, H., Jia, D., Zhu, Q., Shen, B., 2015. Determination, correlation and prediction of the solubilities of niflumic acid, clofenamic acid and tolfenamic acid in supercritical CO2. Fluid Phase Equilib. 392, 95–103. doi:10.1016/j.fluid.2015.02.007

Li, Y., Tang, Z., Jin, J., Zhang, Z., 2013. Binary and ternary solubility of amino- and nitro-benzoic acids in supercritical carbon dioxide. Fluid Phase Equilib. 344, 71–78. doi:10.1016/j.fluid.2013.01.021

Lucas, P., van der Gaag, L. (1991) Principles of Expert Systems. , 426.

MacQuarrie, D. (1976) Statistical Mechanics. Harper & Row, New York.

Mathias, P.M. (1983) A versatile phase equilibrium equation of state. Industrial & Engineering Chemistry Process Design and Development. 22(3), 385–391.

Mathias, P.M., Copeman, T.W. (1983) Extension of the Peng-Robinson equation of state to complex mixtures: Evaluation of the various forms of the local composition concept. Fluid Phase Equilibria. 13, 91–108.

Melhem, A, G., Saini, R., Goodwing, B, M. (1989) A Modified Ping-Robinson Equation of State. Fluid Phase Equilibria, 47.

Méndez-Santiago, J., Teja, A.S. (1999) The solubility of solids in supercritical fluids. Fluid Phase Equilibria. 158–160, 501–510.

Méndez-Santiago, J., Teja, A.S., 1999. The solubility of solids in supercritical fluids. Fluid Phase Equilib. 158–160, 501–510. doi:10.1016/S0378-3812(99)00154-5

Meyer, H., 1899. Zur Theorie der Alkoholnarkose. Arch. für Exp. Pathol. und Pharmakologie 42, 109–118. doi:10.1007/BF01834479

Mhaskar, H.N., Micchelli, C.A. (1994) Dimension-independent bounds on the degree of approximation by neural networks. IBM Journal of Research and Development. 38(3), 277–284.

Mukhopadhyay, M., Rao, G.V.R. (1993) Thermodynamic modeling for supercritical fluid process design. Industrial & Engineering Chemistry Research. 32(5), 922–930.

Müller, E. a, Gubbins, K.E. (2001) Molecular-based equations of state for associating fluids: A review of SAFT and related approaches. Industrial & Engineering Chemistry Research. 40, 2193–2211.

Overton, E., 1899. Ueber die allgemeinen osmotischen Eigenschaften der Zelle, ihre vermutlichen Ursachen und ihre Bedeutung für die Physiologie. S.l.

Panagiotopoulos, A.Z., Reid, R.C. (1986) New Mixing Rule for Cubic Equations of State for Highly Polar, Asymmetric Systems. In pp. 571–582.

Park, C.I., Shin, M.S., Kim, H., 2009. Solubility of climbazole and triclocarban in supercritical carbon dioxide: Measurement and correlation. J. Chem. Thermodyn. 41, 30–34. doi:10.1016/j.jct.2008.08.009

Paula, J.T., Sousa, I.M.O., Foglio, M.A., Cabral, F.A., 2016. Solubility of protocatechuic acid, sinapic acid and chrysin in supercritical carbon dioxide. J. Supercrit. Fluids 112, 89–94. doi:10.1016/j.supflu.2016.02.014

Pelikan EV, 1855. Opyt prilozheniya sovremennykh fizilo-khimischeskikh issledovanii k ucheniyu o yadakh (Experience with the application of modern physiochemical concepts to the study of poisons).

Peng, D.-Y., Robinson, D.B. (1976) A New Two-Constant Equation of State. Industrial & Engineering Chemistry Fundamentals. 15(1), 59–64.

Platt, J.R., 1947. Influence of Neighbor Bonds on Additive Bond Properties in Paraffins. J. Chem. Phys. 15, 419–420. doi:10.1063/1.1746554

Prausnitz, J.M., Lichtenthaler, R.N., Azevedo, E.G. de (1986a) Molecular thermodynamics of fluid-phase equilibria. Prentice-Hall.

Prausnitz, J.M., Lichtenthaler, R.N., Azevedo, E.G. de (1986b) Molecular thermodynamics of fluid-phase equilibria. Prentice-Hall.

Raimondi, L. (1980) A modified Redlich-Kwong equation of state for vapour-liquid equilibrium calculations. Chemical Engineering Science. 35(6), 1269–1275.

Randy D. Weinstein, *, Joseph J. Gribbin, and, Muske, K.R., 2004. Solubility and Salting Behavior of Several β-Adrenergic Blocking Agents in Liquid and Supercritical Carbon Dioxide. doi:10.1021/JE0497202

Redlich, O., Kwong, J.N.S. (1949) On the Thermodynamics of Solutions. V. An Equation of State. Fugacities of Gaseous Solutions. Chemical Reviews. 44(1), 233–244.

Reveco-Chilla, A.G., Cabrera, A.L., de la Fuente, J.C., Zacconi, F.C., del Valle, J.M., Valenzuela, L.M., 2016. Solubility of menadione and dichlone in supercritical carbon dioxide. Fluid Phase Equilib. 423, 84–92. doi:10.1016/j.fluid.2016.04.001

Richardson, B., 1869. Physiological research on alcohols. Med Times Gazzette.

Richet, C., 1893. On the relationship between the toxicity and the physical properties of substances. Compt Rendus Seances Soc Biol.

Rosa, A.B.S., Marceneiro, S., Braga, M.E.M., Dias, A.M. a., de Sousa, H.C., 2015. Solubility of all-trans retinoic acid in supercritical carbon dioxide. J. Supercrit. Fluids 98, 70–78. doi:10.1016/j.supflu.2014.12.027

Roy, K., Kar, S., Das, R.N., 2015. Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment, Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment. doi:10.1016/B978-0-12-801505-6.00001-6

Seidell, A., 1912. A new bromine method for the determination of thymol, salicylates, and similar compounds. Am Chem J 47, 508–526.

Shojaee, S.A., Hezave, A.Z., Aftab, S., Lashkarbolooki, M., Esmaeilzadeh, F., 2013a. Solubility of gabapentin in supercritical carbon dioxide. J. Supercrit. Fluids 78, 1–6. doi:10.1016/j.supflu.2013.02.003

Shojaee, S.A., Rajaei, H., Hezave, A.Z., Lashkarbolooki, M., Esmaeilzadeh, F., 2013b. Experimental measurement and correlation for solubility of piroxicam (a non-steroidal anti-inflammatory drugs (NSAIDs)) in supercritical carbon dioxide. J. Supercrit. Fluids 80, 38–43. doi:10.1016/j.supflu.2013.03.015

Simon, Z., 1974. Specific Interactions. Intermolecular Forces, Steric Requirements, and Molecular Size. Angew. Chemie Int. Ed. English 13, 719–727. doi:10.1002/anie.197407191

Si-Moussa, C., Belghait, A., Khaouane, L., Hanini, S., Halilali, A., 2016. Novel density-based model for the correlation of solid drugs solubility in supercritical carbon dioxide. Comptes Rendus Chim. doi:10.1016/j.crci.2016.09.009

Soave, G. (1972) Equilibrium constants from a modified Redlich-Kwong equation of state. Chemical Engineering Science. 27(6), 1197–1203.

Soave, G.S. (1979) Application of the redlich-kwong-soave equation of state to solid-liquid equilibria calculations. Chemical Engineering Science. 34(2), 225–229.

Sparks, D.L., Hernandez, R., Estévez, L.A., 2008. Evaluation of density-based models for the solubility of solids in supercritical carbon dioxide and formulation of a new model. Chem. Eng. Sci. 63, 4292–4301. doi:10.1016/j.ces.2008.05.031

Sparks, D.L., Hernandez, R., Estévez, L.A., Meyer, N., French, T., 2007. Solubility of Azelaic Acid in Supercritical Carbon Dioxide. J. Chem. Eng. Data 52, 1246–1249. doi:10.1021/je600572z

Stahl, E., Schilz, W., Schütz, E., Willing, E., 1978. A Quick Method for the Microanalytical Evaluation of the Dissolving Power of Supercritical Gases. Angew. Chemie Int. Ed. English 17, 731–738. doi:10.1002/anie.197807311

Stryjek, R., Vera, J.H. (1986) PRSV: An improved peng-Robinson equation of state for pure compounds and mixtures. The Canadian Journal of Chemical Engineering. 64(2), 323–333.

Su, C.-S., Chen, Y.-P., 2008. Measurement and correlation for the solid solubility of non-steroidal anti-inflammatory drugs (NSAIDs) in supercritical carbon dioxide. J. Supercrit. Fluids 43, 438–446. doi:10.1016/j.supflu.2007.08.006

Suleiman, D., Estévez, L.A., Pulido, J.C., García, J.E., Mojica, C., 2005. Solubility of Anti-Inflammatory, Anti-Cancer, and Anti-HIV Drugs in Supercritical Carbon Dioxide. J. Chem. Eng. Data 50, 1234–1241. doi:10.1021/je049551l

Sung, H.-D., Shim*, J.-J., 1999. Solubility of C. I. Disperse Red 60 and C. I. Disperse Blue 60 in Supercritical Carbon Dioxide. doi:10.1021/JE990018T

Tabernero, A., del Valle, E.M.M., Galán, M.Á., 2010. A comparison between semiempirical equations to predict the solubility of pharmaceutical compounds in supercritical carbon dioxide. J. Supercrit. Fluids 52, 161–174. doi:10.1016/j.supflu.2010.01.009

Taft, R.W., 1952. Linear Free Energy Relationships from Rates of Esterification and Hydrolysis of Aliphatic and Ortho-substituted Benzoate Esters. J. Am. Chem. Soc. 74, 2729–2732. doi:10.1021/ja01131a010

Tang, C., Guan, Y.-X., Yao, S.-J., Zhu, Z.-Q., 2014. Solubility of Dexamethasone in Supercritical Carbon Dioxide with and without a Cosolvent. J. Chem. Eng. Data 59, 3359–3364. doi:10.1021/je500345n

Tang, Z., Jin, J., Zhang, Z., Liu, H., 2012. New Experimental Data and Modeling of the Solubility of Compounds in Supercritical Carbon Dioxide. Ind. Eng. Chem. Res. 51, 5515–5526. doi:10.1021/ie2016224

Todeschini, R., Consonni, V., Gramatica, P., 2009. Chemometrics in QSAR, in: Comprehensive Chemometrics. Elsevier, pp. 129–172. doi:10.1016/B978-044452701-1.00007-7

Traube, J., 1904. Theorie der Osmose und Narkose. Pfläger, Arch. für die Gesammte Physiol. des Menschen und der Thiere 105, 541–558. doi:10.1007/BF01682827

Tsai, C.C., Lin, H.M., Lee, M.J., 2014. Solubility of niflumic acid and celecoxib in supercritical carbon dioxide. J. Supercrit. Fluids 95, 17–23. doi:10.1016/j.supflu.2014.07.026

Tumakaka, F., Gross, J., Sadowski, G. (2005) Thermodynamic modeling of complex systems using PC-SAFT. Fluid Phase Equilibria. 228, 89–98.

Tute MS, 1990. History and objectives of quantitative drug design. In: Hansch C, Sammes PG, Taylor JB, editors. Comprehensive medicinal chemistry, vol .4. ed. Oxford: Pergamon Press.

Twu C, H. (1988) A Modified Redlich-Kwong Equation of state for highly polar. Acadimic Periodical Press : Beijing, China, 148.

Usdin, E., McAuliffe, J.C. (1976) A one parameter family of equations of state. Chemical Engineering Science. 31(11), 1077–1084.

Valderrama, J. (1990) A generalized Patel-Teja equation of state for polar and nonpolar fluids and their mixtures. Journal of chemical engineering of Japan.

Valderrama, J.O. (2003) The State of the Cubic Equations of State. Industrial & Engineering Chemistry Research. 42(8), 1603–1618.

Vatanara, A., Najafabadi, A.R., Khajeh, M., Yamini, Y., 2005. Solubility of some inhaled glucocorticoids in supercritical carbon dioxide. J. Supercrit. Fluids 33, 21–25. doi:10.1016/j.supflu.2004.05.002

Wertheim, M.S. (1984) Fluids with highly directional attractive forces. I. Statistical thermodynamics. Journal of Statistical Physics. 35(1–2), 19–34.

Wiener, H., 1947. Structural Determination of Paraffin Boiling Points. J. Am. Chem. Soc. 69, 17–20. doi:10.1021/ja01193a005

Wong, D.S.H., Sandler, S.I. (1992) A theoretically correct mixing rule for cubic equations of state. AIChE Journal. 38(5), 671–680.

Xu, X., Tan, Q., Wang, J., Zhao, X., Tan, L. (2007) A gigantic bird-like dinosaur from the Late Cretaceous of China. Nature. 447(7146), 844.

Yadollah Yamini, *, Jalal Hassan, and, Haghgo, S., 2001. Solubilities of Some Nitrogen-Containing Drugs in Supercritical Carbon Dioxide. doi:10.1021/JE000286N

Yamini, Y., Arab, J., Asghari-khiavi, M., 2003. Solubilities of phenazopyridine, propranolol, and methimazole in supercritical carbon dioxide. J. Pharm. Biomed. Anal. 32, 181–187. doi:10.1016/S0731-7085(03)00016-5

Yamini, Y., Hojjati, M., Kalantarian, P., Moradi, M., Esrafili, A., Vatanara, A., 2012. Solubility of capecitabine and docetaxel in supercritical carbon dioxide: Data and the best correlation. Thermochim. Acta 549, 95–101. doi:10.1016/j.tca.2012.09.010

Yamini, Y., Kalantarian, P., Hojjati, M., Esrafily, A., Moradi, M., Vatanara, A., Harrian, I., 2010. Solubilities of Flutamide, Dutasteride, and Finasteride as Antiandrogenic Agents, in Supercritical Carbon Dioxide: Measurement and Correlation. J. Chem. Eng. Data 55, 1056–1059. doi:10.1021/je900520a

Yamini, Y., Moradi, M., 2011. Measurement and correlation of antifungal drugs solubility in pure supercritical CO2 using semiempirical models. J. Chem. Thermodyn. 43, 1091–1096. doi:10.1016/j.jct.2011.02.020

Yamini, Y., Tayyebi, M., Moradi, M., Vatanara, A., 2013. Solubility of megestrol acetate and levonorgestrel in supercritical carbon dioxide. Thermochim. Acta 569, 48–54. doi:10.1016/j.tca.2013.07.018

Yongyue Sun, Shufen Li, *, and, Quan†, C., 2005. Solubility of Ferulic Acid and Tetramethylpyrazine in Supercritical Carbon Dioxide. doi:10.1021/JE049715Y

Yu, Z.-R., Singh, B., Rizvi, S.S.H., Zollweg, J.A., 1994. Solubilities of fatty acids, fatty acid esters, triglycerides, and fats and oils in supercritical carbon dioxide. J. Supercrit. Fluids 7, 51–59. doi:10.1016/0896-8446(94)90006-X

Zabihi, F., Mirzajanzadeh, M., Jia, J., Zhao, Y., 2016. Measurement and Calculation of Solubility of Quinine in Supercritical Carbon Dioxidez. Chinese J. Chem. Eng. doi:10.1016/j.cjche.2016.10.003

Zabihi, F., Vaziri, A., Akbarnejad, M. M., Ardjmand, M., Otadi, M. and Bozorgmanesh, A. R. (2010) 'A novel mathematical method for prediction of rapid expansion of supercritical solution (RESS) processed ibuprofen powder size distribution', Korean Journal of Chemical Engineering. Springer US, 27(5), pp. 1601–1605. doi: 10.1007/s11814-010-0265-9.

Zadeh, L. a. (1965) Fuzzy sets. Information and Control. 8(3), 338–353.

Zeinolabedini Hezave, A., Esmaeilzadeh, F., 2012. Solubility Measurement of Diclofenac Acid in the Supercritical CO2. J. Chem. Eng. Data 57, 1659–1664. doi:10.1021/je200012x

Zeinolabedini Hezave, A., Khademi, M.H., Esmaeilzadeh, F., 2012. Measurement and modeling of mefenamic acid solubility in supercritical carbon dioxide. Fluid Phase Equilib. 313, 140–147. doi:10.1016/j.fluid.2011.09.031

Zhan, S., Cui, L., Zhao, Q., Chen, L., Wang, J., Chen, S., Ding, S., 2015. Measurement and Correlation of Solubility of Nitrendipine in Supercritical Carbon Dioxide With and Without Ethanol Cosolvent. J. Solution Chem. 44, 1–15. doi:10.1007/s10953-014-0262-4

Zhan, S., Li, S., Zhao, Q., Wang, W., Wang, J., 2017. Measurement and Correlation of Curcumin Solubility in Supercritical Carbon Dioxide. J. Chem. Eng. Data 62, 1257–1263. doi:10.1021/acs.jced.6b00798

Zhen Huang, Wei D. Lu, Sibudjing Kawi, and, Yee C. Chiew*, 2004. Solubility of Aspirin in Supercritical Carbon Dioxide with and without Acetone. doi:10.1021/JE0499465

Zhu, J., Chang, C., Wu, H. and Jin, J. (2015) 'Solubility of polyvinyl alcohol in supercritical carbon dioxide and subcritical 1,1,1,2-tetrafluoroethane', Fluid Phase Equilibria, 404, pp. 61–69. doi: 10.1016/j.fluid.2015.06.031.