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

UNIVERSITE YAHIA FARÈS DE MEDEA
Département de Génie des Procédés et Environnement
Laboratoire de Biomatériaux et Phénomènes de Transport (LBMPT)

THESE DE DOCTORAT EN SCIENCES

Présentée par :

Mr. MOUSSAOUI Mohammed

Ingénieur d'état en Génie Mécanique option Energétique : Université de Médéa, Algérie
Magister en Energétique Numérique : Université de Médéa, Algérie

En vue de l'obtention du titre de ***Doctorat en sciences***
Filière : ***Génie des Procédés***

Thème

Etude et modélisation de quelques propriétés contrôlant des phénomènes chimiques et environnementales

Soutenue le 11/02/2021 devant le jury composé de :

Mme Latifa KHAOUANE	Maitre de Conférences A, Université de Médéa	Présidente
M. Ahmed AZIZI	Maitre de Conférences A, Université de Laghouat	Examinateur
M. Brahim MAHFOUD	Maitre de Conférences A, Université de Bouira	Examinateur
M. Saber HAMIMID	Maitre de Conférences A, Université de Bouira	Examinateur
M. Salah HANINI	Professeur, Université de Médéa	Directeur de thèse
M. Maamar LAIDI	Maitre de Conférences A, Université de Médéa	Co-Directeur de thèse

Remerciements

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

J'adresse mes remerciements à Monsieur ***Salah HANINI****, Professeur à l'université de Médéa, pour m'avoir encadré et conseillé pendant toute la réalisation de ce travail. Son aide appréciable, ses encouragements qui m'ont permis d'avancer dans l'élaboration de ce travail et à qui je témoigne ma profonde reconnaissance.*

J'adresse également mes plus sincères remerciements à Monsieur ***Maamar LAIDI****, Maître de Conférences (A) à l'université de Médéa, pour avoir co-encadré ce travail. Je vous remercie pour votre gentillesse et votre soutien permanent. Il m'a toujours offert son soutien. Son savoir-faire, mais surtout il était toujours présent à tout moment et il a bien veillé pour le bon déroulement de ce travail.*

Mes remerciements les plus sincères vont aussi à M^{me} ***Latifa KHAOUANE****, Maître de Conférences (A) à l'université de Médéa, pour l'honneur qu'elle m'a fait en acceptant de présider le jury de cette thèse.*

Mes vifs remerciements vont à Monsieur ***Ahmed AZIZI****, Maître de Conférences (A) à l'université de Laghouat, d'avoir accepté d'examiner ce travail, je suis très honoré de sa présence parmi le jury de cette thèse.*

Je tiens à remercier vivement Monsieur ***Brahim MAHFOUD****, Maître de Conférences (A) à l'université de Bouira, d'avoir accepté d'examiner ce travail, je suis très honoré de sa présence parmi le jury de cette thèse.*

Mes vifs remerciements vont également à Monsieur ***Saber HAMIMID****, Maître de Conférences (A) à l'université de Bouira, d'avoir accepté d'examiner ce travail, je suis très honoré de sa présence parmi le jury de cette thèse.*

Merci à tous ceux qui m'ont aidé de près ou de loin durant ces années de recherche. Je ne peux conclure sans un mot de remerciement à tous mes collègues.

MOUSSAOUI Mohammed

Dédicaces

Je dédie ce travail ***A*** *mes chers parents,*

A *ma femme,*

A *mon fils : Abdeldjalil,*

A *mes frères et à mes sœurs,*

A *ma famille,*

A *mes collègues au Université Dr. Yahia Farès de Médéa et Université Akli Mohand Olhadj de Bouira,*

A *tous ceux qui m'ont soutenu de près ou de loin.*

MOUSSAOUI Mohammed

الملخص: دراسة ونمذجة بعض الخصائص التي تتحكم في الظواهر الكيميائية والبيئية

تستخدم الموائع فوق الحرجة بشكل عام في عمليات الفصل والاستخلاص المطبقة في الصناعات الغذائية، الكيميائية، الدوائية والصناعات الأخرى. يعتمد تصميم وتحسين عمليات الاستخلاص، الفصل والتنقية بشكل أساسي على معرفة الخصائص فوق الحرجة و قابلية ذوبان المواد الصلبة في الموائع فوق الحرجة. القياس التجريبي للخاصائص الحرارية الفيزيائية لهذه المركبات في الموائع فوق الحرجة صعب ومكلف. لتجنب التجارب المكلفة، نحتاج لتطوير نماذج تنبؤ مرنة وقوية لتقدير هذه الخصائص.

يهدف الجزء الأول من هذا العمل إلى نمذجة الخصائص الحرجة لـ 6700 مركب نقي بناءً على خمسة 5 مواصفات لبنية مادة المركب و اثنان 2 من خصائص الديناميكية الحرارية. لذلك ، تم استخدام أربع طرق وهي (MLR)، (ANN)، (SVM) و(SVM-DA) أظهرت النتائج أن النموذج SVM-DA المطور أكثر دقة مقارنة بالنماذج الأخرى .

وإستندت مساهمتنا الثانية في إقتراح نموذج جديد مبني على كثافة المذيب بثمانية معاملات لحساب بيانات ذوبان المواد الصلبة في ثاني أكسيد الكربون فوق الحرج. إستندنا في دراستنا على مجموعة شاملة من بيانات الذوبان التجريبية لـ 130 من المواد الصلبة المذوبة في ثاني أكسيد الكربون فوق الحرج ممثلة بـ 3269 قيمة تجريبية. بناءً على نتائج المقارنة مع خمسة 05 نماذج تجريبية المتحصل عليها في أبحاث سابقة من حيث الخطأ النسبي المطلق ، تحصلنا على أن النموذج المقترح هو الأفضل إلى حد بعيد مع متوسط قيمة الخطأ النسبي المطلق 7.42٪.

في الجزء الثالث من هذه الدراسة ، تم نمذجة قابلية ذوبان المواد الصلبة في ثاني أكسيد الكربون فوق الحرج (CO_2-SC) باستخدام تقنيات الذكاء الحسابي بناءً على نماذج العلاقة الكمية بين البنية والخصائص (QSPR). في هذه الدراسة تم جمع قاعدة بيانات تحتوي على 3637 من القيم التجريبية لذوبان 145 مادة صلبة في (CO_2-SC) من البحوث العلمية المنشورة سابقًا. من خلال هذه الدراسة تم الحصول على نموذج SVM-QSPR بمتوسط الخطأ النسبي المطلق الكلي AARD بحوالي 1.345٪ وهو أفضل من ذلك الذي تم الحصول عليه باستخدام نموذج ANN-QSPR بـ 2.772٪. كما أظهرت النتائج أن نموذج SVM-QSPR المطور أكثر دقة ويمكن استخدامه كأداة نمذجة قوية بديلة للدراسات السابقة حول قابلية ذوبان المواد الصلبة في ثاني أكسيد الكربون فوق الحرج.

كلمات مفتاحية: قابلية الذوبان، ثاني أكسيد الكربون فوق حرج، النمذجة، الذكاء الحسابي.

Résumé: Etude et modélisation de quelques propriétés contrôlant des phénomènes chimiques et environnementales

Les fluides supercritiques sont très utilisés dans les procédés de séparation et d'extraction appliqués dans les industries alimentaire, chimique, pharmaceutique et autres. La conception et l'amélioration des procédés d'extraction, de séparation et de purification dépendent principalement de la connaissance des propriétés supercritiques et de la solubilité des solides dans les fluides supercritiques (FSC). La mesure expérimentale des propriétés thermochimiques de ces composés en FSC est laborieuse et coûteuse. Pour éviter des expérimentations coûteuses, nous devons développer des modèles de prédiction flexibles et robustes pour estimer ces caractéristiques.

La première partie de ce travail vise à modéliser les propriétés critiques de 6700 composés purs à partir de cinq 5 descripteurs moléculaires et de deux 2 propriétés thermodynamiques. Par conséquent, quatre méthodes ont été utilisées, à savoir la régression linéaire multiple (MLR), réseaux de neurones artificiels (RNAs), machine à vecteurs de support (SVM) et machine à vecteurs de support optimisé par l'algorithme Dragonfly (SVM-DA). Les résultats ont montré que le modèle SVM-DA a donné une meilleure capacité de prédiction par rapport aux autres modèles.

L'objectif de la deuxième contribution est la proposition d'un nouveau modèle semi-empirique avec huit paramètres. Ce modèle est développé en utilisant une large base de données de solubilité expérimentale contenant 130 solides dans le CO_2 supercritique représentés par 3269 points de données expérimentaux disponibles dans la littérature. La comparaison en termes d'erreur relative absolue (ERA), montre que le modèle proposé est le meilleur avec une valeur de l'ERA moyenne de 7.42% par rapport aux meilleurs cinq modèles proposés dans la littérature.

Dans la troisième partie, de cette étude, la solubilité de 145 solides dans le CO_2 supercritique (CO_2-SC) a été modélisée à l'aide des techniques d'intelligence computationnelle basées sur les modèles de relation quantitative structure/propriétés (QSPR). Une base de données contenant 3637 valeurs expérimentales de solubilité a été recueillie à partir des articles publiés antérieurement. L'erreur relative absolue moyenne totale (ERAM) obtenue par le modèle SVM-QSPR a été d'environ 1.345%, ce qui est meilleur que celui obtenu avec le modèle RNA-QSPR de 2.772%. Les résultats montrent que l'approche SVM-QSPR développé dans ce travail s'avère très efficace pour reproduire avec une très bonne précision la solubilité des solutés solides dans CO_2-SC d'une manière simple, fiable et robuste. Elle constitue ainsi une alternative crédible aux modèles classiques.

Mots clés: Solubilité, Fluides supercritiques, Semi-empirique, Modélisation, Réseau de neurones, Support vecteurs machines, QSPR.

Abstract: Study and modeling of some properties controlling chemical and environmental phenomena.

Supercritical fluids are widely used in separation and extraction processes applied in food, chemical, pharmaceutical and other industries. The design and improvement of extraction, separation and purification processes depend primarily on knowledge of supercritical properties and solubility of solids in supercritical fluids (FSC). The experimental measurement of the thermochemical properties of these compounds in FSC is laborious and expensive. To avoid costly experiments, we need to develop flexible and robust prediction models to estimate these characteristics.

The first part of this work aims to model the critical properties of 6700 pure compounds from 5 molecular descriptors and 2 thermodynamic properties. Therefore, four methods were used namely multiple linear regressions (MLR), artificial neural networks (ANNs), support vector machines (SVM) and support vector machines optimized with Dragonfly algorithm (SVM-DA). Results revealed that SVM-DA showed better prediction compared to other models.

In the second contribution, a new model with eight parameters was proposed. This model was tested using large data set of experimental solubility including 130 solids in supercritical CO_2 represented by 3269 experimental data points. And when comparing the overall correlation of performance in terms of absolute relative deviation *ARD*, the proposed model was the best with an average value of 7.42%.

In the third part of this study, the solubility of 145 solids in supercritical CO_2 ($scCO_2$) was modeled using computational intelligence techniques based on quantitative structure property relationship (QSPR) models. A database containing 3637 solubility values was collected from previously published articles. The total average absolute relative deviation *AARD* obtained with the SVM-QSPR model of 1.345%, which is better than that obtained with the ANN-QSPR model of 2.772%. The results show that The SVM-QSPR approach used in this work has been more precise and very effective for simple, reliable and robust calculations of solubility in supercritical CO_2. It thus constitutes a credible alternative to conventional models.

Keywords: Solubility, Supercritical fluids, Semi-empirical, Modeling, Neural network, Support vector machine, QSPR.

Sommaire

Sommaire		I
Abréviations		IV
Nomenclature		V
Liste des figures		VI
Liste des tableaux		IX
Introduction générale		1
Chapitre I.		
La solubilité dans les fluides supercritiques		
I.1.	Introduction	4
I.2.	Définition d'un fluide supercritique	4
I.2.1.	Description de la zone supercritique	5
I.2.2.	Propriétés physico-chimiques des fluides supercritiques	7
I.2.3.	Cas particulier du CO_2-SC	8
I.3.	Techniques expérimentales pour mesurer la solubilité dans CO_2-SC	9
I.3.1.	Les méthodes statiques	9
I.3.1.1.	Les méthodes analytiques	10
I.3.1.2.	Les méthodes synthétiques	11
I.3.1.3.	Les méthodes gravimétriques	12
I.3.1.4.	Les méthodes dynamiques	13
I.4.	Méthodes de Modélisation de la solubilité des solutés solides dans les fluides supercritiques	14
I.4.1.	Modélisation par les équations d'état	14
I.4.1.1.	Forme générale	15
I.4.1.2.	L'équation d'état de Viriel	16
I.4.1.3.	Les équations d'état cubiques	17
I.4.1.4.	Les équations d'état complexes	20
I.4.2.	Modélisation par les modèles semi-empiriques	21
I.4.2.1.	Modèle de Chrastil	23
I.4.2.2.	Modèle d'Adachi et Lu	24
I.4.2.3.	Modèle de Del Valle et Aguilira	24
I.4.2.4.	Modèle de Kumar et Johnston	25
I.4.2.5.	Modèle de Bartle et al.	25
I.4.2.6.	Modèle de Gordillo et al.	26
I.4.2.7.	Modèle de Mendez-Santiago et Teja	26
I.4.2.8.	Modèle de Sung et Shim	27
I.4.2.9.	Modèle de Jouyban et al.	27
I.4.2.10.	Modèle de Sparks et al.	28
I.4.2.11.	Modèle de Garlapati et Madras	28
I.4.2.12.	Modèle de Garlapati et Madras	29
I.4.2.13.	Modèle de Jafari Nedjad et al.	29
I.4.2.14.	Modèle de Ch et Madras	29
I.4.2.15.	Modèle de Bian et al.	29
I.4.2.16.	Modèle de Keshmiri et al.	30
I.4.2.17.	Modèle d'Amooey	30
I.4.2.18.	Modèle de Hozhabr et al.	30
I.4.2.19.	Modèle de Khansary et al.	31
I.4.2.20.	Modèle de Bian et al.	31
I.4.3.	Modélisation par les méthodes d'intelligence computationnelle	31
I.4.3.1.	La modélisation par les RNA	31
I.4.3.2.	Méthode des machines à vecteurs de support (SVM)	34
I.4.3.3.	La modélisation par le système d'inférence adaptative neuro-flou (ANFIS)	35
I.5.	Conclusion	37

Chapitre II.
Techniques d'intelligence computationnelle & Chemoinformatiques

II.1.	Introduction	38
II.2.	Utilités et applications des méthodes QSPR	38
II.3.	Outils de QSPR	39
II.3.1.	Les descripteurs moléculaires	39
II.3.2.	Formulation mathématique des QSPR	41
II.3.3.	Logiciels et interfaces utilisés	41
II.4.	Réseaux de neurones artificiels	42
II.4.1.	Analogie entre neurone biologique et neurone artificiel	43
II.4.1.1.	Le neurone biologique	43
II.4.1.2.	Réseau de neurones artificiels	44
II.4.2.	Concept du réseau de neurones de type perceptron multicouches	46
II.4.3.	Fonction d'activation	47
II.4.4.	Classification des réseaux de neurones	49
II.4.4.1.	Classification selon la description architecturale	49
II.4.4.2.	Classification selon les types d'apprentissages	51
II.4.5.	Procédure d'apprentissage	52
II.4.5.1.	Algorithme de la rétro-propagation	52
II.4.5.2.	Critère d'arrêt pour l'apprentissage	55
II.4.6.	Procédure de modélisation neuronale	55
II.4.6.1.	Sélection des données	55
II.4.6.2.	Collecte des données	56
II.4.6.3.	Séparation des bases de données	56
II.4.6.4.	Normalisation et traitement des données	56
II.4.6.5.	Choix du type de réseau de neurones	57
II.4.6.6.	Choix du nombre de neurones dans la couche cachée	57
II.4.6.7.	Choix de la fonction de transfert	57
II.4.6.8.	Estimation des paramètres ou apprentissage	57
II.4.6.9.	Validation et généralisation	57
II.5.	Adaptive Neuro-Fuzzy Inference System (ANFIS)	58
II.5.1.	Fuzzy systems	59
II.5.2.	Architecture et algorithme	60
II.6.	Machines à Vecteurs de Support (SVM)	62
II.6.1.	Idée de base de la théorie de l'apprentissage statistique	63
II.6.2.	Apprentissage et généralisation	64
II.6.3.	Régression à vecteurs de Support (SVR)	64
II.6.4.	Les fonctions de noyau	66
II.7.	Conclusion	68

Chapitre III.
Contribution à la modélisation des propriétés critiques par l'intelligence computationnelle

III.1.	Introduction	69
III.2.	Matériaux et méthodes	70
III.2.1.	Description de méthode hybride SVM-DA	71
III.3.	Résultats et discussion	71
III.3.1.	Température critique	71
III.3.2.	Pression critique	74
III.3.3.	Volume critique	76
III.3.4.	Facteur acentrique	77
III.3.5.	Domaine d'applicabilité des modèles développés	79
III.3.6.	Évaluation statistique et validation de modèle élaboré	81
III.3.7.	Analyse de sensibilité globale	81
III.4.	Conclusion	84

Chapitre IV.
Contribution à la modélisation de la solubilité des solutés solides dans le CO_2-SC par les modèles semi empiriques

IV.1.	Introduction	86
IV.2.	Matériels et Méthodes	86
IV.2.1.	Élaboration de la base de données	86
IV.2.2.	Structure de la base de données	90
IV.2.3.	Proposition d'un nouveau modèle	90
IV.3.	Résultats et discussions	91
IV.3.1.	Ajustement des paramètres du modèle proposé	91
IV.3.2.	Comparaison globale entre les performances de différents modèles	93
IV.3.3.	Diagramme de dispersion et Vecteurs de validation (α, β, R):	95
IV.3.4.	Discussion des résultats	95
IV.4.	Conclusions	96

Chapitre V.
Contribution à la modélisation de la solubilité des solutés solides dans le CO_2-SC par les techniques d'intelligence computationnelle

V.1.	Introduction	97
V.2.	Méthodologie	98
V.2.1.	Préparation des données	99
V.2.2.	Détermination des descripteurs	103
V.2.3.	Critères d'évaluation de la performance	104
V.3.	Résultats et discussion	105
V.3.1.	Résultats du model RNA-QSPR	105
V.3.2.	Modélisation par les Machines à Vecteurs de Support SVM	108
V.3.3.	Comparaison	109
V.3.4.	Analyse de sensibilité globale	113
V.3.5.	Capacité d'extrapolation	114
V.4.	Conclusion	118

Conclusion générale 119

Abréviations

Abréviation	Signification
AAD	**E**rreur **A**bsolue **M**oyenne (**A**verage **A**bsolute **D**eviation)
ANFIS	**S**ystème d'**I**nférence **N**euro-**F**uzzy **A**daptative
CO_2-SC	**D**ioxyde de **C**arbone **S**upercritique
EdE	**E**quations **d'E**tat
ER	**E**rreur **R**elative
ERAM	**E**rreur **R**elative **A**bsolue **M**oyenne
FIS	**S**ystème d'**I**nférence **F**lou
GA	**A**lgorithmes Génétiques
GC	**C**ontribution de **G**roupe
LSSVM	**M**achine à **V**ecteurs de **S**upport de **M**oindres **C**arrés
QSPR	**R**elation **Q**uantitative **P**ropriété de **S**tructure
RBF	**F**onction de noyau de **B**ase **R**adiale
RNA	**R**éseau de **N**eurones **A**rtificiels
SC	**S**uper**c**ritique
SCF	**F**luides **S**upercritiques
SLT	**T**héorie de l'**A**pprentissage **S**tatistique
SVM	**M**achines à **V**ecteurs de **S**upport
VW	Equation de **V**an der **W**aals

Nomenclature

Lecture	Désignation	Unité
D	Diffusivité	$m^2.s^{-1}$
f	Fugacité	Pa
K	Fonction du noyau (**K**ernel)	(-)
$\mathbf{M_w}$	Masse Molaire	g/mol
NC	Nombre de composés	(-)
$\mathbf{N_i}$	Nombre des points expérimentaux	(-)
P	Pression	Pa
$\mathbf{P_c}$	Pression critique	Pa
$\mathbf{P_{tr}}$	Pression au point triple	Pa
R	Constante universelle des gaz	$J.Mole^{-1}.K^{-1}$
R	Coefficient de corrélation	(-)
$\mathbf{R^2}$	Coefficient de détermination	(-)
T	Température	K
t	Temps	s
$\mathbf{T_c}$	Température critique	K
$\mathbf{T_{tr}}$	Température au point triple	K
V	Volume molaire	m^3.kilo $Mole^{-1}$
w	Poids	(-)
$\mathbf{y_2}$	Solubilité en Fraction molaire	(-)
Z	Facteur de compressibilité	(-)

Lettres Grecques

Lecture	Désignation	Unité
μ	Viscosité	Pa.s
ρ	Densité du solvant supercritique	$kg.m^{-3}$
φ	Coefficient de fugacité	(-)

Exposants

Lecture	Désignation
cal	Calculée
exp	Expérimentale
iso	Isotrope
ref	Référence
SC	Supercritique
sub	sublimation

Liste des figures

Chapitre I.
La solubilité dans les fluides supercritiques

Figure I.1. Représentation 3D des courbes de coexistences en fonction de la température de la densité et de la pression. 5

Figure I.2. Diagramme de phase d'une substance pure. 6

Figure I.3. Cellule d'équilibre liquide-vapeur évoluant vers l'état supercritique. 6

Figure I.4. Diagramme de phase du CO_2. 9

Figure I.5. Appareil de base pour les mesures de solubilité en utilisant une méthode statique/analytique. 10

Figure I.6. La cellule de volume variable utilisée dans la méthode statique/synthétique de mesure de la solubilité. 11

Figure I.7. Cuve de pression utilisé dans la méthode statique/gravimétrique. 13

Figure I.8. Dispositif pour la détermination de la solubilité en utilisant un procédé dynamique. 14

Figure I.9. Etapes de la procédure de modélisation neuronale. 32

Figure I.10. Topologie générale du modèle LS-SVM. 34

Figure I.11. Structure d'ANFIS comprenant deux paramètres d'entrée. 36

Figure I.12. Classification des différents modèles de prédiction de la solubilité des solides dans les fluides supercritiques. 37

Chapitre II.
Techniques d'intelligence computationnelle & Chemoinformatiques

Figure II.1. Pourcentage de chaque type des descripteurs dans la base de données de logiciel DRAGON 7.0. 42

Figure II.2. Schéma d'un neurone biologique. 43

Figure II.3. Schéma d'un neurone artificiel. 44

Figure II.4. Schéma de comparaison entre un neurone biologique et un neurone artificiel. 45

Figure II.5. Réseau de neurones artificiels de type perceptron multicouches. 46

Figure II.6. Les déférentes possibilités de classification des réseaux de neurones. 49

Figure II.7. Réseau multicouches MLP. 49

Figure II.8. Illustration de l'apprentissage supervisé. 51

Figure II.9. Illustration de l'apprentissage non supervisé. 52

Figure II.10. Organigramme d'apprentissage par l'algorithme de rétro-propagation. 53

Figure II.11. La structure typique d'un système flou. 59

Figure II.12. Structure schématique du modèle flou TSK. 61

Figure II.13. L'architecture ANFIS avec deux paramètres d'entrée x et y. 62

Figure II.14. Meilleur compromis entre la complexité du modèle et l'erreur l'apprentissage minimale. 64

Figure II.15. Régression avec le tube insensible ε. 66

Figure II.16. Effet de la transformation du noyau. 67

Chapitre III.
Contribution à la modélisation des paramètres critiques par l'intelligence computationnelle

Figure III.1. Comparaison entre la température critique expérimentale et prévue de l'ensemble de la base de données. a: MLR, b: RNA, c: SVM, d: SVM-DA 76

Figure III.2. Comparaison entre la pression critique expérimentale et prévue de l'ensemble de la base de données. a: MLR, b: RNA, c: SVM, d: SVM-DA 77

Figure III.3. Comparaison entre le volume critique expérimentale et prévue de l'ensemble de la base de données. a: MLR, b: RNA, c: SVM, d: SVM-DA 79

Figure III.4. Comparaison entre le acentrique facteur expérimentale et prévue de l'ensemble de la base de données. a: MLR, b: RNA, c: SVM, d: SVM-DA 81

Figure III.5. Prédiction performance des 04 paramètres par les 04 modèles. 82

Figure III.6. Domaine d'application (William) pour chaque propriété (a) T_C, (b) P_C, (c) V_C et (d) ω. 83

Figure III.7. Erreur Relative Absolu (ERA%) des modèles proposés pour la prédiction de T_C, P_C, V_C et ω d'environ 6700 points de données, 84

Figure III.8. Importance des entrées sur chaque sortie. 85

Figure III.9. Les propriétés expérimentales de certains composés vs prédits par SVM-DA. 86

Figure III.10. Interface graphique. 87

Chapitre IV.
Contribution à la modélisation de la solubilité des solutés solides dans le CO_2-SC par les modèles semi empiriques

Figure IV.1. Organigramme de l'ajustement des paramètres 93

Figure IV.2. Nombre de systèmes les mieux corrélés par chaque modèle. 95

Figure IV.3. Comparaison entre la solubilité expérimentale et prédite de 130 solutés dans CO_2-SC à l'aide d'un modèle semi-empirique. 96

Chapitre V.
Contribution à la modélisation de la solubilité des solutés solides dans le CO_2-SC par les techniques d'intelligence computationnelle

Figure V.1. Procédure de construction et d'optimisation des modèles RNA-QSPR et SVM-QSPR 100

Figure V.2. La régression entre les valeurs expérimentales et prévues de la solubilité de 145 solutés dans CO_2-SC en utilisant RNA. 108

Figure V.3. Comparaison entre les valeurs expérimentales et prédites de la solubilité de 145 solutés dans CO_2-SC à l'aide du modèle SVM. 110

Figure V.4. Comparaison statistique entre le modèle RNA et SVM. 111

Figure V.5. Comparaison entre les modèles développés dans cette étude et ceux précédemment publiés dans la littérature en termes de (ERAM%) et nombre de composés. 112

Figure V.6. Comparaison graphique globale pour chacun des 06 modèles basés sur la densité et notre modèle SVM-QSPR 113

Figure V.7. Interface graphique pour évaluer la solubilité dans CO_2-SC à l'aide d'un modèle SVM. 114

Figure V.8. Analyse de sensibilité entre les variables d'entrée et la sortie calculée. 115

Figure V.9. a: Les valeurs de solubilité prédite par rapport aux valeurs expérimentales pour le système Carvedilol, b: Comparaison entre données expérimentales et SVM-QSPR résultats prévus pour le système Carvedilol par rapport à la densité de CO_2-SC. 115

Figure V.10. a: Les valeurs de solubilité prédite par rapport aux valeurs expérimentales pour le système Cephalexin, b: Comparaison entre données expérimentales et SVM-QSPR résultats prévus pour le système Cephalexin par rapport à la densité de CO_2-SC 116

Figure V.11. a: Les valeurs de solubilité prédite par rapport aux valeurs expérimentales pour le système Cyproheptadine, b: Comparaison entre données expérimentales et SVM-QSPR résultats prévus pour le système Cyproheptadine par rapport à la densité de CO_2-SC. 116

Figure V.12. a: Les valeurs de solubilité prédite par rapport aux valeurs expérimentales pour le système hydrochloride, b: Comparaison entre données expérimentales et SVM-QSPR résultats prévus pour le système hydrochloride par rapport à la densité de CO_2-SC 117

Figure V.13. a: Les valeurs de solubilité prédite par rapport aux valeurs expérimentales pour le système Gabapentin, b: Comparaison entre données expérimentales et SVM-QSPR résultats prévus pour le système Gabapentin par rapport à la densité de CO_2-SC. 117

Figure V.14. a: Les valeurs de solubilité prédite par rapport aux valeurs expérimentales pour le système Phenylephrine Hydrochloride, b: Comparaison entre données expérimentales et SVM-QSPR résultats prévus pour le système Phenylephrine Hydrochloride par rapport à la densité de CO_2-SC. 118

Figure V.15. a: Les valeurs de solubilité prédite par rapport aux valeurs expérimentales pour le système Piroxicam, b: Comparaison entre données expérimentales et SVM-QSPR résultats prévus pour le système Piroxicam par rapport à la densité de CO_2-SC. 118

Figure V.16. a: Les valeurs de solubilité prédite par rapport aux valeurs expérimentales pour le système Spironolactone, b: Comparaison entre données expérimentales et SVM-QSPR résultats prévus pour le système Spironolactone par rapport à la densité de CO_2-SC. 119

Liste des tableaux

Chapitre I.
La solubilité dans les fluides supercritiques

Tableau I.1. Comparaison des propriétés d'un gaz, d'un fluide supercritique et d'un liquide (ordres de grandeur). 7

Tableau I.2. Propriétés critiques d'un certain composés,. 8

Tableau I.3. Les différentes équations d'état cubiques,. 20

Tableau I.4. Avantages et inconvénients des EdE cubiques de type van der Waals,. 21

Chapitre II.
Techniques d'intelligence computationnelle & Chemoinformatiques

Tableau II.1. Correspondance entre les modèles neuronaux biologiques et artificiels. 46

Tableau II.2. Fonctions de transfert usuelles, a=f(n) 48

Tableau II.3. Caractéristiques des algorithmes d'apprentissage implantés dans MATLAB®. 54

Chapitre IV.
Contribution à la modélisation de la solubilité des solutés solides dans le CO_2-SC par les modèles semi empiriques.

Tableau IV.1. Données expérimentales utilisées pour la modélisation semi empirique de la solubilité des solutés solides dans le dioxyde de carbone supercritique. 87

Tableau IV.2. Les données expérimentales de Quinine. 90

Tableau IV.3. Modèles semi-empiriques basée sur la densité pour la corrélation de la solubilité des solutés solide dans CO_2-SC. 91

Tableau IV.4. Comparaison globale de l'erreur de calcul par chacune des 06 modèles étudiées. 94

Chapitre V.
Contribution à la modélisation de la solubilité des solutés solides dans le CO_2-SC par les techniques d'intelligence computationnelle.

Tableau V.1. Données expérimentales utilisées pour la modélisation de solubilité des solutés solides dans le dioxyde de carbone supercritique 102

Tableau V.2. Les trois descripteurs moléculaires 2D utilisés dans le modèle QSPR 106

Tableau V.3. Définitions des mesures de la performance 106

Tableau V.4. Architecture du modèle RNA optimisé 108

Tableau V.5. Résultats statistiques et de validation du modèle RNA-QSPR 109

Tableau V.6. Détails du modèle SVM proposé. 110

Tableau V.7. Analyses statistiques pour le modèle SVM proposé. 110

Tableau V.8. Comparaison entre le modèle SVM-QSPR développé et d'autres modèles intelligents rapportés dans la littérature. 112

Tableau V.9. Modèles basés sur la densité pour la corrélation des solutés solides dans CO_2-SC. 113

Tableau V.10. Comparaison globale pour chacun des 06 modèles basés sur la densité et notre modèle SVM-QSPR. 114

Introduction générale

Introduction générale

De nos jours, l'utilisation des fluides supercritiques dans l'ingénierie des processus gagne une attention croissante en raison de ses divers avantages. C'est également le dioxyde de carbone supercritique (CO_2-SC) est le fluide supercritique (SCF) le plus utilisé car il est non polluant, bénin, non toxique, non inflammable, bon marché et disponible. En tant que solvant vert, le dioxyde de carbone supercritique (CO_2-SC) trouve des applications dans de nombreuses technologies conventionnelles et nouvelles telles que les techniques d'extraction, de séparation, et de traitement par teinture sèche, qui fonctionnent sans émission d'eaux usées, et de nouveaux processus biotechnologiques écologiques respectueux de l'environnement et sans danger, etc. La connaissance précise du comportement en phase des systèmes impliqués et en particulier, de la solubilité des solutés solides dans CO_2-SC à haute pression est cruciale pour la conception, l'optimisation et le fonctionnement de tels procédés [1].

Les solutés solides d'intérêt, tels que les produits pharmaceutiques et les colorants, sont des molécules complexes de poids moléculaire généralement élevé et à faible volatilité. Cela implique que très souvent les données expérimentales, telles que la solubilité, les propriétés critiques, le facteur acentrique ne sont pas facilement disponible dans la littérature [2].

L'utilisation des modèles prédictifs précis dans les applications d'ingénierie telles que celles utilisées dans simulateurs de procédés commerciaux ont été toujours un objectif pour de nombreux travaux de recherche [3].

Trois grandes approches de modélisation ont été utilisées pour modéliser la solubilité d'un soluté solide dans un fluide supercritique pour des fins prédictives ou corrélatives : les équations d'état (EdE), les techniques d'intelligence computationnelle, et les équations semi-empiriques basées sur la densité. Il convient de mentionner que les trois approches reposent davantage sur les données expérimentales [4].

L'utilisation des EdE cubiques nécessite la connaissance de certains paramètres à savoir les propriétés critiques et le facteur acentrique pour calculer la solubilité. En l'absence des valeurs expérimentales de ces paramètres, ils peuvent être estimés en utilisant la méthode de contribution de groupe (GC) [5].

Les difficultés dans l'utilisation des approches classiques et la révolution de l'informatique et le développement de nouvelles méthodes de calcul numérique basé sur l'intelligence artificielle et les sciences cognitives ont donné naissance durant les dernières décennies à de nouvelles méthodes de calcul et de modélisation telles que les réseaux de neurones artificiels (RNA) [4], la machine à vecteurs de support [6], les algorithmes génétiques et la logique floue [7]. L'efficacité, la fiabilité et la robustesse de ces méthodes ont été démontrées par leurs nombreuses applications dans le domaine d'ingénierie.

Le principal avantage de modélisation par l'intelligence computationnelle, comparé aux modèles classiques, est leur capacité à modéliser ces procédés sans avoir une connaissance approfondie des mécanismes mis en jeu. Donc l'investigation de modélisation par l'intelligence computationnelle pour le calcul des propriétés thermochimique telles que les propriétés critiques des substances pures et la solubilité des solutés solides dans le CO_2-SC devient profitable [4].

L'objectif de ce travail, consiste à réaliser une méthodologie de calcul basée sur les méthodes d'intelligences computationnelle et la proposition de nouveaux modèles de prédictions flexibles et robustes afin d'améliorant la qualité de corrélation des propriétés critiques et de la solubilité de solutés solides dans le CO_2-SC.

Pour atteindre l'objectif assigné à cette thèse, cinq chapitres ont été développés. La présente introduction générale a permis de situer le contexte et les motivations de cette recherche.

Le **premier chapitre** est consacré à une généralité sur les fluides supercritiques, leurs propriétés ainsi qu'un aperçu sur les applications et les modèles de prédiction de la solubilité dans les fluides supercritiques.

Dans le **deuxième chapitre**, nous parlons brièvement sur les différentes approches d'intelligences computationnelles.

Le **troisième chapitre** est consacré à une contribution à la modélisation des propriétés critiques et facteur acentrique par différentes approches d'intelligences computationnelles (réseaux de neurones artificiels (RNA), les machines à vecteurs de support (SVM)) et la méthode de régression linaire multiple (MLR)

Le **chapitre quatre** est consacré à une contribution à la modélisation semi-empirique de la solubilité des solutés solides dans le dioxyde de carbone supercritique, étude

comparative de tous les résultats obtenus par cinq modèles et puis proposition d'un nouveau modèle semi-empirique basé sur la densité.

Le **dernier chapitre** traite une troisième contribution à la modélisation d'intelligences computationnelles de la solubilité des solutés solides dans le dioxyde de carbone supercritique pour une base de donnée de 145 composés présentant 3637 point expérimentaux, une proposition d'un nouveau modèle semi-empirique basé sur la densité a été effectué suivie par une étude comparative de tous les résultats obtenus.

Les **conclusions** tirées de ce travail ainsi que les études qui peuvent être envisagées dans le futur sont présentés à la fin de cette thèse.

Chapitre I.
La solubilité dans les fluides supercritiques.

Chapitre I

La solubilité dans les fluides supercritiques.

I.1.	Introduction	4
I.2.	Définition d'un fluide supercritique	4
I.3.	Techniques expérimentales pour mesurer la solubilité dans SC-CO_2	9
I.4.	Modélisation de la solubilité des solutés solides dans les fluides supercritique	14
I.5.	Conclusion	37

I.1. Introduction

Les états usuels de la matière sont : gaz, liquide et solide. Entre ces trois états, existe des comportements et des points particuliers liés aux transitions de phases : la fusion, la sublimation, le point triple où coexistent les trois états et le point critique au-delà duquel les fluides ne sont ni liquide ni gaz mais supercritique.

Pour le développement des technologies de l'industrie chimique, la sélection des solvants est d'une grande importance. Depuis longtemps, le champ des recherches s'est limité aux liquides. Plus récemment, des tentatives ont été faites pour utiliser, comme nouveaux solvants, les fluides supercritiques dont les propriétés peuvent être contrôlées aisément au moyen de la pression et de la température.

Ces fluides se sont rapidement révélés d'un grand intérêt pour l'extraction ou la séparation. Par la suite, les applications se sont étendues à de nombreux domaines, comme la réaction ou la cristallisation.

Nous rappelons dans un premier temps sur les fluides supercritiques, la description de la zone supercritique, ses propriétés physico-chimiques et le cas particulier de CO_2-SC. Nous exposons ensuite, les applications des fluides supercritiques, techniques de mesures expérimentales et les méthodes de modélisations de la solubilité des solutés solides dans les fluides supercritique.

I.2. Définition d'un fluide supercritique

Les fluides supercritiques dans ces conditions de température et de pression sont caractérisés pour la première fois en 1822, suite aux travaux menés par le physicien Cagniard de La Tour [8]. Ce dernier observe que, dans un espace clos, la séparation entre les phases liquide et gazeuse disparaît au-delà d'une certaine température et d'une certaine pression.

Il définit alors le point critique, qui marque l'apparition d'un état intermédiaire entre le liquide et le gaz, appelé alors "état supercritique".

Cette notion sera ensuite discutée, puisqu'il ne s'agit pas d'un état de la matière comme peuvent l'être le solide, le liquide ou le gaz. Il s'agit cependant d'une expression couramment utilisée dans le jargon scientifique, nous essaierons de l'utiliser avec parcimonie et nous la mettrons entre guillemets lorsque nous souhaiterons évoquer ce fluide dans ces hautes pressions et températures. A noter que contrairement aux autres transitions de phases (liquide-gaz, liquide-solide ou solide-gaz), le passage dans le domaine SC ne s'accompagne pas d'une discontinuité des propriétés physico-chimiques du fluide.

I.2.1. Description de la zone supercritique

La matière qui nous entoure peut se présenter sous trois états thermodynamiques : état solide, état liquide et état gazeux. Les changements d'états sont progressifs et localisés et fonction de la pression et de la température. La figure (I.1) permet de situer la région critique dans une représentation en trois dimensions, fonction de la densité, de la température et de la pression. Il faut se placer au-dessus de l'isotherme critique pour se trouver dans le domaine supercritique, on voit qu'il est possible de passer de densités très faibles à des densités très élevées de façon parfaitement continue. Or, les propriétés de solvatation des fluides sont dépendantes de la densité et varient énormément le long des isothermes supercritique, c'est cette propriété qui est utilisée dans de nombreuses applications.

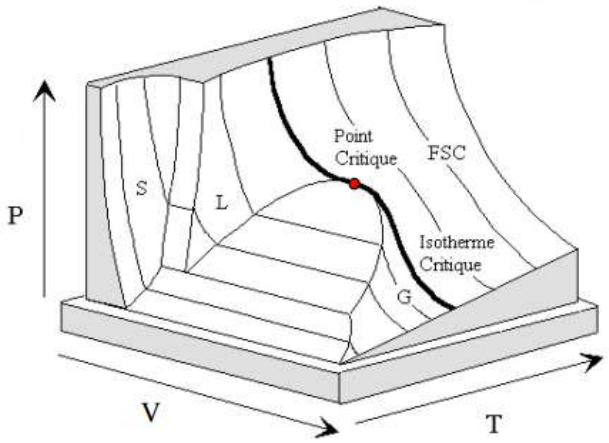


Figure I.1. Représentation 3D des courbes de coexistences en fonction de la température de la densité et de la pression, [9].

Nous exposons aussi le diagramme de phase (Figure I.2) qui montre que tout corps pur possède des courbes d'équilibre qui décrivent ces changements d'état. Le point triple correspond à un couple de température et de pression donné (T_{tp} ; P_{tp}) pour un corps pur, et

c'est dans ces conditions uniques que les trois états (liquide, solide et gaz) coexistent. La figure (I.3) montre une cellule d'équilibre liquide-vapeur à haute pression et son évolution à des conditions supercritiques, Avec l'effet de la température et de la pression, on observe un phénomène d'expansion de la phase liquide, ce qui entraine une disparition progressive de la ligne séparant la phase liquide et la phase gazeuse. Au point critique, une seule phase homogène est obtenue.

Pour ce même corps pur, il existe un couple de température et de pression (T_c ; P_c) où la phase liquide et gazeuse ont la même densité, définissant ainsi un point particulier du diagramme, appelé le point critique. Au-delà de ce point, se situe le domaine SC de ce composé, où seule une phase existe. Il n'existe pas de changement d'état entre les phases gazeuse ou liquide et SC, la transition se fait de manière continue et homogène (une phase n'apparait pas localement dans une autre donc sans apparition de surface et donc de tension de surface), c'est pourquoi ces transitions sont matérialisées en pointillés dans le diagramme de phase.

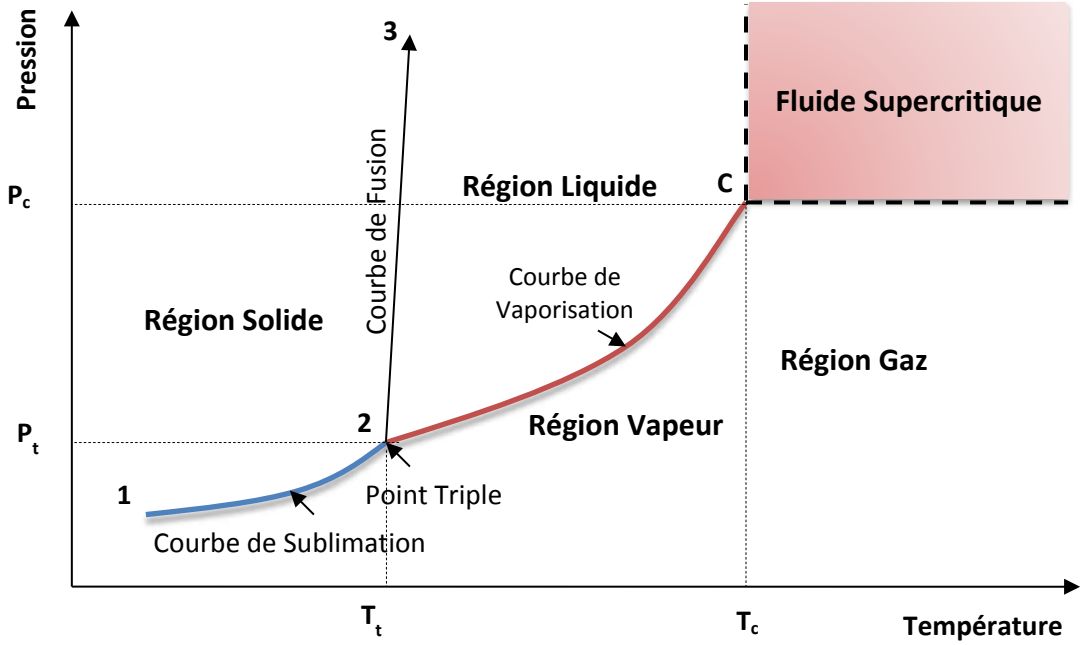


Figure I.2. Diagramme de phase d'une substance pure, [10].

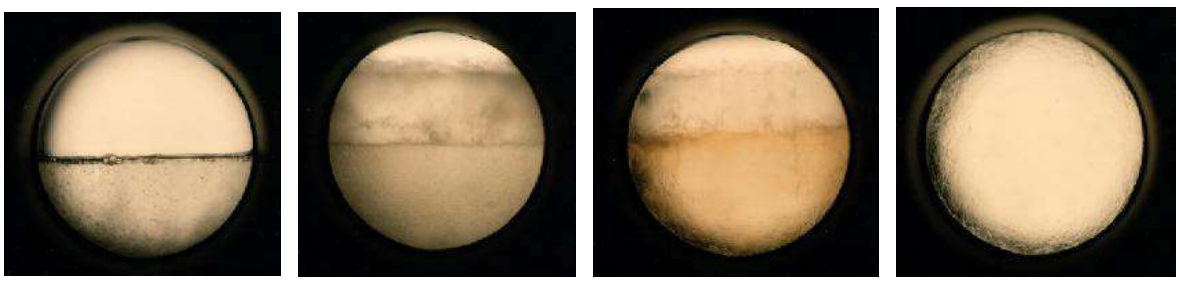


Figure I.3. Cellule d'équilibre liquide-vapeur évoluant vers l'état supercritique, [11].

I.2.2. Propriétés physico-chimiques des fluides supercritiques

Dans les conditions supercritiques, le fluide a des propriétés intermédiaires à celles d'un gaz et d'un liquide (Tableau I.1). Les fluides supercritiques possèdent une viscosité et une diffusivité proches des gaz, mais leur masse volumique les rapproche du liquide. Cette masse volumique élevée confère aux fluides supercritiques la propriété d'être de bons solvants. Leur viscosité dynamique, proche de celle des gaz, favorise les transferts de matière. De plus, grâce à leur diffusivité plus élevée que celle des liquides ainsi que grâce à l'absence de tension de surface, ils ont une meilleure pénétration dans les milieux poreux.

Les fluides supercritiques offrent la particularité d'avoir des propriétés ajustables non seulement par la température, mais aussi par la pression ; ceci leur a valu d'être qualifiés de solvants a "géométrie variable". A température constante, une augmentation de la pression entraîne une augmentation de la masse volumique et donc une augmentation de leur pouvoir solvant.

L'influence de la température sur la solubilité est plus complexe que celle de la pression. Aux pressions élevées, quand le fluide est relativement incompressible, la solubilité augmente avec la température suivant ainsi l'augmentation de la tension de vapeur du soluté. Par contre, à des pressions intermédiaires proches de la P_c, le fluide est hautement compressible et une faible augmentation de la température provoque une forte diminution de la masse volumique et donc une diminution de la solubilité du composé. Il peut être ainsi utilisé comme solvant à un moment donné à l'instant suivant, simplement en diminuant sa pression par exemple : l'étape de séparation solvant/produit est réduite à une détente, ce qui évite des purifications par distillations, extractions liquide-liquide ou autres procédés.

Tableau I.1. Comparaison des propriétés d'un gaz, d'un fluide supercritique et d'un liquide (ordres de grandeur) [10].

	Masse Volumique ρ ($kg.m^{-3}$)	**Viscosité μ (Pa.s)**	**Diffusivité D ($m^2.s^{-1}$)**
Gaz 20°C ; 1 bar	$0.1 < \rho < 2$	$10^{-5} < \mu < 3.10^{-5}$	$10^{-5} < D < 4.10^{-5}$
Fluide Supercritique $T_c ; P_c$	$200 < \rho < 500$	$10^{-5} < \mu < 3.10^{-5}$	$10^{-8} < D < 10.10^{-8}$
$T_c ; 4P_c$	$400 < \rho < 900$	$3.10^{-5} < \mu < 9.10^{-5}$	2.10^{-8}
Liquide 20°C ; 1 bar	$600 < \rho < 1600$	$2.10^{-4} < \mu < 3.10^{-3}$	$2.10^{-10} < D < 2.10^{-9}$

En résumé, le fluide supercritique possède plusieurs avantages par rapport au liquide:

- Grand coefficient de diffusion et l'absence de tension de surface.
- Faible coefficient de viscosité.
- Bonne capacité volumique.
- Bonne vitesse de dissolution.
- Solubilité, variable en modifiant la température ou la pression, ce qui facilite la récupération de la substance dissoute.
- Propriétés physiques adaptables par « simple » changement de pression et/ou de température.

Il existe un certain nombre de composés présentant des conditions supercritiques plus ou moins accessibles (tableau I.2). Le CO_2 et l'eau sont les deux fluides les plus utilisés. Néanmoins, l'eau présente des inconvénients comme des coordonnées supercritiques élevées et un effet oxydant, ce qui permet au CO_2 d'être le fluide supercritique de choix pour un grand nombre de domaines d'applications, [12].

Tableau I.2. Propriétés critiques d'un certain composés, [13].

Solvent	T_c (°C)	P_c (MPa)	Solvent	T_c (°C)	P_c (MPa)
Acetone	235.1	4.70	Hexafluroethane	20	3.06
Ammonia	132.6	11.3	Methane	-82.6	4.60
Carbone dioxide	31.1	7.38	Methanol	239.6	8.09
Cyclohexane	280.5	4.07	n-hexane	234.5	3.01
Diethyl ether	193.7	3.64	Propane	96.8	4.25
Diflioromethane	78.6	5.83	Propylene	91.9	4.60
Diflioroethane	113.7	4.50	Sulfur hexafluoride	45.7	3.76
Dimethyl ether	127	5.24	Tetrafluoromethane	-45.4	3.74
Ethane	32.3	4.87	Toluene	318.8	41.1
Ethylene	9.4	5.04	Trifluoromethane	26.3	4.86
Ethyne	35.3	6.14	water	374.3	22.1

I.2.3. Cas particulier du CO_2-SC

Le CO_2 supercritique connait une utilisation importante au niveau industriel dans de nombreux domaines d'applications, notamment pour : les applications pharmaceutiques, le traitement des textiles, la polymérisation, l'extraction de produits naturels, les applications alimentaires, le nettoyage, les applications de chromatographie par fluide supercritique et la mise en œuvre des polymères.

Le CO_2 est utilisé couramment pour plusieurs raisons :

- Il est facile à manipuler dans les conditions critiques puisque la T_c et la P_c sont relativement basses : P_c = 73,8 bar et T_c = 31,1°C (Figure I.4).
- Non inflammable, inodore, le plus souvent chimiquement inerte, non polluant et non toxique.
- Il est peu couteux (peu cher à la fabrication et sans cout d'élimination comparativement aux solvants organiques).

- Il est disponible à une grande pureté.
- Il solubilise un grand nombre de solutés non polaires ou faiblement polaires dans le domaine SC, la précipitation de ce soluté se fait ensuite par détente (retour du CO_2 a l'état gazeux) d'où l'absence de solvant résiduel à la fin du traitement et la récupération facile du CO_2.
- Il présente un grand pouvoir de diffusion ainsi qu'une densité proche du liquide dans le domaine SC.
- Le pouvoir solvant est important vis-à-vis des composés qui présentent une faible masse molaire et diminue avec l'augmentation de la masse molaire du composé.

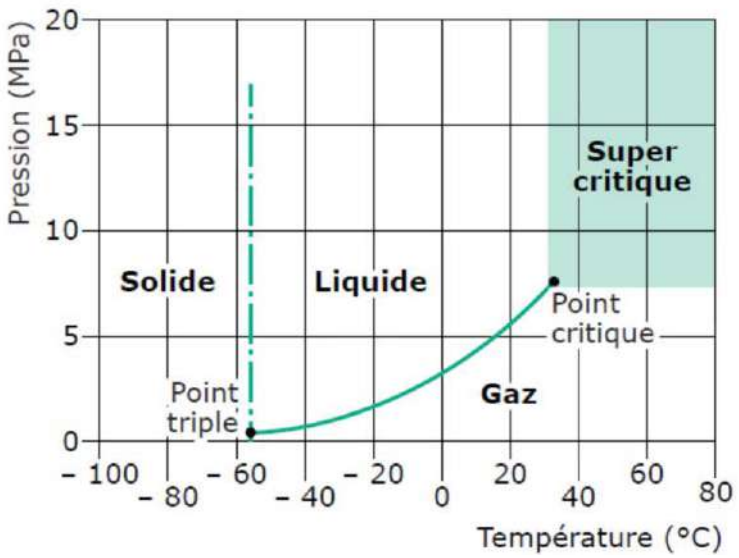


Figure I.4. Diagramme de phase du CO_2, [10].

I.3. Techniques expérimentales pour mesurer la solubilité dans CO_2-SC

Typiquement la solubilité est définie comme la fraction molaire ou la fraction massique de soluté dans le fluide supercritique qui est en équilibre avec la masse de soluté. La mesure de la solubilité au début a été effectuée par Francis (1954) [14] et, depuis 1989, diverses méthodes pour mesurer la solubilité dans les fluides supercritiques ont été examinées dans la littérature. Ces méthodes peuvent être divisées en deux grandes catégories : statique et dynamique.

I.3.1. Les méthodes statiques

Dans ces méthodes, le soluté est autorisé à être en contact statique pour longtemps avec le fluide supercritique afin d'atteindre l'équilibre. Selon l'échantillonnage et du type de récipient haute pression utilisé, il y a trois méthodes : analytique, synthétique et gravimétrique.

I.3.1.1. Les méthodes analytiques

Dans cette méthode, un récipient d'équilibre à volume constantes utilisée dans laquelle un composé est équilibré avec une quantité connue de fluide supercritique. Un petit échantillon de la phase fluide est prélevé et analysé pour la concentration du soluté. Le récipient de l'échantillon est rempli de soluté, et puis le vide est appliqué pour éliminer l'air [15].

Des quantités connues de solvant (si nécessaire) et de dioxyde de carbone sont ensuite pompés dans le système. Le fluide est autorisé à contacter le soluté par une pompe de recirculation pour réduire le temps d'équilibrage. Une pompe magnétique est utilisée, car il a un petit volume intérieur. Une fois l'équilibre est atteint, un petit volume d'échantillon est étiré à l'aide d'une boucle d'échantillon fixé à la soupape de commutation. L'échantillon peut être envoyé directement à une Chromatographie liquide à haute pression pour l'analyse en ligne ou collectées pour l'analyse hors ligne en utilisant d'autres techniques. Solubilité du soluté, y_2, en fractions molaires est calculée en tant que le vide est appliqué pour éliminer l'air [15].

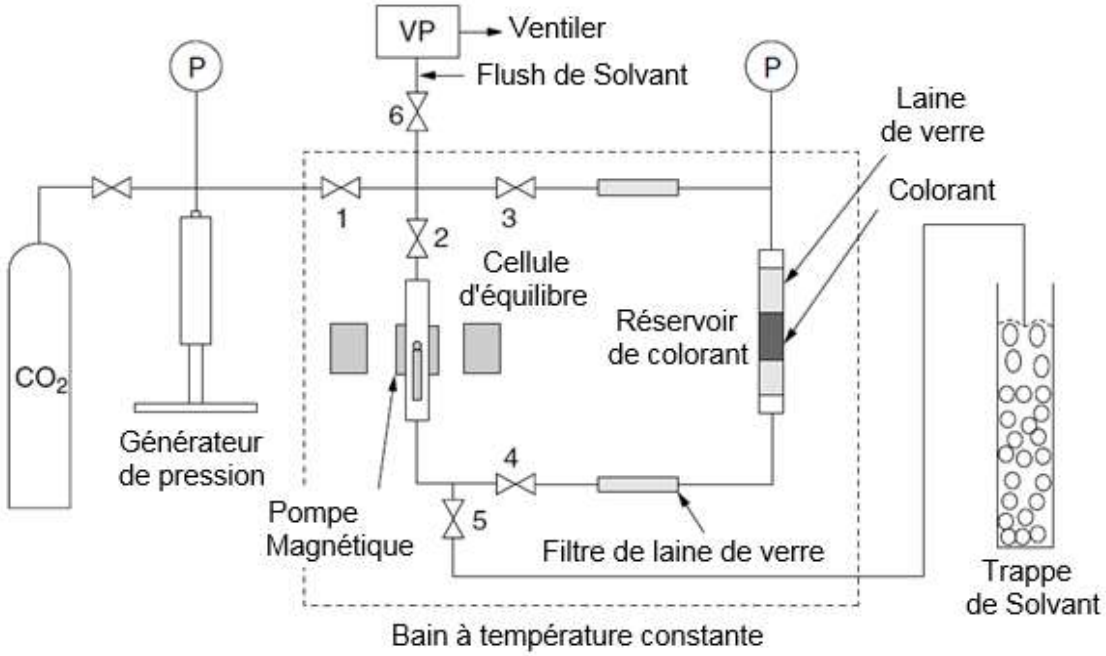


Figure I.5. Appareil de base pour les mesures de solubilité en utilisant une méthode statique / analytique, [15].

VP est une pompe à vide et une connexion pour le solvant de rinçage ;
1-6 sont des vannes ;
P comprend des jauges de pression.

Étapes de l'opération :

a) Evacuer l'ensemble du dispositif avec toutes les soupapes ouvertes à l'exception des vannes 1 et 5;
b) Isoler l'appareil en fermant la vanne 6 ;

c) Fournir CO_2 à partir du cylindre de l'appareil à une pression souhaitée par pression de fonctionnement générateur et une vanne d'ouverture ;
d) Fermer la vanne 1 et faire circuler CO_2 en utilisant une pompe magnétique pour environ 2 heures pour atteindre l'équilibre ;
e) Fermer les vannes 3 et 4 pour isoler le CO_2 + le mélange de (soluté, fluide) ;
f) Ouvrir la vanne 5 pour libérer le mélange de (solvant, fluide) dans le piège ;
g) Circuler le solvant pour rincer le soluté déposé sur les murs ;
h) Analyser la solution dans le piège pour la concentration de soluté.

I.3.1.2. Les méthodes synthétiques

Une cellule de volume variable est utilisée pour ajuster les conditions de fonctionnement. Typiquement, le récipient est équipé d'une fenêtre en saphir pour des observations visuelles. Une quantité fixe de soluté est dissoute en un montant connu de fluide supercritique. Les conditions sont alors ajustées pour réduire progressivement la solubilité du mélange et provoquer la précipitation. Le début de la précipitation, qui est, du point de trouble, est enregistré comme une mesure de l'état de solubilité. Pour mesurer le point de trouble, une propriété physique peut être utilisé, dont d'abord dérivée par rapport aux variables modulés subit un changement net lors de la transition de phase se produit. Pour un soluté/fluide chargement donné, nombreux points de solubilité peut être mesurée par la variation de la température et de la pression de façon indépendante. Cette méthode ne nécessite pas de prélèvement de l'échantillon ou de son analyse, [16].

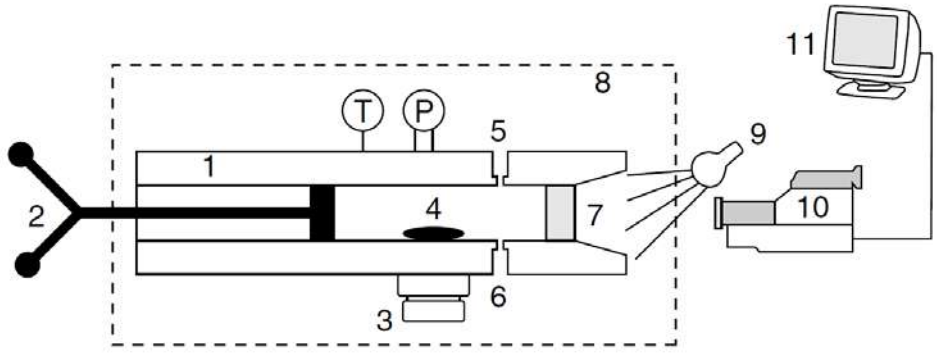


Figure I.6. La cellule de volume variable utilisée dans la méthode statique/synthétique de mesure de la solubilité, [16].

Les composants de la cellule de volume variable sont les suivants :

(1) Cellule de volume variable. (2) Pompe à vis. (3) Agitateur magnétique.
(4) Bar magnétique. (5) Ligne d'échantillonnage supérieure.
(6) Ligne d'échantillonnage inférieure. (7) la fenêtre de saphir. (8) Bain thermostat.
(9) Source de lumière. (10) Caméra du vidéo. (11) Un moniteur.

Le procédé de synthèse est très commode pour la détermination des équilibres de phase binaire et des limites de phases dans des mélanges à plusieurs composants. Les mesures peuvent être effectuées en un temps relativement court. Une autre caractéristique intéressante de cette méthode est la faible masse du soluté nécessaire pour enquêter sur un large éventail de composition. Des méthodes de synthèse ne sont pas utilisables pour des solutés qui ont le même indice de réfraction que le CO_2. La détermination du point de trouble augmente avec la différence entre les deux indices de réfraction. Pour les systèmes opaques ou de solutés avec près-CO_2 indices de réfraction, des techniques alternatives non visuels ont été proposées telles que l'utilisation de rayons X transmis pour déterminer une transition de phase.

I.3.1.3. Les méthodes gravimétriques

Dans ce procédé, le soluté est maintenue dans un petit flacon à l'intérieur de la cuve sous pression contenant un fluide supercritique [17]. Le couvercle de la fiole est tel que le soluté dissous peut facilement pénétrer à travers elle, mais les particules de soluté ne peuvent pas. Dans une expérience typique, un soluté de quantité connue est placé dans un flacon et le récipient est rempli d'un fluide.

Ensuite, le fluide atteint le flacon, le soluté se dissout, et l'amène à. Au fil du temps, la majeure partie du fluide atteint la même concentration d'équilibre en tant que le fluide à l'intérieur du flacon. Le récipient est ensuite uniquement dépressurisé, et la masse de soluté restant dans la fiole est mesurée par gravimétrie.

Bien que le procédé soit relativement simple, il peut être efficacement utilisé que pour les matières solides qui présentent une solubilité supérieure à 10^{-3} en fraction molaire, qui ne fondent pas dans les conditions expérimentales, et qui absorbent une quantité négligeable de dioxyde de carbone. La modification par Isse et al. [18] étend cette méthode pour les matières solides de solubilité d'ordre 10^{-5} en fraction molaire. Cette méthode n'a pas été largement adoptée en raison de la faible précision.

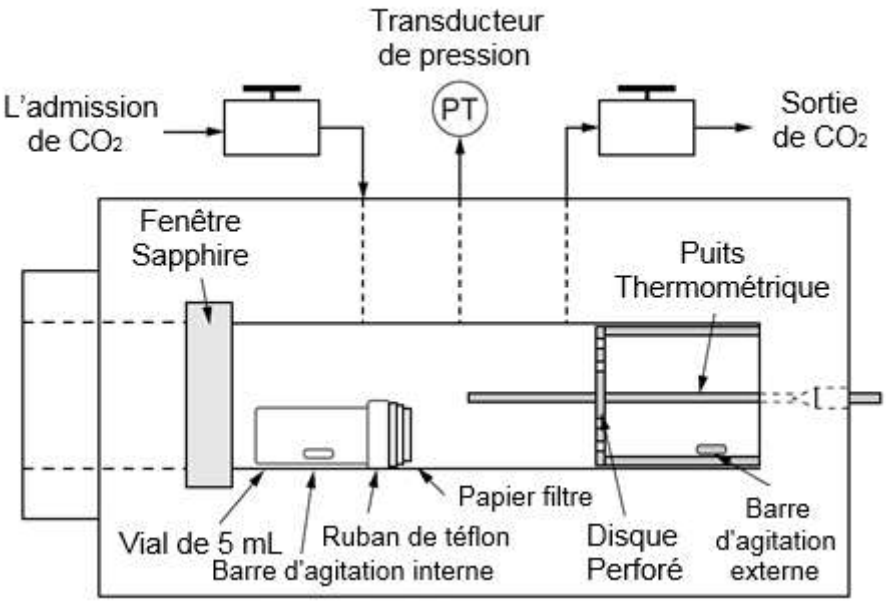


Figure I.7. Cuve de pression utilisé dans la méthode statique/gravimétrique [17].

I.3.1.4. Les méthodes dynamiques

Dans cette méthode, le récipient est rempli d'un soluté, et un flux continu supercritique est circulé à travers le récipient. Le taux du courant doux est utilisé de telle sorte que le courant de sortie est supposé atteindre l'équilibre, qui est ensuite analysé pour la concentration de soluté par chromatographique, spectroscopique, gravimétrique, et d'autres techniques. Ici, on remplit soluté dans le récipient d'échantillon. Le flux continu de fluide supercritique ou le mélange de fluide avec co-solvant sont maintenus.

Pour l'analyse en ligne, par exemple en utilisant un détecteur UV, le mélange de fluide est transmis directement par l'analyseur pour la correction la ligne de base, puis le fluide est autorisé à passer à travers le récipient d'échantillon. Pour l'analyse hors ligne, l'échantillon est collecté dans le solvant ou dans un piège froid pendant une période de temps donnée, puis analysés pour quantité de soluté.

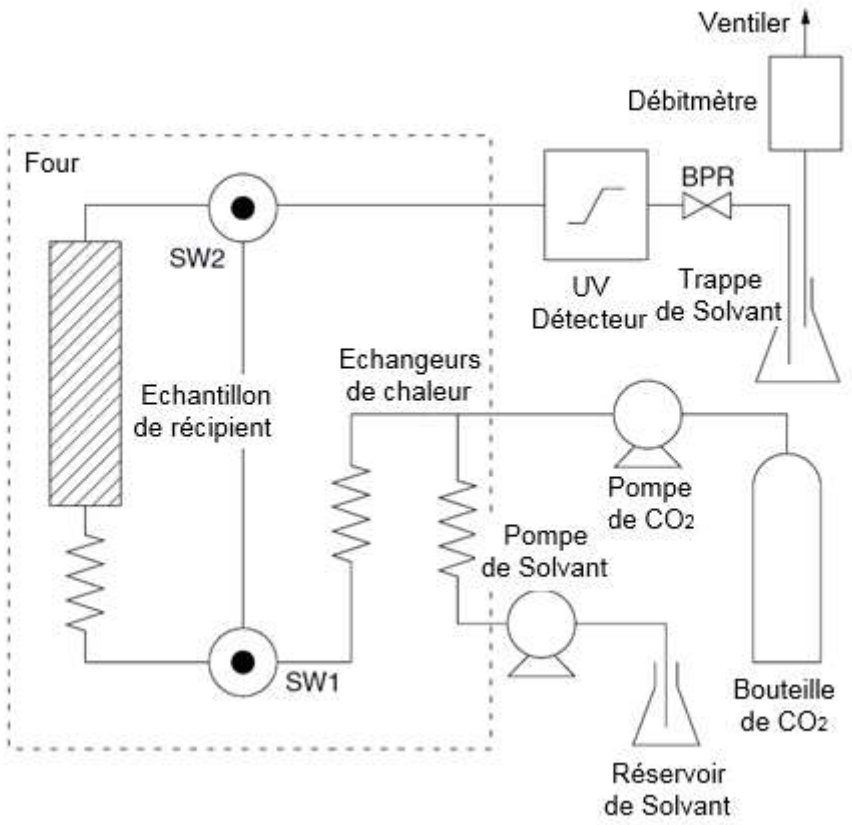


Figure I.8. Dispositif pour la détermination de la solubilité en utilisant un procédé dynamique, [19].

BPR : régulateur de contre-pression pour maintenir une pression constante dans le système.
SW : des vannes de commutation pour détourner les flux.

I.4. Méthodes de Modélisation de la solubilité des solutés solides dans les fluides supercritiques

Trois grandes approches de modélisation ont été utilisées pour modéliser la solubilité d'un soluté solide dans les fluides supercritiques : équations d'état (EdE), techniques d'intelligences computationnelles et les équations semi-empiriques. Il convient de mentionner que ces trois approches reposent davantage sur les données expérimentales et par conséquent, leur utilisation en dehors de la gamme de données expérimentales à partir de laquelle les modèles ont été obtenus n'est pas sûre.

I.4.1. Modélisation par les équations d'état

Une équation d'état exprime mathématiquement la relation entre la pression, la température et le volume molaire d'un système donné, afin de décrire les propriétés thermodynamiques des fluides ainsi que la prédiction des équilibres de phases. Il existe une très grande variété d'équations d'état dont la plupart sont applicables dans des conditions

précises et pour des composés particuliers, mais principalement elles peuvent être classées selon les trois catégories distinctes suivantes, [20] :

- Equations d'état du type du Viriel qui sont les plus simples et qui représentent des déviations assez modestes par rapport à l'idéalité de la phase vapeur.
- Les équations d'état cubiques qui représentent raisonnablement bien le comportement des phases liquide et vapeur pour les systèmes faisant intervenir des molécules non polaires, pour des intervalles de température et de pression assez limités. Cependant elles sont assez simples à manipuler mathématiquement.
- Les équations d'état applicables pour de larges intervalles de température et de pression : elles sont du type de Lee-Kesler qui est une version assez générale de celles du type de Benedict-Webb-Rubbin (BWR). Elles sont assez complexes à manipuler numériquement.

Aussi selon la nature des équations les classes suivantes peuvent être distinguées :

- Les équations d'état de nature empirique, qui contiennent un grand nombre de paramètres spécifiques aux constituants et aux conditions de pression et de température dans lesquelles ils sont étudiés engendrant des domaines où elles se révèlent assez précises.
- Les équations d'état théoriques, qui sont basées sur la thermodynamique statistique. Le nombre de paramètres de ces équations est assez faible et présentent un sens physique réel. Elles peuvent être utilisées dans un large intervalle de conditions mais sont moins précises que les précédentes et sont assez complexes à appliquer.
- Les équations semi-empiriques, qui combinent les caractéristiques des deux types précédents. Les plus connues sont les équations d'état cubiques. Elles s'appliquent avec précision dans le domaine des hautes pressions et températures. Cependant les équations les plus connues et les plus souvent utilisées dérivent directement de l'équation de Van der Waals (VW) (1873).

L'application des équations d'état a été étendue aux cas des mélanges de composés d'où l'introduction de règles de mélanges destinées à prendre en charge la présence simultanée de composés différents dans le système. Ceci, comme mentionné, auparavant soulève le problème du choix de la combinaison équation d'état-règle de mélanges la plus adéquate. Une multitude de règles de mélange sont également proposées dans la littérature.

I.4.1.1. Forme générale

Pour un corps pur, Van der Waals proposa l'équation suivante, [21] :

$$P = \frac{RT}{V-b_i} - \frac{a_i}{V^2} \qquad \text{(I.1)}$$

Les deux paramètres a_i et b_i, sont caractéristiques du corps pur considéré, sont constants et sont exprimés comme suit :

$$a_i = \Omega_a \frac{R^2 T_{c,i}^2}{P_{c,i}} \quad \text{(I.2a)}$$

$$b_i = \Omega_b \frac{RT_{c,i}}{P_{c,i}} \quad \text{(I.2b)}$$

Avec :

$$\Omega_a = \frac{27}{64} \quad et \quad \Omega_b = \frac{1}{8}$$

Cette équation nécessite la connaissance des deux coordonnées critiques du corps pur *i*, la température critique ($T_{c,i}$) et la pression critique ($P_{c,i}$). Elle comprend un terme de répulsion qui prend en compte le volume des molécules par l'intermédiaire du paramètre *b* (covolume) et un terme d'attraction qui dépend du paramètre *a*.

Le terme équation d'état cubique vient du fait que le volume V est élevé aux puissances de 1, 2 et 3 dans cette équation. La majorité des équations à deux paramètres peuvent être exprimées comme suit :

$$P = \frac{RT}{V-b} + \frac{a}{v^2+ubv+wb^2} \quad \text{(I.3)}$$

Avec *u, v* et *w* pouvant prendre des valeurs entières dans l'intervalle [-2, 0].

Les paramètres *a* et *b* dépendent des forces d'attraction et de répulsion respectivement, qui existent entre les molécules. Le paramètre *b* est connu comme le covolume, et les deux paramètres sont calculés à partir des propriétés critiques des corps purs, comme montré dans la section suivante.

Une expression équivalente à l'équation (I.3) peut être exprimée comme suit :

$$Z^3 - (1 + b' - Ub')Z^2 + \left(a' + wb'^2 - Ub' - Ub'^2\right)Z - a'b' - wb'^2 - wb'^3 = 0 \quad \text{(I.4)}$$

Avec Z étant le facteur de compressibilité et $a' = \frac{aP}{(RT)^2}$; $b' = \frac{bP}{(RT)}$

I.4.1.2. L'équation d'état de Viriel

L'équation d'état de Viriel est un développement polynomial infini en pression ou en densité (inverse de volume), [22].

Cette équation a été proposée en premier par Thiesen en 1885 sur des bases empiriques, et représente le comportement volumétrique d'un fluide réel peu dense, à partir de la référence gaz parfait, [23].

La forme de cette équation, tronquée à l'ordre de 2 est :

$$Z = 1 + \frac{B}{v} + \frac{C}{v^2} + \cdots \quad \text{(I.5)}$$

Où les coefficients B, C, … sont appelés respectivement second, troisième, … coefficient du Viriel et sont associés respectivement aux interactions entre particules élémentaires deux à deux, trois à trois, …. Pour un fluide pur, les coefficients du Viriel ne dépendent que de la température. On peut aussi exprimer le développement en facteur de compressibilité en fonction de la pression sous la forme :

$$Z = 1 + B'P + C'P^2 + D'P^3 + \cdots \quad \text{(I.6)}$$

Dont les paramètres sont liés aux coefficients du Viriel par des relations dont nous ne citons que les deux premières :

$$B' = \frac{B}{RT} \qquad et \qquad C' = \frac{C - B^2}{R^2T^2}$$

Ces développements en série ne sauraient être considérés comme une équation d'état, puisqu'ils comportent un nombre infini de termes dont la valeur est inconnue et la sommation impossible. Par contre, leur troncature a donné naissance aux « équations d'état du Viriel » que l'on caractérise par l'ordre de la troncature et par la nature du développement dont elles sont dérivées, selon qu'il soit en volume ou en pression. Ainsi l'équation de van der Waals détaillée dans le prochain paragraphe s'écrit :

$$Z = 1 + \frac{\left(b - \frac{a}{RT}\right)}{v} + \frac{b}{v^2} + \frac{b}{v^3} + \cdots \quad \text{(I.7)}$$

Soit, la contribution du gaz parfait (le 1) complétée par des termes proportionnels à *b*, la correction du volume liée à la taille finie des molécules.

Il existe différentes méthodes pour déterminer les coefficients des équations d'état du Viriel, notamment les coefficients B et B', soit à partir de données expérimentales, soit à partir de corrélations, [21]. L'équation du Viriel est souvent issue du développement du Viriel tronqué au second ou au troisième terme. Elle est appliquée uniquement aux systèmes monophasiques gazeux, puisqu'elle est limitée théoriquement et en pratique aux basses pressions.

I.4.1.3. Les équations d'état cubiques

a. L'équation d'état de Van der Waals

Historiquement le développement des EdE a été largement empirique. Il y a plus d'un siècle, Van der Waals proposa sa première version de son EdE. Depuis, plusieurs modifications ont été proposées pour améliorer la corrélation et/ou la prédiction des

propriétés volumétriques, thermodynamiques et les équilibres de phases. Bien que l'équation de Van der Waals ne soit pas précise pour la plupart des applications, elle est considérée comme la contribution majeure dans le domaine depuis la première tentative de représentation du comportement PVT par Boyle au 17ème siècle. En effet l'équation de Van der Waals n'est qu'un cas particulier de la forme générale des EdE cubiques (δ=0, ε=0, η=b, Θ=a), [24].

$$Z = \frac{V}{V-b} - \frac{\left(\frac{\Theta}{RT}\right)V(V-\eta)}{(V-b)(V^2+\delta V+\varepsilon)} = Z^{rep} + Z^{att} \qquad \text{(I.8)}$$

Qui donne :

$$Z = \frac{V}{V-b} - \frac{a}{V} \qquad \text{(I.9)}$$

Le premier terme est le terme de répulsion, 'sphère dure', contenant le paramètre de covolume b, qui prend en compte le volume propre des molécules et qui doit être inférieur à V tandis que le deuxième est le terme d'attraction contenant le paramètre a , qui représente une mesure des forces d'attraction entre les molécules. L'EdE de van der Waals était la première à prédire la coexistence de phases liquides et vapeur. Les concepts développés par van der Waals sur la distinction entre les forces de répulsion provoquées par taille moléculaire et les forces de cohésion provoquées par l'attraction moléculaire demeurent toujours en tant que base de plusieurs théories au sujet de la prédiction des propriétés des liquides et des simulations sur ordinateur basées sur la mécanique statistique. En outre, ces idées sur l'existence d'une équation continue reliant les états liquide et vapeur, selon le principe des états correspondants, et sur l'élargissement de l'EdE des corps purs aux mélanges par l'utilisation des règles de mélange ont contribué considérablement aux développements actuels dans ce domaine.

Toutes les équations cubiques sont des améliorations de celle de Van der Waals. Dans le but de mieux représenter les propriétés volumétriques des fluides.

b. L'équation d'état de Redlich et Kwong

Redlich et Kwong (1949) ont proposé une modification empirique du terme d'attraction de Van der Waals en introduisant une variation du paramètre a avec la température (tableau I.2). Ceci permet d'améliorer le calcul des masses volumiques, en particulier en phase vapeur. Par contre, les prédictions de la masse volumique liquide restent

médiocres. Cependant, il a été remarqué que les équations d'état cubiques 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. L'équation d'état de Soave

Soave (1972) a appliqué l'équation d'état de Redlich-Kwong dans calcul des tensions de vapeur des hydrocarbures en modifiant l'expression du paramètre a devenu fonction de la température et du facteur acentrique de Pitzer.

d. L'équation d'état de Peng et Robinson

En 1976, Peng et Robinson ont apporté en plus de la modification du terme d'attraction **a**, une nouvelle expression pour le terme de répulsion **b**. Le progrès est sensible pour le calcul des masses volumiques en phase liquide. En outre, le facteur de compressibilité critique est mieux estimé.

Les équations de Peng-Robinson (PR) et de Soave-Redlich-Kwong (SRK) sont largement utilisées dans l'industrie, particulièrement dans le domaine de raffinage. Leurs avantages résident dans le fait qu'elles nécessitent peu de données expérimentales (les coordonnées critiques et le facteur acentrique), un temps de simulation relativement court et surtout qu'elles conduisent à une bonne estimation des équilibres liquide-vapeur. Par ailleurs, ces équations présentent d'importantes limitations ; par exemple, les densités des liquides ne sont pas correctement évaluées.

L'application des équations cubiques est limitée aux zones relativement diluées. Dans cette zone, les résultats obtenus avec l'équation de Peng-Robinson et, dans une moindre mesure, avec celle de Redlich-Kwong (RK) se superposent aux données extraites de NIST. Il en est de même pour l'équation de Redlich-Kwong-Soave (SRK). Dans la phase dense, les résultats sont plus contrastés ;

Peng-Robinson (1976) ont surévalué légèrement la masse volumique contrairement à Redlich-Kwong. Les équations de Redlich-Kwong-Soave et Peng-Robinson sont les plus utilisées et sont les mieux adaptées pour calculer la densité en phase dense comme en phase légère.

Tableau I.3. Les différentes équations d'état cubiques, [25].

Equation	u	w	a	b
Van des Waals	0	0	$\frac{27\,R^2T_c^2}{64\,P_c}$	$\frac{RT_c}{8\,P_c}$
Redlich-Kwong	1	0	$\frac{0.42748\,R^2T_c^{2.5}}{P_cT^{0.5}}$	$\frac{0.08664\,RT_c}{P_c}$
Redlich-Kwong-Soave	1	0	$\frac{0.42748\,R^2T_c^2}{P_c}\,[1+f(w)\,(1-T_r^{0.5})]^2$ $\boldsymbol{avec}\ f(w)=0.48+1.574w-0.176w^2$	$\frac{0.08664\,RT_c}{P_c}$
Peng-Robinson	2	-1	$\frac{0.45724\,R^2T_c^2}{P_c}\,[1+f(w)\,(1-T_r^{0.5})]^2$ $\boldsymbol{avec}\ f(w)=0.37464+1.54226w-0.26992w^2$	$\frac{0.07780\,RT_c}{P_c}$

I.4.1.4. Les équations d'état complexes

Les équations d'état complexes se basent sur la théorie moléculaire qui se divise en deux catégories : les EdE pour les chaînes moléculaires et les EdE pour les fluides associatifs.

a) Les EdE pour les chaînes moléculaires

Les EdE qui entrent dans cette catégorie sont par exemple l'EdE de la théorie de la chaîne dure perturbée (PHCT) de Beret et Prausnitz (1975) qui traite les molécules comme chaînes de segments. Pour réduire la complexité mathématique de cette EdE Kim et al. (1986) ont proposé une version simplifiée de la théorie de la chaîne dure perturbée (SPHCT). L'EdE basée sur la théorie de la chaîne dure perturbée (PHCT) s'écrit selon la forme générale, [26]:

$$Z = Z^{hc} + \frac{a}{RTV} \qquad \text{(I.10)}$$

Le terme Z^{hc} représente le coefficient de compressibilité de la chaîne dure (hard chain).

L'EdE de la théorie de la chaîne anisotrope perturbée (PACT) prend en considération les effets des différences de la taille et la forme moléculaire, et les forces intermoléculaires y compris les forces anisotropes dipolaires et quadripolaires. En termes de coefficient de compressibilité, l'équation d'état de PACT peut être écrite sous la forme :

$$Z = 1 + Z^{rep} + Z^{iso} + Z^{ani} \qquad \text{(I.11)}$$

Les exposants *rep*, *iso* et *ani* désignent les contributions des forces de répulsion, les interactions isotropes et les interactions anisotropes respectivement.

b) Les EdE pour les fluides associatifs

Parmi les nombreuses EdE qui entrent dans cette catégorie on peut citer l'EdE de la théorie statistique des fluides associatifs (SAFT) proposée par Chapman et al. (1989) [27], l'EdE cubique plus association (CPA) proposée par Kontogeorgis et al. (1996) [28] qui englobe une équation d'état cubique (SRK) plus un terme à fondement théorique basé sur la théorie de perturbation thermodynamique tenant compte de l'association, l'EdE de la théorie statistique de la chaîne perturbée de fluides associatifs (PC-SAFT) proposée par Gross et Sadowski (2001) [29], et l'EdE de la théorie de la chaîne associée anisotrope perturbée (APACT). Bien que la formulation mathématique de l'EdE puisse être écrite de différentes manières, dans toutes les EdE de cette catégorie, on trouve un terme qui tient compte de l'association des molécules à travers un ou deux sites liant de la molécule.

Tableau I.4. Avantages et inconvénients des EdE cubiques de type van der Waals, [2].

Avantages	Inconvénients
a. La forme polynomiale cubique par rapport au volume facilite relativement le calcul	**a.** Les données PVT réels suivent plutôt une forme de degré 4 par rapport au volume
b. Présentent un comportement aux limites correct, V→b quand P→∞	**b.** Les deux termes, attractif et répulsif, sont imprécis, comme le montre la simulation moléculaire
c. Les imprécisions connues des termes attractifs et répulsifs s'annulent quand les EdE sont employées pour calculer les propriétés des fluides en particulier l'ELV.	**c.** Les EdE cubiques ne peuvent pas représenter toutes les propriétés de fluide dans tous les domaines *de P* et *de T*
d. Pour la plupart des applications, les EdE cubiques peuvent être ajustées pour donner des valeurs précises des propriétés volumétriques ou thermodynamiques	**d.** La dépendance de la température de la constante du terme attractif n'est pas bien connue ; le Covolume *b* semble dépendre de la densité, mais cette dépendance n'est pas encore établie.
e. L'extension des EdE aux mélanges est relativement facile par l'utilisation des règles de combinaison et de mélange de toutes sortes de complexité	**e.** Puisque les interactions entre des molécules différentes sont inconnues, la plupart des règles de mélange et de combinaison sont empiriques, et nécessitent des paramètres d'interaction binaires à ajuster.
f. Les EdE cubiques conviennent à l'application de règles de mélange modernes qui incluent les modèles de l'énergie libre de Gibbs ou des paramètres dépendant de la concentration	**f.** Dans les applications aux mélanges complexes, plusieurs paramètres d'interaction binaires pourraient être nécessaires, même avec l'utilisation de règles de mélange modernes.

I.4.2. Modélisation par les modèles semi-empiriques

La solubilité d'un solide pur non volatil dans un fluide supercritique (y_2), est déterminée à partir des relations thermodynamiques classiques en comparant les fugacités dans la phase solide et dans la phase supercritique pour chaque composant. Ce critère

d'isofugacité pour ce système binaire est réduit à une équation, car le solvant est supposé insoluble dans la phase solide :

$$f_i^S = f_i^{SC} \quad \text{(I.12)}$$

La fugacité de la phase solide est donnée par :

$$f_2^S = P_2^{sub}\phi_2^{sat}\, exp[\,\tfrac{V_2^S}{RT}\,(P - P_2^{sub})\,]\,] \quad \text{(I.13)}$$

La fugacité du soluté solide dans la phase supercritique est donnée par :

$$f_2^{SC} = y_2\phi_2^{SC}P \quad \text{(I.14)}$$

Ici, ϕ_2^{SC} est le coefficient de fugacité du soluté solide dans la phase supercritique.

Combinaison d'équations. (I.12) - (I.14) donne l'expression de la solubilité de la fraction molaire :

$$y_2 = \frac{P_2^{sub}}{P}\frac{\phi_2^{sat}}{\phi_2^{SC}}\exp\left[\frac{V_2^S}{RT}(P - P_2^{sub})\right] \quad \text{(I.15)}$$

Équation (I.15) montre que la solubilité est le produit du solubilité idéale $\frac{P_2^{sub}}{P}$ et un terme de correction connu sous le nom de facteur de rehaussement qui reflète la non-idéalité de la phase supercritique par le rapport $\frac{\phi_2^{sat}}{\phi_2^{SC}}$ et l'effet de la pression sur la fugacité du solide, comme il est connu pour les liquides, par le terme exponentiel appelé facteur de Poynting. De plus, pour les solides de très faible volatilité, comme c'est le cas avec de nombreux médicaments solides, la pression de sublimation est très basse et les valeurs du coefficient de fugacité saturée du solide, ϕ_2^{sat}, sont très proches de un. L'expression de la solubilité solide du soluté est ensuite réduite à

$$y_2 = \frac{P_2^{sub}}{P}\frac{1}{\phi_2^{SC}}\exp\left[\frac{V_2^S}{RT}(P - P_2^{sub})\right] \quad \text{(I.16)}$$

En fait, cette équation donne la solubilité sous une forme implicite parce que ϕ_2^{SC} est une fonction de y_2, T, P et Z^{SCF} (le coefficient de compressibilité de la phase supercritique), elle-même fonction de y_2, T et P.

$$\phi_2^{SC} = \phi(\mathrm{T}, \mathrm{P}, Z^{SCF}, y_2) \quad \text{(I.17)}$$

$$Z^{SCF} = Z(T, P, y_2) \quad \text{(I.18)}$$

Les équations. (I.16)-(I.18) constituent la base de calcul de la solubilité d'un soluté solide dans un SCF par un EdE. La qualité des prévisions ou des corrélations dépend de l'EdE, et même pour le même EdE, cela dépend aussi des règles de mélange. De plus, cette méthode nécessite les propriétés du soluté, et les calculs sont très sensibles à la précision des propriétés du soluté, [30,31]. Si on prend le logarithme de l'équation (I.16) :

$$\ln y_2 = \ln P_2^{sub} - \ln P - \ln \phi_2^{SC} + \frac{V_2^s}{R}.\frac{P}{T} - \frac{V_2^s P_2^{sub}}{R}.\frac{1}{T} \quad \text{(I.19)}$$

Pour T et P donnés, les propriétés volumétriques du solvant sont régies par l'équation (I.20) :

$$P = ZRT\rho_1 \quad \text{(I.20)}$$

La combinaison des équations (I.19) et (I.20), on peut relier la solubilité du soluté à la densité du solvant et au coefficient de compressibilité du solvant pur à la température et à la pression considérée :

$$\ln y_2 = \ln P_2^{sub} - \ln(ZRT\rho_1) - \ln \phi_2^{SC} + V_2^s\, Z\rho_1 - \frac{V_2^s P_2^{sub}}{R}.\frac{1}{T} \quad \text{(I.21)}$$

Bien que l'équation (I.21) donne la solubilité sous une forme implicite complexe, comme mentionné précédemment, elle montre comment la solubilité du soluté pourrait être liée à la densité du solvant supercritique. Un modèle basé sur la densité peut être considéré comme une simple approximation empirique du côté droit de l'équation (I.21) qui ne nécessite aucun paramètre de soluté contrairement aux modèles EdE.

Puisque les propriétés volumétriques (P, T, ρ) à l'état supercritique sont régies par la règle de phase ainsi que la relation est :

$$Z = \frac{PV}{RT} = \frac{P}{RT\rho} \quad \text{(I.22)}$$

Ou Z est le coefficient de compressibilité et R la constante universelle des gaz parfaits.

Ceci signifie qu'à une température donnée, la pression est proportionnelle à la densité :

$$P = ZRT\rho \quad \text{(I.23)}$$

Les modèles fondés sur la densité pour la corrélation de la solubilité d'un soluté solide dans une phase dense ont considérablement évolué depuis les travaux de Stahl et al. (1978) [32].

La section suivante rapporte l'évolution de la modélisation semi-empirique de la solubilité de solutés solides dans les fluides supercritiques dans les quatre dernières décennies.

I.4.2.1. Modèle de Chrastil

Le modèle de Chrastil (1982) (CH) [33], est le premier modèle basé sur la densité du solvant à être proposé. Il est basé sur la théorie qui suppose qu'à l'équilibre un complexe de solvant est formé par association du soluté dissous et k molécules du solvant (k est le nombre d'association). La théorie de Chrastil a menée au développement d'un modèle qui relie la

solubilité du corps dissous et la densité du solvant pur. Cette relation est exprimée par l'équation (I.24) [33]:

$$\ln y_2 = a_0 + a_1 \ln \rho_1 + \frac{a_2}{T} \qquad \text{(I.24)}$$

y_2 la solubilité en fraction molaire du soluté solide, ρ_1 la densité du solvant supercritique, T la température absolue, a_0, a_1 et a_2 les paramètres ajustables du modèle. Le paramètre a_1 est équivalent au nombre d'association, tandis que le paramètre a_2 multiplié par la constante universelle des gaz R représente approximativement la chaleur totale de la réaction qui est la somme de la chaleur de dissolution et de la chaleur de vaporisation ($\Delta H_{totale} = -a_2 R$).

L'équation de Chrastil comporte de nombreuses limitations. Si la solubilité du soluté est très importante (supérieure à 100-200 $kg.m^{-3}$) ou si la plage de température étudiée est trop vaste, l'équation doit être corrigée pour fournir des résultats corrects.

I.4.2.2. Modèle d'Adachi et Lu

Adachi et Lu (1983) (AL) [34] ont modifié l'équation de Chrastil pour améliorer la modélisation de la solubilité de triglycérides. Il faut noter que dans les modèles de Chrastil, le nombre d'association est supposé constant et indépendant de la densité ou de la température. Adachi et Lu (1983) ont modifié le nombre d'association en polynôme du second ordre de la densité du fluide supercritique. Ils ont proposé l'équation:

$$\ln y_2 = a_0 + (a_1 + a_2\rho_1 + a_3\rho_1^2)\ln \rho_1 + \frac{a_4}{T} \qquad \text{(I.25)}$$

En employant l'équation (I.25), ils ont constaté une réduction significative de l'erreur entre la solubilité expérimentale et calculée.

I.4.2.3. Modèle de Del Valle et Aguilira

L'étude de la solubilité menée sur des huiles végétales dans le CO_2 supercritique, Del Valle et Aguilera (1988) (dVA) [35] ont montré que la variation de $(a_0 + a_2/T)$ en fonction de (1/T) (équation (I.25)) pour une série d'isothermes est une droite de pente a, représentant une fonction des enthalpies de dissolution et de la vaporisation du corps dissous. Pour tenir compte du changement de l'enthalpie de la vaporisation avec la température, l'équation modifiée de Chrastil est donnée ci-dessous :

$$\ln y_2 = a_0 + a_1 \ln \rho_1 + \frac{a_2}{T} + \frac{a_3}{T^2} \qquad \text{(I.26)}$$

I.4.2.4. Modèle de Kumar et Johnston

Le modèle de Kumar et Johnston (1988) (KJ) [36] provient de leur observation de la relation qui existe entre le logarithme de la solubilité *($ln(y_2)$)* et le logarithme la densité *($ln(\rho_1)$)* et dans certains cas entre le logarithme de la solubilité *($ln(y_2)$)* et la densité *(ρ_1)* dépendent du système et ni l'une ni l'autre ne peut être généralisée. Ainsi leur modèle se présente sous la forme suivante:

$$\ln y_2 = a_0 + a_1\rho_1 + \frac{a_2}{T} \qquad \text{(I.27)}$$

De même, le paramètre a_2 est lié à la chaleur totale de la réaction qui est la somme de la chaleur de dissolution et de la chaleur de vaporisation ($\Delta H_{totale} = -a_2 R$).

I.4.2.5. Modèle de Bartle et al.

Bartle et al. (1991) (BR) [37] ont proposé un nouveau modèle empirique pour décrire la relation entre la solubilité et la densité du solvant supercritique. Ils ont noté qu'une expression linéaire pourrait exister entre le facteur d'amélioration de la solubilité (E) du soluté dissous et la densité du solvant comme suit :

$$\ln E = k_0 + k_1\rho_1 \qquad \text{(I.28)}$$

Le facteur d'amélioration E qui représente le rapport de la solubilité réelle sur la solubilité idéale est défini comme :

$$E = y_2 P / P_2^{sub} \qquad \text{(I.29)}$$

Où P_2^{sub} est la pression de sublimation du soluté solide. Les paramètres k_0 et k_1 du modèle sont obtenus par régression des données expérimentales. Selon ces auteurs, ce formalisme permet la corrélation de la solubilité du corps dissous aux différentes températures par une seule équation simple. Cependant, l'équation ci-dessus est strictement limitée aux systèmes pour lesquels la pression de sublimation du corps dissous est disponible. Pour cette raison, Bartle et al. ont reformulé l'équation (I.29) en introduisant une pression de référence, P^{ref}, de 1 bar et d'une densité de référence, ρ_1^{ref}, de 700 kg/m^3. Cette dernière a été choisie puisque elle figure habituellement dans le domaine des données expérimentales à toutes les températures. En appliquant la définition du facteur d'amélioration de l'équation (I.28) et la relation $\ln P_2^{sub} = A + B/T$ pour la pression de sublimation, le modèle de Bartle et al. s'écrit :

$$\ln \frac{y_2 P}{P^{ref}} = a_0 + a_1(\rho_1 - \rho_1^{ref}) + \frac{a_2}{T} \qquad \text{(I.30)}$$

Dans ce modèle, le paramètre a_2 peut être directement utilisé pour estimer la chaleur de vaporisation du corps dissous ($\Delta H^{vap} = -a_2 . R$). Basé sur les valeurs de ΔH^{vap} et l'enthalpie totale de la réaction, les enthalpies de dissolution peuvent être calculées pour chaque système soluté solide/solvant.

I.4.2.6. Modèle de Gordillo et al.

Le modèle empirique de Gordillo et al. (1999) (GO) [38] est presque identique à celui de Yu et al. (1994) :

$$\ln y_2 = d_0 + d_1 P + d_2 P^2 + d_3 PT + d_4 T + d_5 T^2 \qquad \text{(I.31)}$$

Les constantes d_0 à d_5 sont les paramètres ajustables du modèle. Les auteurs du modèle ont signalé la supériorité de leur modèle par rapport aux modèles de Chrastil, del Valle et Aguilira et Yu, pour la corrélation des données expérimentales de la solubilité de la pénicilline G dans le CO_2 supercritique.

I.4.2.7. Modèle de Mendez-Santiago et Teja

Le modèle proposé par Méndez-Santiago et Teja (1999) (MT) [39] est basé sur la théorie des solutions diluées pour laquelle le facteur d'amélioration est une fonction de la densité du solvant. Le modèle est exprimé par l'équation :

$$T \ln E = A + B\rho_1 \qquad \text{(I.32)}$$

Les constantes A et B sont les paramètres ajustables à partir des données expérimentales. L'équation de Méndez-Santiago et Teja est une abréviation de l'expression donnée par Harvey (1990) :

$$T \ln E = -[A + B(\rho_1 - \rho_{c,1})] + \frac{V_2 (P - P_2^{sub})}{R} - T \ln \varphi_1 \qquad \text{(I.33)}$$

Où $\rho_{c,1}$ est la densité critique du solvant, V_2 est le volume molaire du soluté solide et φ_1 est le coefficient de fugacité du solvant. Bien que Harvey (1990) a mentionné que le tracé de $T \ln E = f(\rho_1)$ devrait aboutir à une droite si les deux derniers termes de son équation sont négligés, il n'a, cependant, pas testé la validité de cette approximation contre des données expérimentales. Méndez-Santiago et Teja ont constaté la dominance du premier terme et ils ont montré que les constantes A et B sont indépendantes de la température de telle sorte que les données de solubilité à différentes températures se situent sur une ligne droite en accord avec l'équation (I.32). Par conséquent, les auteurs du modèle l'ont proposé comme un excellent test d'uniformité des données de la solubilité pour différentes isothermes.

Bien que Méndez-Santiago et Teja n'aient signalé aucune limite supérieure pour la densité dans l'Eq (I.33) en termes de comportement linéaire, ils ont noté une limite pour les faibles densités égale environ à 0.5 la densité critique du solvant. Dans plusieurs cas, la pression de sublimation du solide n'est pas disponible, pour cela Méndez-Santiago et Teja (1999) ont proposé une autre équation pour le cas où la pression de sublimation est inconnue. Ce modèle est obtenu en substituant la pression de sublimation du soluté solide par une expression à deux constantes du type Clausius–Clapeyron pour la pression de vapeur et qui conduit à l'expression :

$$T \ln y_2 P = A' + B' \rho_1 + C' T \qquad \text{(I.34)}$$

Où $A', B' et C'$ sont les constantes ajustables par régression des données expérimentales.

I.4.2.8. Modèle de Sung et Shim

Sung et Shim (1999) (SS) [15] ont montré que la température a une influence sur la solubilité et ont indiqué que les isothermes de solubilité était linéaire mais leurs pentes diminuaient avec la température. Le nouveau modèle de KJ modifié, fonction de la température est représenté sous la forme suivante :

$$\ln y_2 = (C_0 + \frac{C_1}{T}) \ln \rho_1 + \frac{C_2}{T} + C_3 \qquad \text{(I.35)}$$

Les paramètres C_0 à C_3 sont obtenus par régression des données expérimentales

I.4.2.9. Modèle de Jouyban et al.

Après avoir passé en revue la littérature sur les données expérimentales de la solubilité des solutés solides dans les fluides supercritiques et les modèles existants Jouyban et al. (2002) (JB) [40] ont constaté : *(a)* la relation non linéaire entre le logarithme de la solubilité et pression en conditions isothermes, *(b)* la relation non linéaire entre le logarithme de la solubilité et la température en conditions isobares et *(c)* la relation linéaire entre le logarithme de la solubilité et le logarithme de la densité du solvant dans un certain domaine de pression et de la température. Ils ont alors proposé un modèle empirique qui tient compte de ces différentes constatations :

$$\ln y_2 = M_0 + M_1 P + M_2 P^2 + M_3 P T + M_4 \frac{T}{P} + M_5 \ln \rho_1 \qquad \text{(I.36)}$$

Les paramètres M_0 à M_5 sont obtenus par régression des données expérimentales

I.4.2.10. Modèle de Sparks et al.

Sparks et al. (2008) (SP) [41] ont évalué la capacité des modèles de Chrastil, Del Valle et Aguilera, Adachi et Lu, Méndez-Santiago et Teja, et Bartle et al., à corréler avec précision la solubilité des solutés solides dans les fluides supercritiques. Ils ont choisi dans cette évaluation six systèmes binaires (solide + CO_2 supercritique) avec trois isothermes chacun. Ils ont constaté l'échec de ces modèles quand la densité réduite du solvant est inférieure à un. En outre les modèles de type Chrastil présentent une incohérence dimensionnelle. Ils ont alors reformulé ces modèles sous formes adimensionnelles et ont proposé un nouveau modèle qui porte les solutions à ces deux critiques. Ce modèle est basé sur ceux d'Adachi-Lu et del Valle-Aguilera dont ils présentent, selon les auteurs, une amélioration par rapport au modèle de Chrastil. Cependant il existe des cas ou le modèle d'Adachi-Lu est meilleur que le modèle de del Valle-Aguilera et vice versa, car le premier corrige l'effet de la densité sur la solubilité tandis que le deuxième corrige l'effet de la température sur la solubilité.

La combinaison des deux modèles sous formes adimensionnelles donne:

$$c_2^* = \rho_{r,1}^{(a_0+a_1\rho_{r,1}+a_2\rho_{r,1}^2)} \exp\left(b_0 + \frac{b_1}{T_r} + \frac{b_2}{T_r^2}\right) \quad \text{(I.37)}$$

$$\text{Avec } c_2^* = \frac{c_2}{\rho_{c,1}};\ \rho_{r,1} = \frac{\rho_1}{\rho_{c,1}};\ T_r = \frac{T}{T_{c,1}} \quad \text{(I.38)}$$

a_0, a_1, a_2, b_0, b_1 et b_2 sont les paramètres ajustables à partir des données expérimentales ; c_2 la solubilité, en gramme par litre, du soluté solide dans la phase supercritique ; $\rho_{c,1}, T_{c,1}$ sont respectivement la masse volumique critique et la température critique du solvant.

I.4.2.11. Modèle de Garlapati et Madras

Garlapati et Madras (2009) (GM1) [42] ont modifié le dernier modèle de Chrastil. Ils ont trouvé une relation linéaire entre ln (y_2) et ln (ρ_1 T) dans une certaine gamme de densité et de température et ont relié la solubilité d'un écartement dans SCFS à la densité et à la température

$$\ln y_2 = G_0 \ln \rho_1 + \frac{G_1}{T} + G_2 \quad \text{(I.39)}$$

Où G_0-G_2 sont des paramètres ajustables à partir données expérimentales.

I.4.2.12. Modèle de Garlapati et Madras

Dans une deuxième étude, Garlapati et Madras (2010) (GM2) [43] ont utilisé une base de données de 15 composés avec un total de 727 points expérimentaux pour comparer leur nouveau modèle, cette fois de cinq paramètres, au modèle de Jouyban et al. (2002) sur lequel ils se sont basés. Leur modèle est le suivant :

$$\ln y_2 = M_0 + (M_1 + M_2\rho_1)\ln\rho_1 + \frac{M_3}{T} + M_4 \ln(\rho_1 T) \qquad \text{(I.40)}$$

Où M_0-M_4 sont des paramètres constantes ajustables à partir données expérimentales.

I.4.2.13. Modèle de Jafari Nedjad et al.

Jafari Nedjad et al. (2010) (NJ) [44] eux aussi se sont basés sur les constatations de Jouyban et al. (2002) pour proposer un modèle empirique à quatre paramètres :

$$\ln y_2 = J_0 + J_1 P^2 + J_2 T^2 + J_3 \ln\rho_1 \qquad \text{(I.41)}$$

J_0-J_3 sont les paramètres du modèle ajustables à partir des données expérimentales.

I.4.2.14. Modèle de Ch et Madras

Basé sur la théorie d'association Ch et Madras (2010) (ChM) [45] ont présenté un modèle à quatre paramètres :

$$y_2 = \left(\frac{P}{P^*}\right)^{(k-1)} \exp\left(\frac{a}{T} + b\rho_1 + c\right) \qquad \text{(I.42)}$$

Où P^* est une pression de référence égale à 0.1 MPa, et k, a, b et c sont les paramètres du modèle ajustable à partir des données expérimentales.

I.4.2.15. Modèle de Bian et al.

Dans leur étude Bian et al. (2011) [46] ont utilisé une base de données de 54 composés avec un total de 1321 points expérimentaux pour comparer 11 modèles semi-empiriques ((Chrastil, Adachi-Lu, del Valle-Aguilera, Kumar-Johnston, Méndez Santiago-Teja, Gordillo, Jouyban, Sparks, Garlapati-Madras, et Ch-Madras). Ils ont ensuite utilisé ces résultats ainsi que les constatations des études précédentes pour proposer un nouveau modèle qui a donné la meilleure qualité de corrélation pour la base de données considérée. Leur modèle comporte cinq paramètres :

$$c_2 = \rho_1^{(e_0 + e_1\rho_1 + e_2/\ln T)} \exp\left(\frac{a + m\rho_1}{T} + b\right) \qquad \text{(I.43)}$$

Ou e_0, e_1, e_2, a, m, et b sont les six paramètres du modèle et C_2 la solubilité du soluté solide dans le CO_2 supercritique en g/litre.

I.4.2.16. Modèle de Keshmiri et al.

Afin de fournir des calculs de solubilité plus précis en utilisant moins de paramètres ajustables, Keshmiri et al. (2014) [47] ont proposée l'équation empirique suivante :

$$\ln y_2 = a_0 + \frac{a_1}{T} + a_2 P^2 + \left(a_3 + \frac{a_4}{T}\right)\ln \rho_1 \tag{I.44}$$

a_0, a_1, a_2, a_3, a_4 : sont les paramètres du modèle ajustable à partir des données expérimentales.

I.4.2.17. Modèle d'Amooey

La corrélation des données expérimentales sur la solubilité à l'aide des équations précédentes a donné des résultats médiocres. Ali Akbar Amooey (AAA) (2014) [48] a tenté de présenter une approche permettant de corréler la solubilité des solutés dans le CO_2 avec une précision élevée, par comparaison avec d'autres méthodes empiriques utilisant la densité et la température.

L'équation (I.45) présente la corrélation pour l'estimation de la solubilité du médicament en fonction de la densité de CO_2 (ρ) et de la température (T) au moyen d'analyses de régression multiple. Les unités de densité (ρ) et de température (T) sont exprimées respectivement en $Kg.m^{-3}$, Kelvin. Cette corrélation a neuf paramètres ajustables (a_0 - a_8).

$$\ln y_2 = \frac{a_0 + a_1/\rho_1 + a_2/\rho_1^{\,2} + a_3 \ln T + a_4 (\ln T)^2}{1 + a_5/\rho_1 + a_6 \ln T + a_7 (\ln T)^2 + a_8 (\ln T)^3} \tag{I.45}$$

I.4.2.18. Modèle de Hozhabr et al.

Hozhabr et al. (2014) [49] ont réarrangé l'équation de Mendez-Santiago et Teja et ont inséré une nouvelle constante. La nouvelle équation empirique établie pour corréler les données de solubilité dans SCF est représentée comme suit:

$$\ln y_2 = a_0 + \frac{a_1}{T} + \frac{a_2 \rho_1}{T} - a_3 \ln P \tag{I.46}$$

Où a_0 - a_3 sont les constantes du modèle pouvant être obtenues en ajustant des données de solubilité expérimentales. La nouvelle équation montre la relation non linéaire entre ln y_2 et ln P dans des conditions isothermes, une relation non linéaire entre ln y_2 et la température dans des conditions isobares, une relation linéaire entre ln y_2 et la densité dans une certaine plage de pression et de température.

I.4.2.19. Modèle de Khansary et al.

Khansary et al. (2015) [50] ont basé sur les relations fondamentales existantes et ont été considéré comme des règles empiriques pour le nouveau modèle développé:

a. Relation non linéaire entre les conditions inisothermiques $ln(y_2)$ et la pression.
b. Relation non linéaire entre les conditions inisobariques de $ln(y_2)$ et de la température.
c. Relation non linéaire entre $ln(y_2)$ et la température lorsque la densité est constante.
d. Relation non linéaire entre les conditions inisothermiques $ln(y_2)$ et la densité.
e. Relation linéaire entre $ln(y_2)$ et $\ln \rho$.

Afin de fournir des calculs de solubilité plus précis, le développement d'un modèle précis a été proposé en comparaison avec d'autres modèles de la littérature.

$$\ln y_2 = \frac{a_0}{T} + a_1 P + \frac{a_2 P^2}{T} + \left(a_3 + a_4 P\right)\ln \rho_1 \tag{I.47}$$

Où $a_0 - a_4$ sont les paramètres ajustables.

I.4.2.20. Modèle de Bian et al.

Bian et al. (2016) [51] ont sélectionné 3 séries de données expérimentales, afin d'obtenir les relations linéaires. Selon le système considéré, il est possible d'observer l'un des comportements suivants:

a. Relation linéaire entre $ln(y_2)$ et $ln(\rho_1)$ dans des conditions isothermiques.
b. Relation linéaire entre $ln(y_2)$ et ρ_1/T dans des conditions isothermes.
c. Relation linéaire entre $ln(y_2)$ et $\rho_1\ ln(\rho_1)$ dans des conditions isothermes.

Afin d'obtenir une équation flexible, capable de décrire tous les problèmes de comportements décrits, l'équation empirique suivante est proposée pour corréler la solubilité des composés solides dans le CO_2-SC:

$$\ln y_2 = a_0 + \frac{a_1}{T} + \frac{a_2 \rho_1}{T} + \left(a_3 + a_4 \rho_1\right)\ln \rho_1 \tag{I.48}$$

I.4.3. Modélisation par les méthodes d'intelligence computationnelle

I.4.3.1. La modélisation par les RNA

La connaissance expérimentale de la solubilité de solutés solides dans le dioxyde de carbone supercritique est le plus souvent représentée par les grandeurs mesurables telles que la température et la pression ainsi que la composition des phases en équilibre. D'autre part, les relations qui lient les variables d'état d'un système donné, décrivant la solubilité de solutés solides dans le dioxyde de carbone supercritique, sont régies par la règle de phases qui définit la variance ou le degré de liberté du système.

Bien que les RNA font partie des modèles de boites noires qui représentent une relation empirique entre des ensembles d'entrées et de sorties, le choix de variables d'entrées et de sorties est décisif pour que le modèle neuronal se rapproche des modèles déterministes.

Par conséquent la sélection des données expérimentales de la solubilité de solutés solides dans le dioxyde de carbone supercritique et des données supplémentaires des corps purs, le choix des variables d'entrées et de sorties et le prétraitement des données constituent la première étape de la modélisation neuronale des équilibres de phases.

L'étape suivante concerne le choix des paramètres du réseau de neurones qui vont permettre au réseau d'apprendre à partir des données sélectionnées la relation entre les entrées et les sorties choisies.

La troisième étape est consacrée à l'opération d'apprentissage et le test de généralisation du réseau choisi. Enfin la performance du réseau peut être, éventuellement, comparée avec celles des autres types de modélisation.

La figure (I.9) résume les étapes de la méthodologie utilisée dans ce travail pour la modélisation neuronale des propriétés critiques et de la solubilité de solutés solides dans le dioxyde de carbone supercritique.

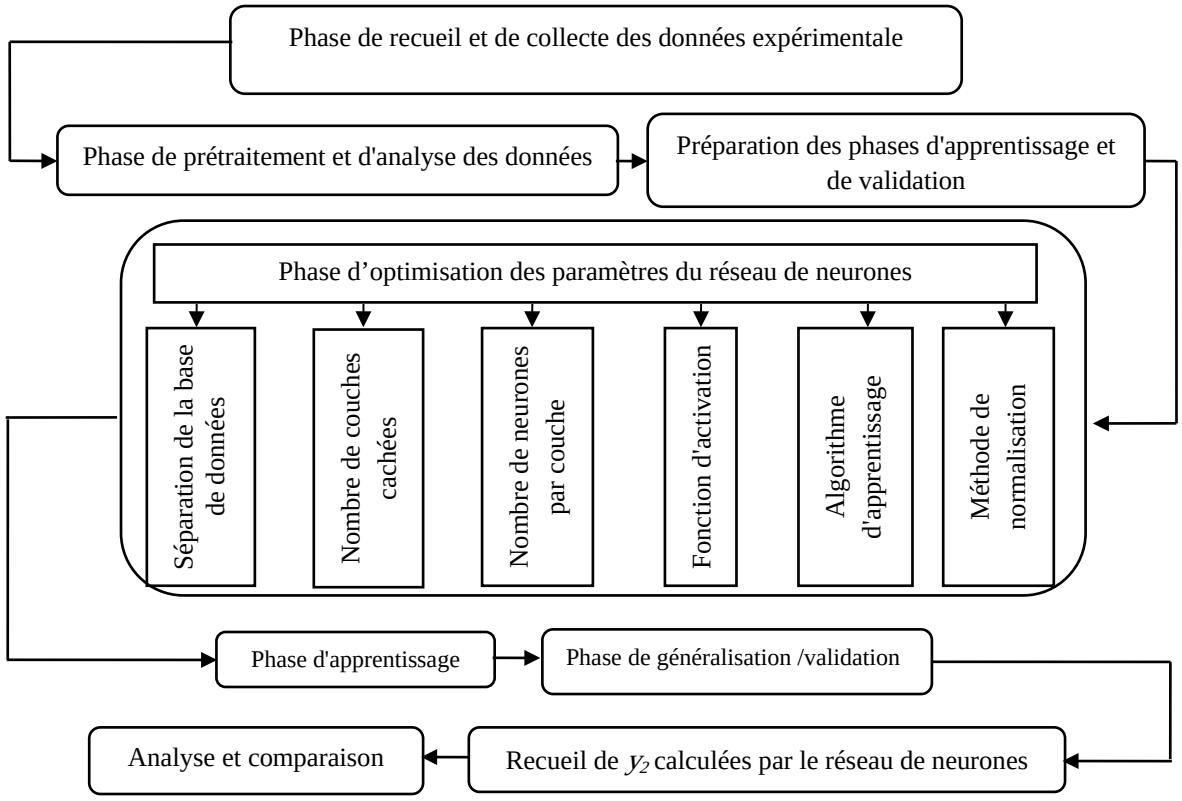


Figure I.9. Etapes de la procédure de modélisation neuronale, [2].

Durant les dernières années, la littérature scientifique relative aux réseaux de neurones artificielles s'est enrichie considérablement par la publication des travaux relatant des

applications diverses telles que la modélisation de la solubilité, nous citons les travaux suivants :

Abdallah el hadj et al. [4] ont travaillé avec la méthode hybride basée sur les neurones d'optimisation des réseaux et des particules (EOP) et a été appliqué aux données de la littérature pour développer et valider un modèle qui peut prédire avec précision la solubilité de binaire systèmes (CO_2 + médicaments solides). RNA a été utilisée pour la modélisation le processus non linéaire. Le PSO a été utilisé pour deux fins: remplacer la rétropropagation standard pour former l'RNA et d'optimiser le processus. L'apprentissage et la validation a été axée sur l'utilisation d'un vecteur de validation. Après ils ont fait une comparaison en termes de l'ERAM% entre les résultats prévus pour chaque binaire pour toute les plages de température et de pression et les résultats prédits par les modèles basés sur la densité (CH, BR et MST) et un ensemble d'équations d'état (VP-SAFT, PR et RSK), montre que le modèle RNA–PSO avec configuration optimisée, les cinq neurones en entrée et les couches cachées et un neurone dans la couche de sortie corrèlent beaucoup mieux la solubilité des médicaments solides dans CO_2-SC. Une stratégie de contrôle a également été faite. Les entrées ont un effet important sur la solubilité et a permis l'estimation de certaines propriétés du solide à partir de la solubilité donnée avec une bonne précision sans nécessité de la méthode de contribution du groupe, [4].

Khazaie Poul et al. (2016) [52] ont utilisé les RNA optimisés comportant 7 entrées, température de composé, pression, température critique, pression critique, densité, poids moléculaire et facteur acentrique, ont été utilisé pour la prédiction de la solubilité de trois colorants dispersés dans le CO_2-SC et l'éthanol comme co-solvant. L'analyse de sensibilité numérique et l'équation de Garson ont été utilisées pour déterminer le degré d'efficacité de différentes variables d'entrées sur l'efficacité du modèle proposé. Les résultats ont montré que le modèle de RNA proposé avait un coefficient de corrélation, un coefficient d'efficacité du modèle de Nash–Sutcliffe et un taux de discordance d'environ 0.998, 0.992 et 1.053, respectivement.

Aminian (2017) [53] a proposé un nouveau modèle basé sur le réseau neuronal pour corréler la solubilité des solutés dans le CO_2-SC. La solubilité de 15 solutés dans le CO_2-SC à différentes températures et pressions est estimée. 4 modèles de réseau de neurones ont été testés pour étudier la capacité de généralisation de chaque réseau. Les résultats de différents modèles semi-empiriques, à savoir CH, KJ, BR et MST, les modèles d'équation d'état de Peng-Robinson et Soave-Redlich-Kwong sont comparés à ceux estimés en utilisant le modèle

proposé. L'erreur absolue moyenne (EAM) de 5,42% par rapport à ceux des modèles basés sur la densité et à l'équation d'état, représentant la précision raisonnable du modèle développé pour estimer la solubilité des solutés dans le processus d'extraction par fluide supercritique.

I.4.3.2. Méthode des machines à vecteurs de support (SVM)

En anglais Support Vector Machine (SVM) et Least Square Support Vector Machine (LSSVM), les machines à vecteurs de support est une approche intelligente a été introduite pour la première fois par Vapnik (1998) [54] sont basées sur la théorie de l'apprentissage statistique (SLT) et sur le principe de minimisation du risque structurel (SRM). Les SVM obtiennent la solution en résolvant la programmation quadratique (QP), tout en évitant les minima locaux et constituent un avantage par rapport aux autres techniques de régression [55]. Comme mentionné précédemment, il peut être très difficile de trouver le modèle SVM final car il nécessite la résolution d'un ensemble d'équations non linéaires (programme quadratique). Dans un souci de simplification, Suykens et Vandewalle (1999) [56] ont proposé une version modifiée de SVM appelée machine à vecteurs de support de moindre carrée (LS-SVM) qui aboutissait à un ensemble d'équations linéaires au lieu d'un programme quadratique. L'algorithme standard LS-SVM a été décrit comme suit :

$$y(x) = \sum_{k=1}^{N} \alpha_k k(x, x_k) + b \qquad \text{(I.49)}$$

où (b, α) est la solution du système linéaire. Dans la littérature, un certain nombre d'introductions complètes à SVM sont disponibles, et la théorie de LSSVM a également été décrite en détail [57].

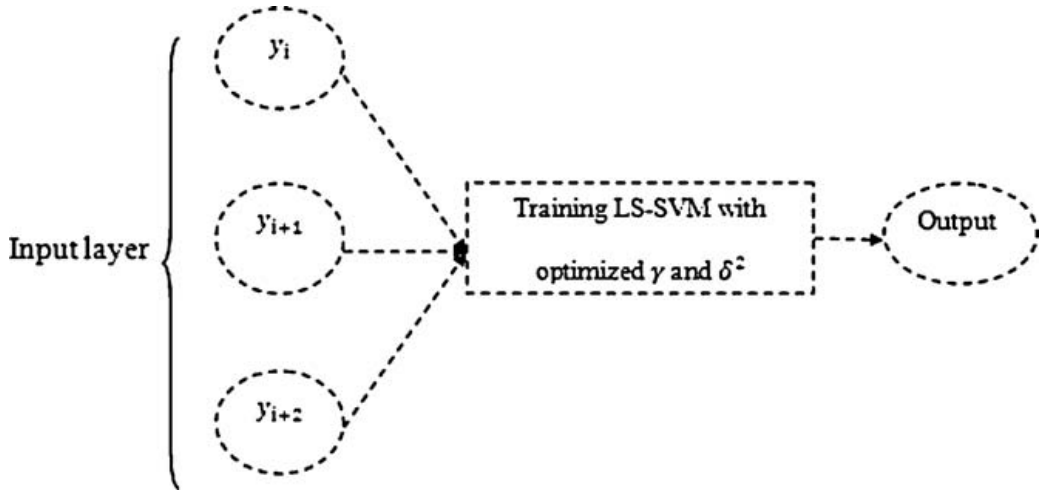


Figure I.10. Topologie générale du modèle LS-SVM, [58].

Le passage en revue de la littérature réalisée dans ce travail fait sortir et remarquer qu'il n y a pas assez de travaux qui fait utilisent ces deux méthodes SVM et LSSVM. Les deux études suivantes sont les plus récentes :

Mehdizadeh et Movagharnejad (2011) [58] ont utilisé la modélisation par LSSVM et ils ont basée sur l'algorithme génétique qui est utilisée pour prédire la solubilité de 25 solutés différents dans le dioxyde de carbone supercritique. Ce modèle comprend trois entrées, dont la température, la pression et la densité du dioxyde de carbone supercritique, ainsi qu'une sortie unique représentant la solubilité de différents solutés dans le CO_2-SC. Les prévisions du modèle ont été comparées aux résultats de sept corrélations semi-empiriques bien connues (CH, dVA, MST, YU, GO, JB et NJ).

Les résultats ont montré que la méthode actuelle présentait une erreur relative moyenne d'environ 4,92% pour 25 solutés alors que la meilleure équation semi-empirique (modèle de JB) aboutissait à une erreur relative moyenne d'environ 13,60% pour les mêmes solutés.

Duan et al. (2018) [59] ont fait une corrélation pour déterminer la solubilité du médicament dans des fluides supercritiques, c'est-à-dire le CO_2 supercritique, la technique de LSSVM, a été utilisée pour estimer la solubilité de 33 composés médicamenteux différents dans le CO_2 supercritique. La solubilité correspondante a été déterminée en fonction de la température, de la pression, de la densité de CO_2-SC et de deux propriétés physiques du médicament, à savoir le poids moléculaire et le point de fusion. Les résultats obtenus à partir du modèle LSSVM utilisé ont été comparés à huit corrélations (CH, BR, SS, AL, SP, GO, MST et Keshmiri). L'erreur relative absolue moyenne obtenue et le coefficient de régression pour le test du modèle LSSVM étaient respectivement de 5,61% et 0,9975. Par conséquent, le modèle développé dans ce travail peut être appliqué de manière fiable aux médicaments étudiés en ne connaissant que leurs propriétés physiques.

I.4.3.3. La modélisation par le système d'inférence adaptative neuro-flou (ANFIS)

En anglais Adaptive Neuro-Fuzzy Inference System (ANFIS), et une méthodes intelligentes comprenant le RNA, modèles de (ANFIS), LS-SVM, sont d'une robustesse incroyable et des outils fiables d'analyse et d'interprétation des données pouvant être utilisés pour prévoir les problèmes de régression et de classification. Le concept de logique floue (FL), introduit pour la première fois par Zadeh (1965) pour gérer les systèmes complexes et non linéaires, a suscité beaucoup d'attention dans la modélisation et la résolution de différents problèmes dans différents domaines de la science. L'algorithme ANFIS est composé de

plusieurs réseaux à réaction directe, qui utilisent des combinaisons de RNA et de FL pour mapper un domaine d'entrée sur un domaine de sortie [60]. Cet algorithme a un grand potentiel pour la résolution et la modélisation de problèmes non linéaires en raison de sa capacité à coupler le pouvoir verbal de la logique floue et le pouvoir numérique des RNA [61,62]. Cet algorithme présente également une capacité remarquable d'apprentissage, de tri, de création et d'extension. Il a une structure basée sur des règles, qui permet d'exploiter des règles floues à partir de données numériques ou de créer ces règles conformément aux données correspondantes. Cet algorithme présent également plusieurs inconvénients, le principal étant que la phase d'apprentissage de l'algorithme et l'évaluation des paramètres connexes prennent généralement beaucoup de temps Chang F-J et Chang Y-T (2006). La figure (I.11) illustre la structure d'un algorithme ANFIS avec deux paramètres d'entrée X_1, X_2 et un paramètre de sortie Y.

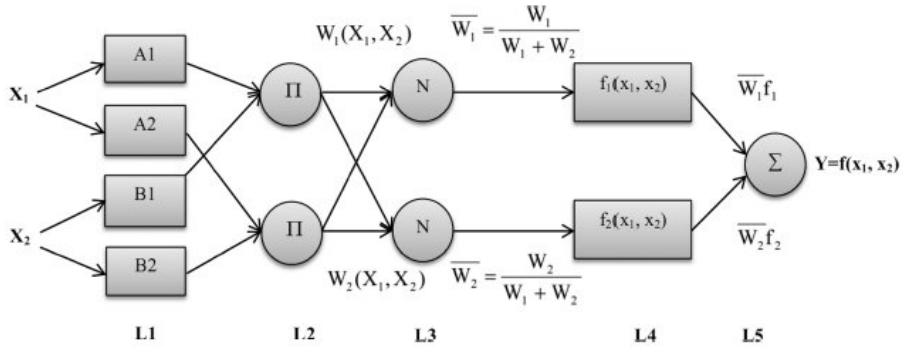


Figure I.11. Structure d'ANFIS comprenant deux paramètres d'entrée Safari et al. (2014)[62].

Dadkhah et al. (2017) [63] ont réalisé une communication où un système d'inférence neuro-fuzzy adaptatif (ANFIS) a été mis au point pour estimer la solubilité de plusieurs espèces solides dans le CO_2 supercritique. Un total de 795 points de données ont été recueillis dans la littérature et utilisés dans le processus de développement et d'évaluation du modèle. L'algorithme génétique (GA) a été mis en œuvre pour optimiser le rayon d'influence du système d'inférence fuzzy (FIS) initial. Les paramètres d'ajustement de la structure FIS ont été optimisés en utilisant une nouvelle méthode basée sur la conjugaison de la méthode hybride et de la méthode d'optimisation d'essaims de particules (PSO).

La figure (I.12) résume les déférentes méthodes utilisées pour la modélisation des propriétés critiques et de la solubilité de solutés solides dans le dioxyde de carbone supercritique

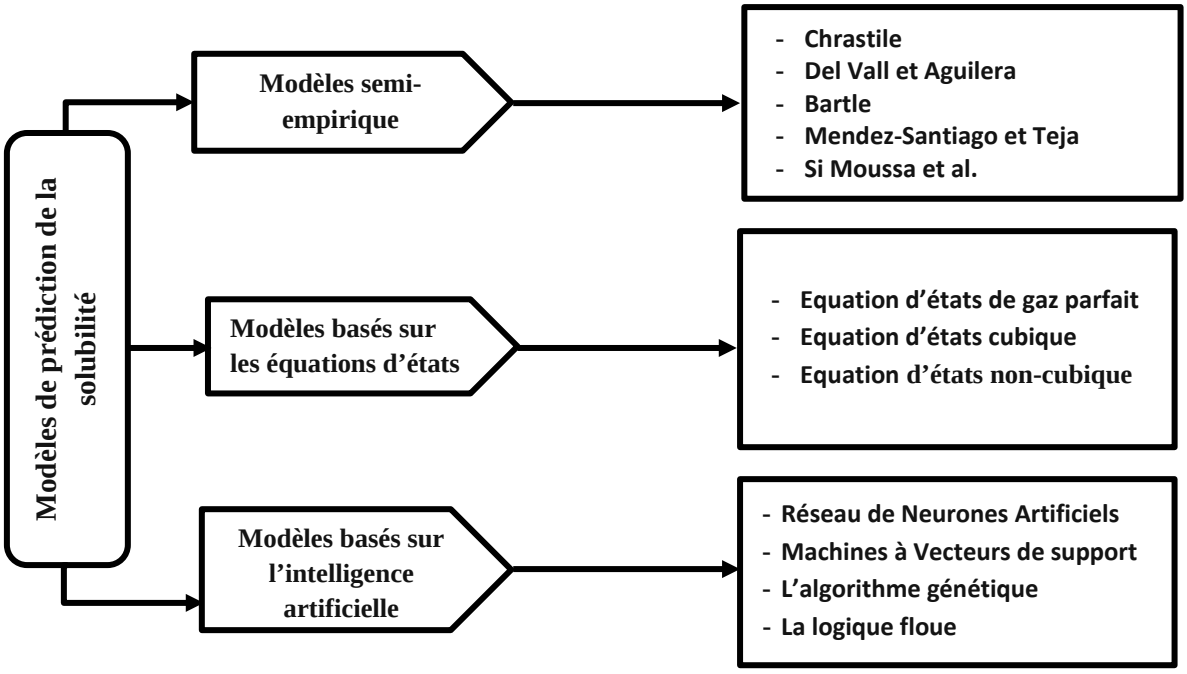


Figure I.12. Classification des différents modèles de prédiction de la solubilité des solides dans les fluides supercritiques.

I.5. Conclusion

Cette généralité montre que la technologie des fluides supercritiques n'a pas cessé d'évoluer proposant plusieurs procèdes utiles pour l'industrie pharmaceutique. Le point crucial dans toutes les technologies à base de fluide supercritique, notamment l'ingénierie des particules pharmaceutiques, est la connaissance de la solubilité des composés dans le fluide supercritique à différentes températures et pressions. À cet égard, de nombreux chercheurs et analystes de laboratoire ont mesuré et rapporté solubilité à l'équilibre de différentes substances dans les fluides supercritiques en particulier le dioxyde de carbone dans un large domaine de température et de pression.

Étant donné que la mesure de la solubilité de tous les composés dans le dioxyde de carbone supercritique dans de larges gammes de température et de pression est impossible, les modèles de prédiction et de corrélation de la solubilité des composés ont été largement utilisés au cours des deux dernières décennies.

Chapitre II. Techniques d'intelligence computationnelle & Chemoinformatiques.

Chapitre II
Techniques d'intelligence computationnelle & Chemoinformatiques.

II.1.	Introduction	38
II.2.	Utilités et applications des méthodes QSPR	38
II.3.	Outils de QSPR	39
II.4.	Réseaux de neurones artificiels	42
II.5.	Adaptive Neuro-Fuzzy Inference System (ANFIS)	58
II.6.	Support Vector Machine (SVM)	62
II.7.	Conclusion	68

II.1. Introduction

La chémoinformatique est la discipline dont le but est l'étude et la résolution des problèmes relatifs à la chimie en appliquant des méthodes informatiques. Cette discipline est devenue une véritable interface entre la chimie et l'informatique. En d'autres termes, la grande quantité de données issues des expériences ne peut être traitée et analysée que par des méthodes informatiques [64]. Notre intérêt se portera sur l'application de cette technique inhérente à la prédiction de la solubilité à partir de la structure moléculaire.

Les relations quantitatives structure-propriétés (en anglais QSPR : Quantitative Structure Property Relationship) sont un exemple typique de cette catégorie. Les relations entre les structures des molécules et leurs propriétés sont généralement établies à l'aide de méthodes de modélisation par apprentissage statistique. Les techniques usuelles reposent sur la caractérisation des molécules par un ensemble de descripteurs, nombres réels mesurés ou calculés à partir des structures moléculaires. Il est alors possible d'établir une relation entre ces descripteurs et la grandeur modélisée. Parmi les méthodes d'apprentissage qui sont nombreuses, les réseaux de neurones artificiels (RNA ou ANN en anglais : Artificial Neural Networks) et la machines à vecteurs de support (SVM).

II.2. Utilités et applications des méthodes QSPR

Initiées sur la base de relations empiriques durant la seconde partie du 19ème siècle, les QSPR ont été mathématiquement formalisées dans les années 1930 et leur acceptation, en tant que discipline à part entière, a été définitivement acquise au début des années 1960 suite aux travaux de Hansch et Fujita [65] qui ont proposé un modèle mathématique pour corréler l'activité biologique et la structure chimique. Depuis lors, un grand nombre de modèles a été conçu pour la prédiction de diverses propriétés d'intérêt chimique, physique, ...etc.

Il y'a de nombreuses utilités dans l'utilisation des méthodes *QSPR*. Nous donnons ci-dessous, à titre d'exemple, quelques-unes :

- Prédire les propriétés physico-chimiques et l'activité biologique.
- Comprendre et rationaliser les mécanismes d'action dans une série de produits chimiques.
- Economiser le coût de développement de produits (par exemple dans l'industrie pharmaceutique, pesticides, produits d'hygiène personnelle,…etc.).
- Réduire ou même remplacer les tests longs et coûteux sur les animaux.

II.3. Outils de QSPR

Les modèles *QSPR* permettent de quantifier la relation entre la structure et la propriété physicochimique et d'apprécier l'impact de diverses propriétés moléculaires sur cette propriété. L'élaboration de tels modèles *QSPR* repose sur l'utilisation de trois outils : une ou plusieurs données thermochimiques à prédire, des descripteurs moléculaires pour représenter les propriétés des molécules et des méthodes d'apprentissage statistique qu'on introduira dans une partie à part.

II.3.1. Les descripteurs moléculaires

De nombreuses recherches ont été menées, au cours des dernières décennies, pour trouver la meilleure façon de représenter l'information contenue dans la structure des molécules. Ainsi, les recherches ont abouti à la représentation de la structure moléculaire par un ensemble de valeurs numériques appelées descripteurs moléculaires. On en dénombre aujourd'hui plus de 3000 types, qui quantifient des caractéristiques physico-chimiques ou structurelles de molécules. Ils peuvent être obtenus de manière expérimentale, mais les descripteurs calculés, et non mesurés, sont à privilégier : ils permettent en effet d'effectuer des prédictions sans avoir à synthétiser les molécules, ce qui est un des objectifs de la modélisation. Les descripteurs moléculaires peuvent être divisés en deux grandes catégories [66]: les descripteurs expérimentaux (par exemple : la polarisabilité, la densité électronique, le potentiel d'ionisation) et les descripteurs moléculaires théoriques. Ces derniers sont à nouveaux classés en trois catégories en fonction de leur dimension.

Les descripteurs 1D: Ils sont accessibles à partir de la formule brute de la molécule et décrivent des propriétés globales du composé. Il s'agit par exemple de sa composition, c'est-à-dire les atomes qui le constituent, ou de sa masse molaire.

Les descripteurs 2D: Ils sont calculés à partir de la formule développée de la molécule. Ils peuvent être de plusieurs types :

- Les indices constitutionnels caractérisent les différents composants de la molécule. Il s'agit par exemple du nombre de liaisons simples ou multiples, du nombre de cycles…

- Les indices topologiques peuvent être obtenus à partir de la structure 2D de la molécule, et donnent des informations sur sa taille, sa forme globale et ses ramifications. L'indice de Wiener par exemple, permet de caractériser le volume moléculaire et la ramification d'une molécule. Un autre indice, nommé "indice de Randic", est l'un des descripteurs les plus utilisés. Il peut être interprété comme une mesure de l'aire de la molécule accessible au solvant.

Les descripteurs 3D: Ces descripteurs 2D reflètent les propriétés physiques dans la plupart des cas, mais sont insuffisants pour expliquer de façon satisfaisante certaines propriétés ou activités, telles que les propriétés thermochimiques. Aussi, des descripteurs représentant l'information dérivée de la représentation en trois dimensions des molécules ont pu être calculés grâce au développement des techniques instrumentales. Ils sont évalués à partir des positions relatives de ses atomes dans l'espace, et décrivent des caractéristiques plus complexes. Ces descripteurs apportent davantage d'informations, et sont nécessaires à la modélisation de propriétés ou d'activités qui dépendent de la structure 3D. On distingue plusieurs familles importantes de descripteurs 3D:

- **Les descripteurs géométriques** les plus importants sont le volume moléculaire, la surface accessible au solvant, le moment principal d'inertie.
- **Les descripteurs électroniques** permettent de quantifier différents types d'interactions inter et intramoléculaires, de grande influence sur les propriétés physicochimiques. Par exemple, les énergies de la plus haute orbitale moléculaire occupée et de la plus basse vacante sont des descripteurs fréquemment sélectionnés. Par ailleurs, le moment dipolaire, le potentiel d'ionisation, et différentes énergies relatives à la molécule sont d'autres paramètres importants.
- **Les descripteurs spectroscopiques:** les molécules peuvent être caractérisées par des mesures spectroscopiques, par exemples par leurs fonctions d'onde vibrationnelles. En effet, les vibrations d'une molécule dépendent de la masse des atomes et des forces d'interaction entre ceux-ci ; ces vibrations fournissent donc des informations sur la structure de la molécule et sur sa conformation. Les spectres infrarouges peuvent être obtenus soit de manière expérimentale, soit par calcul théorique, après recherche de la géométrie optimale de la molécule. Par exemple, les descripteurs de type MoRSE sont calculés à partir d'une simulation du spectre infrarouge.

II.3.2. Formulation mathématique des QSPR

L'axiome central de la modélisation QSPR 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, les équations *QSPR* prennent la forme générale suivante :

$$\textit{Réponse} = \textit{fonction (Descripteurs)}$$

II.3.3. Logiciels et interfaces utilisés

DRAGON 7.0

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 Roberto Todeschini et al. 2009 [67] (la référence la plus complète pour la théorie des descripteurs).

Interface utilisateur: Dragon offre une interface utilisateur graphique conviviale (sur Windows et Linux) et une interface en ligne pour le traitement par lots de grands ensembles de données.

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.

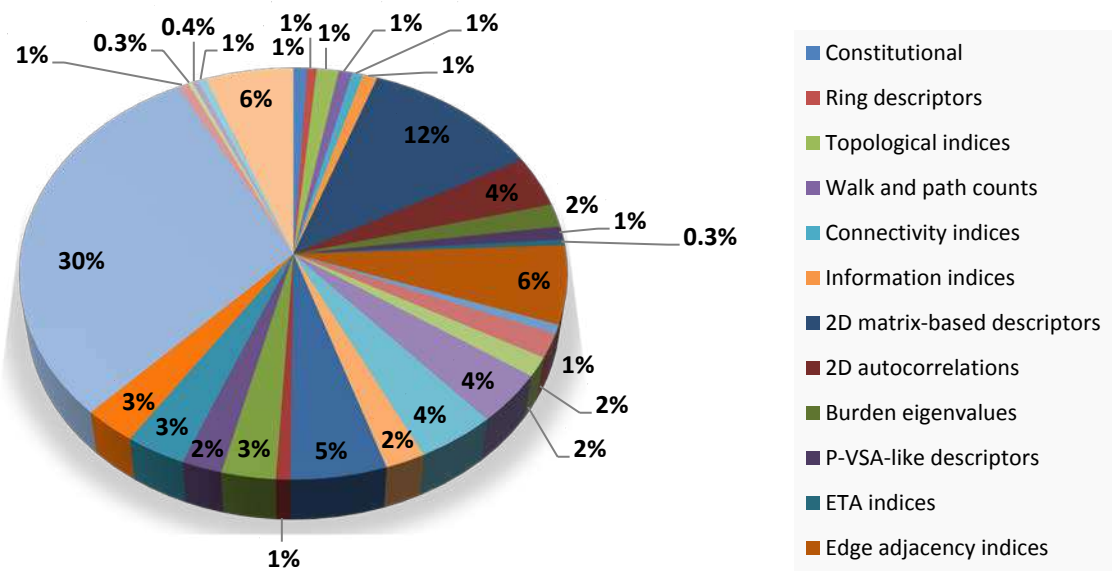


Figure II.1. Pourcentage de chaque type des descripteurs dans la base de données de logiciel DRAGON 7.0.

II.4. Réseaux de neurones artificiels

Malgré la puissance grandissante des calculateurs et la mise en œuvre d'approches théoriques de plus sophistiquées; plusieurs problèmes sont posés dans le cadre de ces recherches résistant encore aux algorithmes et aux méthodes classiques. Ces problèmes sont souvent dépendants de l'environnement et ne répondent pas à des modèles physiques connus.

En abordant par cette définition d'un RNA, reproduite du livre de Haykin, [68] on peut alors, imaginer les vastes domaines d'application potentielle des RNA, concevoir l'intérêt qui leur a été réservé durant les deux dernières décennies, traduit par le nombre énorme de publications, et enfin comprendre pourquoi ils sont devenus un outil robuste de calcul scientifique dans plusieurs disciplines de sciences et d'ingénierie.

En raison de la disponibilité et la richesse bibliographique qui traite les fondements et les applications des RNA il est inutile de présenter avec détails toutes les notions théoriques des RNA, le lecteur peut consulter les excellents ouvrages de Dreyfus [69], Engelbrecht [70].

Par conséquent, on se limite dans ce chapitre à une présentation succincte des notions fondamentales des RNA et du vocabulaire nécessaire à la compréhension des parties suivantes de cette thèse. Nous examinons les facteurs majeurs affectant l'application réussie des RNA, en particulier les perceptrons multicouches (MLP, Multi Layer Perceptron) comme approximateurs universels de fonctions.

II.4.1. Analogie entre neurone biologique et neurone artificiel

L'idée de réseaux neuraux artificiels a été inspirée de la manière par laquelle les neurones biologiques traitent les informations. Ce concept est utilisé pour mettre en œuvre des logiciels de simulations des processus massivement parallèles qui impliquent des éléments de traitement connectés dans une architecture de réseau.

II.4.1.1. Le neurone biologique

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 100 milliards de neurones interconnectés. Bien que les neurones ne soit pas tous identiques, leur forme et certaines caractéristiques permettent de les répartir en quelques grandes classes.

Le neurone est une cellule constituée principalement de trois parties: dendrites, le corps cellulaire et l'axone.

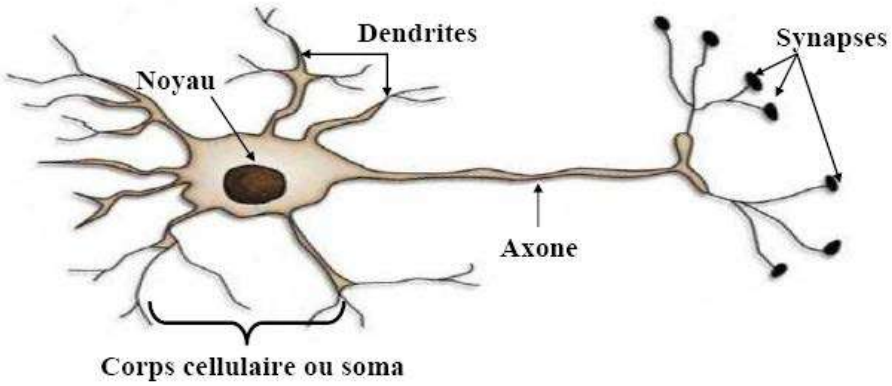


Figure II.2. Schéma d'un neurone biologique.

L'apprentissage dans le cerveau humain se fait à travers un réseau des neurones qui sont connectés par l'axone, les synapses et des dendrites, figure (II.2).

Le neurone biologique est donc une cellule composée d'un corps cellulaire et d'un noyau. Le corps cellulaire se ramifie pour former ce que l'on nomme les dendrites. C'est par les dendrites que l'information est acheminée de l'extérieur vers le soma, corps du neurone.

L'information traitée par le neurone chemine ensuite le long de l'axone (unique) pour être transmise aux autres neurones. La jonction entre deux neurones est appelée la synapse.

II.4.1.2. Réseau de neurones artificiels

Un réseau de neurones artificiels est un processeur parallèle de traitement d'informations distribuées, qui présente une propension naturelle à la mémorisation et à l'exploitation de connaissances relatives à l'environnement dans lequel il est immergé, connaissances acquises à partir de l'expérience.

Le neurone artificiel (figure II.3) reçoit les entrées qui sont analogues aux impulsions électrochimiques que les dendrites de neurones biologiques reçoivent d'autres neurones. Il est donc un opérateur mathématique très simple qui traduit la représentation mathématique d'un neurone biologique. Selon, Dreyfus [69] un neurone est une fonction algébrique non linéaire, paramétrée, à valeur bornée". Il possède des entrées qui peuvent être les sorties d'autres neurones, ou des entrées de signaux extérieurs, et une sortie. La valeur de la sortie résulte du calcul de la somme des entrées, pondérées par des coefficients (dits poids synaptiques) et du calcul d'une fonction non linéaire bornée de cette somme pondérée. Cette fonction (dite fonction d'activation) est le plus souvent une fonction sigmoïde, mais peut également être une gaussienne, un échelon.

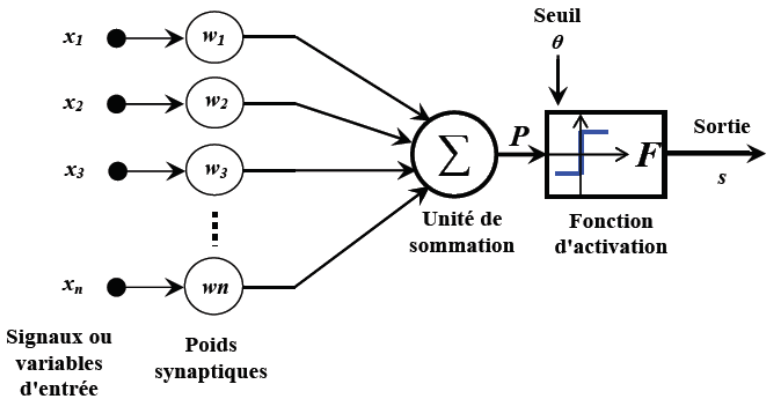


Figure II.3. Schéma d'un neurone artificiel, [69].

Un neurone artificiel peut être défini par les trois éléments suivants:

1. **Entrées "X_i"** : Typiquement, les valeurs d'entrées sont des stimuli externes en provenance de l'environnement ou des sorties d'autres neurones artificiels. Elles peuvent être discrétisées grâce à un ensemble tel que {0, 1} ou bien être des valeurs réelles.
2. **Poids synaptiques "W_i"**: La première chose que fait un neurone formel est de calculer la somme pondérée de ses entrées:

$$p = \sum_{i=1}^{n} w_i x_i \qquad \text{(II.1)}$$

Les poids sont des valeurs réelles qui déterminent la contribution de chaque entrée. Un algorithme d'apprentissage permet de déterminer le meilleur ensemble de poids pour le problème considéré.

3. **Seuil "θ"**: Le seuil est un nombre réel qui est soustrait de la somme pondéré des valeurs d'entrées. Pour des raisons de simplicité, le seuil peut être vu comme une entrée supplémentaire associée à un poids. La fonction seuil interne à chaque neurone artificiel assure que la valeur de sommation des potentiels présynaptiques ne dépassera pas certaines limites raisonnables (en général l'intervalle 0 - 1).

L'analogie entre un réseau de neurones artificiels et le cerveau est suscitée par deux remarques [71]:

- Un réseau de neurones artificiels acquiert la connaissance de son environnement par l'intermédiaire d'un apprentissage qui "stimule" la plasticité du cerveau. En cours d'apprentissage, le réseau peut être amené à modifier sa structure en créant ou supprimant des neurones ou des liaisons entre neurones.
- La connaissance acquise par un réseau de neurones est encodée par les forces ou intensités évolutives des connexions établies entre neurones formels. Les forces des connexions liant les neurones réfèrent aux valeurs des poids ou coefficients synaptique qui définissent les paramètres du modèle interne du réseau.

La figure (II.4) résume la correspondance entre le neurone artificiel et le neurone biologique. On remarque dans cette description schématique, que chaque composant ou constituant du neurone artificiel trouve son inspiration biologique.

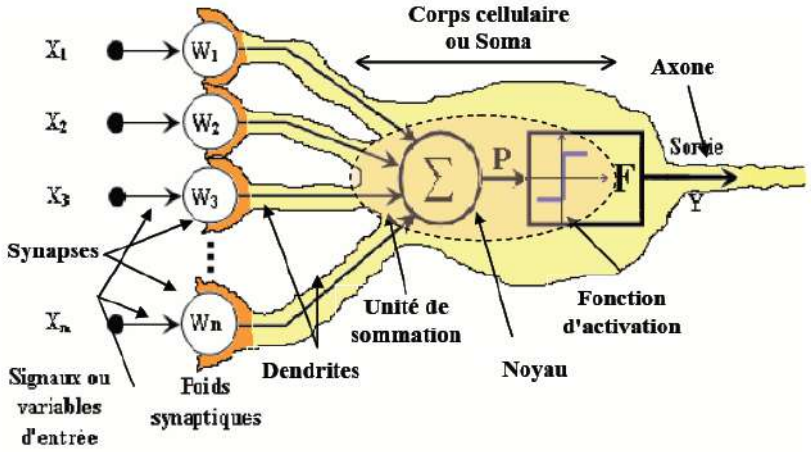


Figure II.4. Schéma de comparaison entre un neurone biologique et un neurone artificiel.

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.

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

II.4.2. Concept du réseau de neurones de type perceptron multicouches

Les neurones artificiels sont arrangés en couches (figure II.5), où la couche d'entrée reçoit des signaux d'entrée (u_i) du monde réel et chaque couche successive reçoit les sorties pondérées ($w_{ij}u_i$) de la couche précédente comme son entrée résultant ainsi en un RNA de type perceptron multicouches, dans lequel chaque sortie d'une couche est alimentée en avant à la couche qui la succède où elle est traitée. Les sorties de la dernière couche constituent les sorties au monde réel.

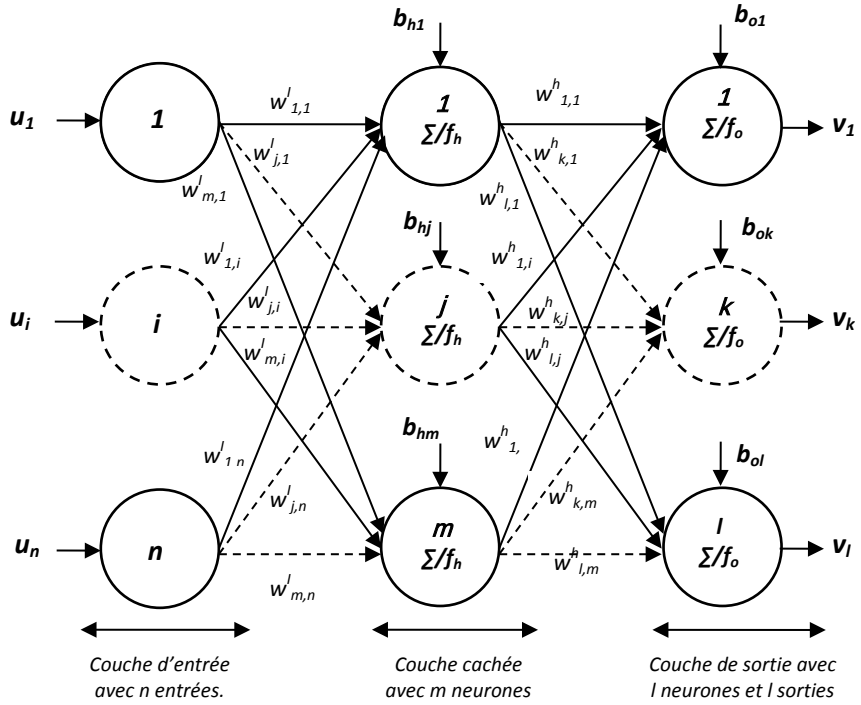


Figure II.5. Réseau de neurones artificiels de type perceptron multicouches [69].

Dans un tel RNA de type perceptron multicouches un neurone, dans une couche cachée ou une couche de sortie, a deux tâches :

- Il additionne les biais aux entrées pondérés de plusieurs connexions et applique ensuite une fonction d'activation (dite aussi fonction de transfert ou encore fonction de seuillage) à cette somme comme donnée par l'équation (II.2) (pour le neurone *j* de la couche cachée de la figure (II.5) :

$$z_j = f_h\left(\sum_{i=1}^{n} w_{ji}^{I} u_i + b_{hj}\right); \qquad j = 1, 2, \ldots, m \tag{II.2}$$

- Il propage la valeur résultante par les connexions partantes aux neurones de la couche suivante où elle subit le même processus tel que donné par l'équation (II.3) (par exemple les sorties de la couche cachée alimentée au neurone *k* de la couche de sortie donnent la sortie v_k) :

$$v_k = f_O\left(\sum_{j=1}^{m} w_{kj}^{h} z_j + b_{ok}\right); \qquad k = 1, 2, \ldots, l \tag{II.3}$$

En combinant des équations (II.2) et (II.3) on obtient la relation entre la sortie v_k et les entrées u_i du réseau de neurones:

$$v_k = f_O\left(\sum_{j=1}^{m} w_{kj}^{h} f_h\left(\sum_{i=1}^{n} w_{ji}^{I} u_i + b_{hj}\right) + b_{ok}\right); \quad k = 1, 2, \ldots, l \tag{II.4}$$

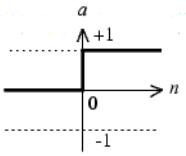
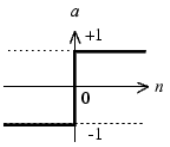
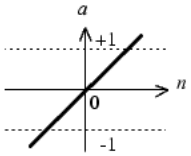
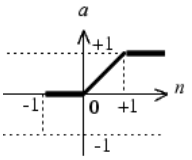
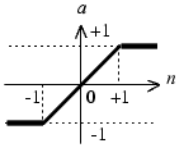
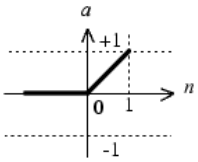
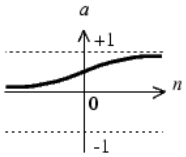
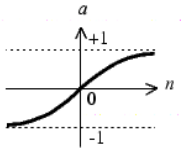
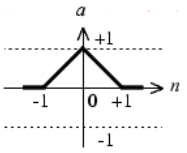
Le nombre de neurones dans la couche d'entrée et la couche de sortie est fixé par le nombre de variables indépendantes et dépendantes respectivement. Le modélisateur définit le nombre de couches cachées et le nombre de neurones dans chaque couche cachée. Le développement du modèle est réalisé par un processus d'apprentissage dans laquelle un ensemble de données et des variables indépendantes est présenté à la couche d'entrée du réseau. Les sorties de la couche de sortie comprennent une prédiction des variables dépendantes. Par conséquent, l'objectif de la modélisation neuronale est de réduire au minimum les erreurs de prédiction de l'ensemble des données de validation présentées au réseau à la fin de l'étape d'apprentissage.

II.4.3. Fonction d'activation

La fonction d'activation calculé 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®[72].

Tableau II.2. Fonctions de transfert usuelles, *a=f(n)*

Nom de la fonction	**Relation entrée / sortie**	**Graphe de la fonction**	**Fonction MATLAB® correspondante**
Seuil	$a = 0$ si $n < 0$ $a = 1$ si $n \geq 0$		hardlim
Seuil symétrique	$a = -1$ si $n < 0$ $a = 1$ si $n \geq 0$		hardlims
Linéaire	$a = n$		purelin
Linéaire saturée	$a = 0$ si $n < 0$ $a = n$ si $0 \leq n \leq 1$ $a = 1$ si $n > 1$		satlin
Linéaire saturée symétrique	$a = -1$ si $n < -1$ $a = n$ si $-1 \leq n \leq 1$ $a = 1$ si $n > 1$		satlins
Linéaire positive	$a = 0$ si $n < 0$ $a = n$ si $n \geq 0$		poslin
Logistique (Sigmoïde logarithmique)	$a = \frac{1}{1+e^{-n}}$		logsig
Tangente hyperbolique	$a = \frac{e^{+n} - e^{-n}}{e^{+n} + e^{-n}}$		tansig
Triangulaire de base	$a = 1 - \lvert n \rvert,\ -1 \leq n \leq +1$ $a = 0,$ *autrement*		tribas

Selon Samarasinghe [73] la fonction linéaire est plus appropriée pour les neurones de sorties alors que les fonctions sigmoïdes sont plus appropriées pour les neurones des couches cachées .

II.4.4. Classification des réseaux de neurones

Selon l'architecture on peut classer les réseaux de neurones en trois grandes catégories :

- Les réseaux de neurones statiques.
- Les réseaux de neurones dynamiques.
- Les réseaux de neurones évolutifs (auto-organisée).

Et selon l'apprentissage on peut classer les réseaux de neurones en deux grandes catégories :

- Apprentissage supervisé.
- Apprentissage non supervisé.

La figure (II.6) résume les différentes possibilités suivantes lesquelles on peut classer les réseaux de neurones.

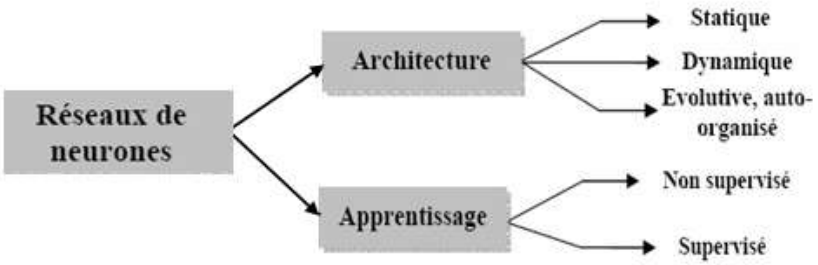


Figure II.6. Les déférentes possibilités de classification des réseaux de neurones.

II.4.4.1. Classification selon la description architecturale

L'architecture du réseau peut être décrite par le nombre de couches et le nombre de neurones dans chaque couche.

L'architecture d'un MLP (figure II.7) 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.

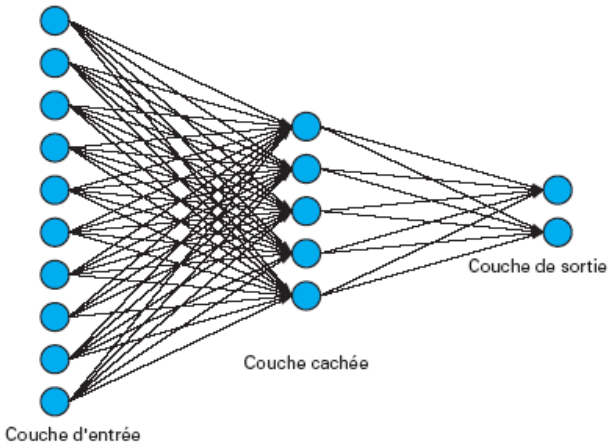


Figure II.7. Réseau multicouches MLP [69].

L'architecture définit le fonctionnement du réseau, il existe trois types de réseaux à architectures différentes :

a. Les réseaux de neurones 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 retour "feed-back" d'information c'est-à-dire l'information circule dans un 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 : Multi Layer Perceptron).
- Les réseaux modulaires.
- Le modèle de Kohonen.

Dans notre étude on va utiliser les réseaux statiques multicouches.

b. 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.

c. Réseaux à architecture auto-organisée

Les réseaux auto organisés sont des réseaux de neurones qui changent leur structure interne pendant l'utilisation. Ainsi les neurones se regroupent topologiquement suivant la représentation des exemples issus de l'espace d'entrée.

II.4.4.2. Classification selon les types d'apprentissages

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. 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.

On distingue deux types d'apprentissage: l'apprentissage supervisé et l'apprentissage non supervisé, selon qu'on fournit ou non au modèle les sorties désirées.

a. 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.8).

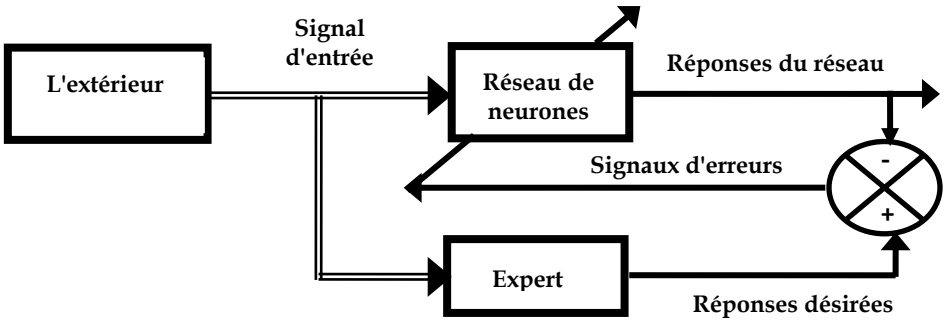


Figure II.8. Illustration de l'apprentissage supervisé [74].

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

b. Apprentissages non supervise

Contrairement à l'apprentissage supervisé effectué sous contrôle d'un expert, l'apprentissage non supervisé est autodidacte.

L'ensemble des exemples d'apprentissages ne comprend que des stimuli. Aucune réponse désirée n'est associée. Cet apprentissage consiste à détecter automatiquement des régularités qui figurent dans les exemples présentés et modifier les poids des connexions pour que les exemples ayant les mêmes caractéristiques de régularités provoquent la même sortie. Ce type d'apprentissages est représenté schématiquement en (figure II.9).

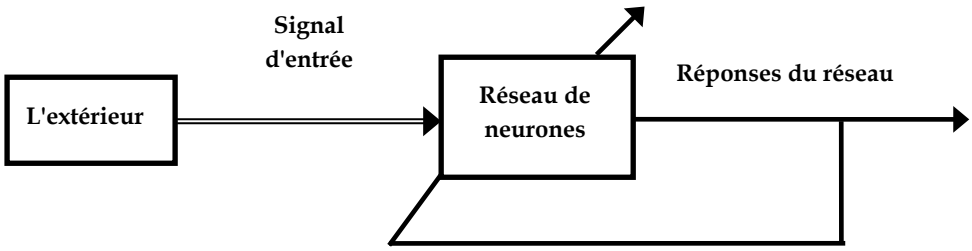


Figure II.9. Illustration de l'apprentissage non supervisé [74].

II.4.5. Procédure d'apprentissage

Dans notre étude, comme nous disposons des valeurs de sortie, nous avons choisi l'apprentissage supervisé, mettant en jeu un algorithme d'optimisation. Ce dernier cherche à minimiser, par des méthodes d'optimisation non linéaire, une fonction de coût qui constitue une mesure de l'écart entre les réponses réelles du réseau et ses réponses désirées, [69].

Cette optimisation se fait de manière itérative, en modifiant les poids en fonction du gradient de la fonction de coût : le gradient est estimé par une méthode spécifique, dite la règle du gradient de l'erreur (delta rule). C'est une des règles les plus utilisées pour l'apprentissage supervisé de réseaux de neurones.

Initialement développée pour résoudre des problèmes de traitements adaptatifs du signal, elle a conduit ensuite à un algorithme connu sous le nom de l'algorithme de rétropropagation du gradient de l'erreur ou encore "Backpropagation"[75].

II.4.5.1. 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. L'organigramme d'apprentissage par l'algorithme de rétro-propagation est présenté comme suit figure (II.10), [76]:

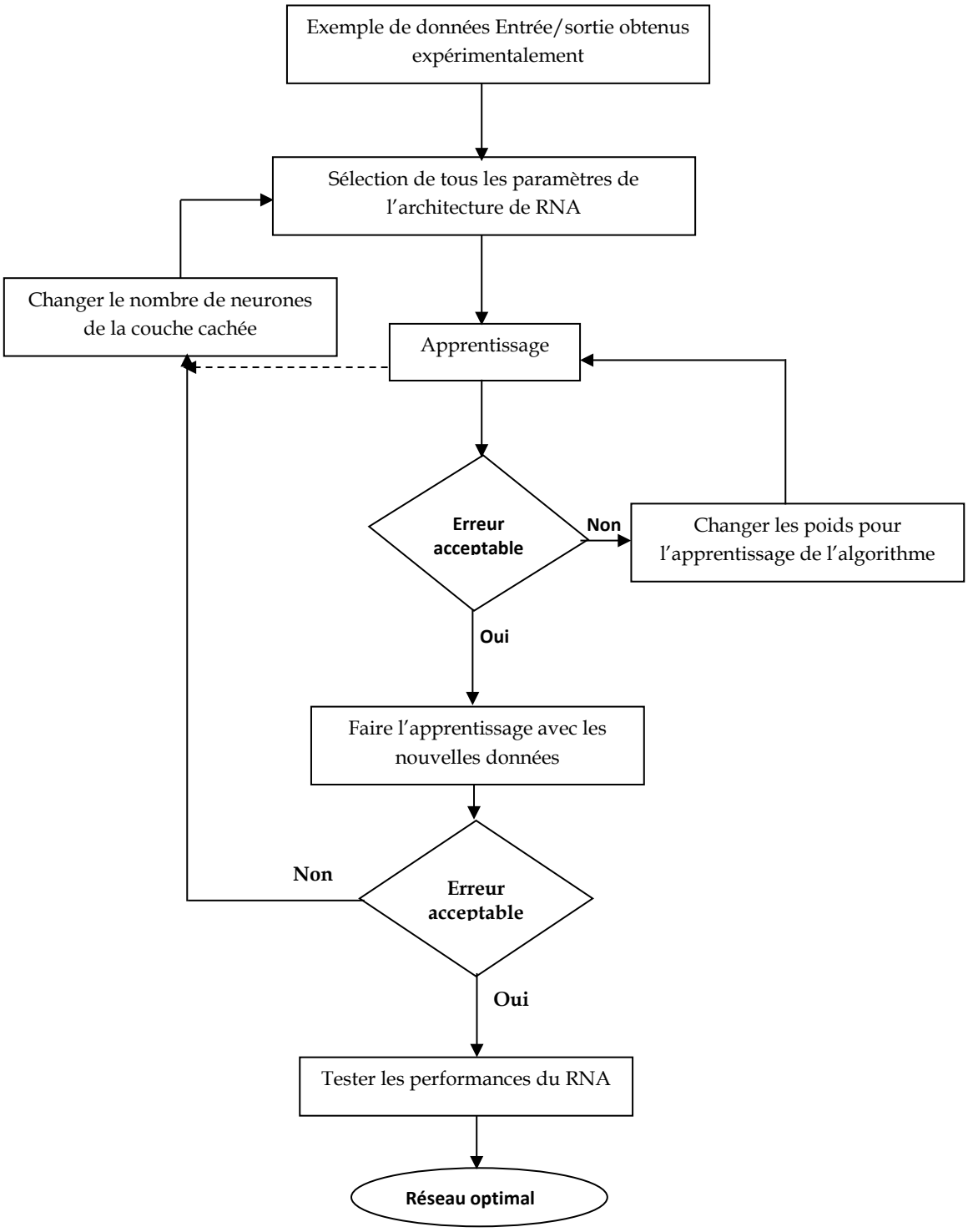


Figure II.10. Organigramme d'apprentissage par l'algorithme de rétro-propagation [76].

Depuis son introduction, l'algorithme de rétro-propagation a été largement étudié et plusieurs modifications y ont été apportées. L'algorithme de base décrit ci-dessus converge très

lentement pour les réseaux multicouches. Les variations apportées à l'algorithme rétro-propagation ont pour objectifs l'accélération de la convergence du processus d'apprentissage et l'amélioration de la capacité de généralisation, [72]. De nombreuses techniques existent, plus ou moins rapide, performant et gourmand en mémoire vive.

Le tableau (II.3) résume les algorithmes d'apprentissage par rétro-propagation implanté dans la boite à outils des RN du programme MATLAB®.

Tableau II.3. Caractéristiques des algorithmes d'apprentissage implantés dans MATLAB®.

Algorithme	Caractéristiques
Descente du gradient	
Descente du gradient simple (***traingd)***	Descente du gradient de base. Réponse lente; peut être utilisée pour apprentissage en mode séquentiel
Descente du gradient avec moment (***traingdm)***	Descente du gradient avec moment. Le moment permet à un réseau de répondre non seulement au gradient local, mais aussi aux tendances récentes dans la surface d'erreur. Sans moment un réseau peut être coincé dans un minimum local peu profond. Plus rapide que ***traingd***, peut être utilisée que pour apprentissage en mode séquentiel
Descente du gradient avec taux d'apprentissage variable (***traingda)***	Taux d'apprentissage variable. Plus rapide que ***traingd***, mais ne peut être utilisée que pour apprentissage en mode batch. Utilise des règles heuristiques pour la variation du taux d'apprentissage
Descente du gradient avec moment et taux d'apprentissage variable ***(traingdx)***	Descente du gradient avec inertie (momentum) et taux d'apprentissage variable
Descente des gradients conjugués	
Scaled conjugate gradient ***(trainscg)***	Le seul algorithme de gradients conjugués qui n'exige aucune recherche de ligne. Très bon algorithme d'apprentissage en général.
Quasi-Newton	
BFGS quasi-Newton ***(trainbfg)***	Il exige le stockage d'une approximation de la matrice Hessienne et nécessite plus de calcul à chaque itération que les algorithmes de gradients conjugués, mais possède habituellement une convergence plus rapide.
Levenberg-Marquardt	
Levenberg-Marquardt ***(trainlm)***	C'est l'algorithme d'apprentissage le plus rapide pour les réseaux de taille modérée. Il a la caractéristique de réduction de mémoire pour l'utilisation quand l'échantillon d'apprentissage est grand.
Régularisation	
Régularisation Bayésienne ***(trainbr)***	Mise à jour des poids et des biais selon l'optimisation de Levenberg-Marquardt. Il réduit au minimum une combinaison d'erreurs quadratiques et des poids et détermine ensuite la combinaison correcte afin de produire un réseau avec une capacité de généralisation meilleure. Ce processus est appelé la régularisation Bayésienne. Il réduit la difficulté de déterminer l'architecture de réseau optimum.

II.4.5.2. Critère d'arrêt pour l'apprentissage

L'une des méthodes qui permet de contrôler la capacité d'un réseau de neurones consiste à arrêter l'apprentissage "à temps". De ce fait différents critères permettant de décider quand stopper l'algorithme d'apprentissage ont été développés :

- Quand l'erreur d'apprentissage a atteint un seuil fixé;
- Après un nombre fixé de cycles d'apprentissage;
- Quand une estimation de l'erreur et de généralisation est minimum.

II.4.6. Procédure de modélisation neuronale

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 multicouches (MLP). 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.4.6.1. Sélection des données

Lorsqu'on conçoit un modèle neuronal, 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, [Dreyfus, 2005].

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.

II.4.6.2. 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 donnée reconstituée doit être représentative des données susceptibles d'intervenir en phase d'utilisation du système neuronal.

II.4.6.3. 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.4.6.4. Normalisation et traitement 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. Cependant, les deux méthodes suivantes sont les plus utilisées ;

- La normalisation qui a pour but d'avoir données ayant une distribution normale de moyenne égale à zéro et un écart type égal à un. Cette méthode a l'équation de normalisation suivante:

$$Valeur\ normalisée = \frac{Valeur\ \ réelle - Valeur\ moyenne}{Ecart\ Type} \qquad \text{(II.5)}$$

  La fonction ***prestd*** de MATLAB® réalise la normalisation selon l'équation (II.5), la fonction inverse de dénormalisation est réalisée par la fonction ***poststd***.
- La normalisation qui a pour but d'avoir les données normalisées comprises entre -1 et +1 en utilisant l'équation de normalisation suivante:

$$Valeur\ normalisée = \frac{2(Valeur\ \ réelle - Minimum)}{Maximum - Minimum} - 1 \qquad \text{(II.6)}$$

  La fonction ***premnmx*** de MATLAB® réalise la normalisation selon l'équation (II.6), la fonction inverse de dénormalisation est réalisée par la fonction ***postmnmx***.

II.4.6.5. Choix du type de réseau de neurones

Le réseau de type MLP est le réseau le plus utilisé pour l'approximation des fonctions des données statiques. Il est d'ailleurs le réseau par défaut pour les interfaces graphiques de réseaux de neurones, ***nntool*** et ***nftool***, de MATLAB®. Par conséquent, le réseau de type MLP, fonction ***newff*** de MATLAB®, a été notre choix.

II.4.6.6. Choix du nombre de neurones dans la couche cachée

Il n'y a pas de règles générales qui fixent le nombre de couches cachées d'un MLP pour un problème donné. Par conséquent la stratégie d'essai et erreur (trial and error) est la plus raisonnable [77]. Le critère d'arrêt est la capacité de généralisation (phase de validation du réseau) et une fonction de performance donnée [78].

II.4.6.7. 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.4.6.8. 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 [79].

II.4.6.9. 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.5. Adaptive Neuro-Fuzzy Inference System (ANFIS)

Adaptive Neuro-Fuzzy Inference System (ANFIS) est une combinaison de systèmes flous et de réseaux neuronaux. Dans ANFIS, un réseau de neurones est utilisé pour décider des paramètres du système flou. La nécessité d'une optimisation manuelle des paramètres du système flou est principalement supprimée par l'ANFIS.

Les paramètres du système sont ajustés automatiquement avec l'utilisation du réseau neuronal. Par exemple ajuster les limites des fonctions d'appartenance, conduisant à des performances améliorées sans intervention de l'opérateur.

Le système neuro-flou avec la capacité d'apprentissage des réseaux de neurones et avec les avantages du système flou basé sur des règles peut améliorer considérablement les performances et peut fournir un mécanisme pour incorporer des observations passées dans le processus de classification ou de régression.

Dans les réseaux neuronaux, la formation construit essentiellement le système. Cependant, en utilisant un schéma neuro-flou, le système est construit par des définitions de logique floue et est ensuite affiné à l'aide d'algorithmes d'apprentissage de réseau neuronal.

Les réseaux de neurones et la logique floue sont des estimateurs universels. Ils peuvent estimer n'importe quelle fonction avec n'importe quelle précision prescrite si suffisamment de neurones cachés et de règles floues sont disponibles. Les algorithmes de rétro-propagation et de descente de gradient sont utilisés pour ajuster les paramètres des fonctions d'appartenance (ensembles flous) et les poids de défuzzification (réseaux de neurones) pour les réseaux de neurones flous. Pour mettre à jour les paramètres, l'ANFIS applique ces deux techniques.

L'ANFIS est un FIS (Fuzzy Inference System) implémenté dans le cadre d'un réseau de neurones flou adaptatif. Dans l'ANFIS, la représentation des connaissances d'un FIS est combinée au pouvoir d'apprentissage des RNA.

L'objectif fondamental de l'ANFIS est d'intégrer les fonctionnalités des systèmes flous et des réseaux de neurones. L'avantage du flou est que les connaissances préalables sont dénotées dans un ensemble de contraintes pour diminuer l'espace de recherche d'optimisation. L'adaptation de la rétro-propagation à un réseau structuré de manière à automatiser le réglage paramétrique de contrôle flou est utilisée à partir du réseau neuronal.

ANFIS utilise un algorithme de descente de gradient pour affiner les paramètres de prémisse qui définissent les fonctions d'appartenance. Pour les paramètres conséquents qui décrivent les coefficients de chaque équation, la méthode des moindres carrés est utilisée par l'ANFIS pour les identifier.

Cette approche combine l'algorithme de descente de gradient et la méthode des moindres carrés, ainsi appelée méthode d'apprentissage hybride. Pour obtenir une bonne généralisation des données invisibles, la taille de l'ensemble de données d'apprentissage doit être au moins équivalente au nombre de paramètres ajustables dans ANFIS.

Les neurones de l'ANFIS ont différentes structures:

- La fonction Membership est décrite par des trapèzes souples paramétrés (Generalized Bell Functions).
- Les fonctions sont des régressions linéaires et une multiplication avec w, c'est-à-dire des poids normalisés ω et une sortie (somme algébrique).
- La normalisation se fait par somme et division arithmétique.
- Les règles sont différentiables T-norme généralement produit.

II.5.1. Fuzzy systems

Le terme "système" est généralement compris comme un ensemble de composants en interaction avec une structure bien définie et organisés comme un tout complexe qui peut être distingué de l'environnement "externe". Un système communique avec l'environnement par ce que l'on appelle des entrées et des sorties.

Les systèmes flous sont des structures basées sur des techniques floues orientées vers le traitement de l'information, où l'utilisation de la théorie des ensembles classique et de la logique binaire est impossible ou difficile. Dans la littérature, des termes tels que système flou, modèle flou, système basé sur des règles floues, contrôleur flou ou mémoire associative floue sont utilisés de manière interchangeable selon le type d'application. Leur principale caractéristique est la représentation symbolique des connaissances sous forme de règles conditionnelles floues (si-alors).

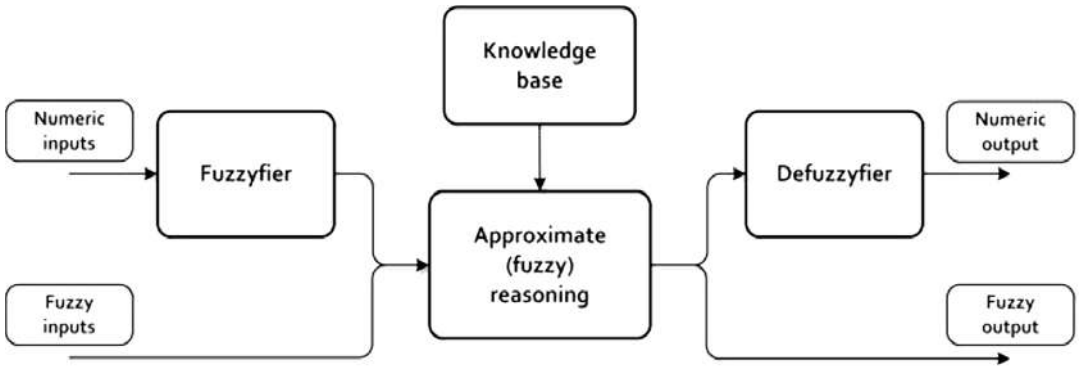


Figure II.11. La structure typique d'un système flou.

La structure typique d'un système flou tel qu'illustré à la Figure (II.11) se compose de quatre blocs fonctionnels: le fuzzifier, le moteur d'inférence floue, la base de connaissances et le défuzzifier. Les valeurs linguistiques (définies par des ensembles flous) et les données nettes (numériques) peuvent être utilisées comme entrées pour un système flou. Si des données nettes sont appliquées, le processus d'inférence est précédé par la fuzzification, qui attribue l'ensemble flou approprié à l'entrée non floue.

Les valeurs des variables d'entrée sont mappées en valeurs linguistiques de la variable de sortie au moyen de la méthode appropriée de raisonnement approximatif (moteur d'inférence) en utilisant la connaissance d'expert, qui est représentée comme un ensemble de règles conditionnelles floues (base de connaissances). En plus des valeurs linguistiques, les données numériques peuvent être requises comme sortie système floue. Dans de tels cas, des méthodes de défuzzification sont utilisées, qui attribuent les données nettes représentatives à l'ensemble flou de sortie résultant.

Les applications pratiques des systèmes flous comprennent des problèmes pour lesquels la description mathématique complète n'est pas disponible, ou pour lesquels l'utilisation du modèle précis (non flou) n'est pas économique ou très peu pratique. La capacité de traiter des informations inexactes fait d'un système flou un excellent outil, par exemple, pour les processus de contrôle, l'identification du système, l'aide à la décision et le traitement du signal et de l'image. Plus de détails mathématiques peuvent être trouvés dans. Néanmoins, le principal problème de la logique floue est qu'il n'y a pas de procédure systématique pour la conception d'un contrôleur flou. D'un autre côté, un réseau de neurones a la capacité d'apprendre de l'environnement (paires d'entrée-sortie), d'auto-organiser sa structure et de s'y adapter de manière interactive. C'est une raison simple derrière l'existence de l'ANFIS. Un système qui est capable de méthodologies auto-organiser la structure du réseau et d'adapter les paramètres du système flou pour prédire différents types de propriétés.

II.5.2. Architecture et algorithme

ANFIS est un type de réseau neuronal axé sur le système d'inférence floue Takagi-Sugeno. Il s'agit d'une technique d'IA utilisant actuellement des processus hydrologiques. ANFIS est d'abord introduit par Jang (1993) et est basé sur le modèle flou de Sugeno du premier ordre. L'ANFIS utilise couramment soit la rétropropagation, soit une combinaison de rétropropagation et d'estimation des moindres carrés pour la prédiction d'un paramètre de fonctions d'appartenance (MF). Le but le plus important de l'intégration de réseaux de

neurones avec des systèmes flous est d'exercer la capacité d'apprentissage du réseau de neurones, tandis que la capacité d'apprentissage est un avantage en termes de système flou. D'un autre côté, il y a plus d'avantages pour un système combiné sous l'aspect d'un réseau neuronal.

Le système d'inférence floue de type Takagi-Sugeno est utilisé dans ANFIS où la sortie de chaque règle peut être un terme constant ou peut être une combinaison linéaire de variables d'entrée ajoutées à un terme constant. La moyenne pondérée de la sortie de chaque règle est la sortie finale. Un réseau adaptatif est un réseau à anticipation multicouche. Ce réseau a cinq couches. Les couches d'entrée et de sortie de ce réseau représentent respectivement les descripteurs et la réponse.

La couche cachée contient les fonctions de nœud comme les MF et les règles, dans lesquelles chaque nœud remplit une fonction particulière. Dans ce système, "les nœuds carrés" (nœuds adaptatifs) ont des paramètres; Alors que "les nœuds du cercle" (nœuds fixes) n'ont pas de paramètres (voir Figure II.13.). Par exemple, dans un modèle flou de Sugeno avec deux entrées, x et y, et une sortie, un ensemble de règles communes avec deux règles floues "si-alors" est le suivant:

Règle I: si $x=A_1$ *et* $y=B_1$, alors

$$f_1 = p_1x + q_1y + r_1 \qquad \text{(II.7)}$$

Règle II: si $x=A_2$ *et* $y=B_2$, alors

$$f_2 = p_2x + q_2y + r_2 \qquad \text{(II.8)}$$

Avec:

f	Un paramètre de sortie (réponse),
p, q, r	Sont des paramètres linéaires
A, B	Sont des paramètres non linéaires.

Ensuite, le réseau de neurones adaptatifs et son fonctionnement identique au FIS sont présentés dans la figure (II.12) et figure (II.13) respectivement.

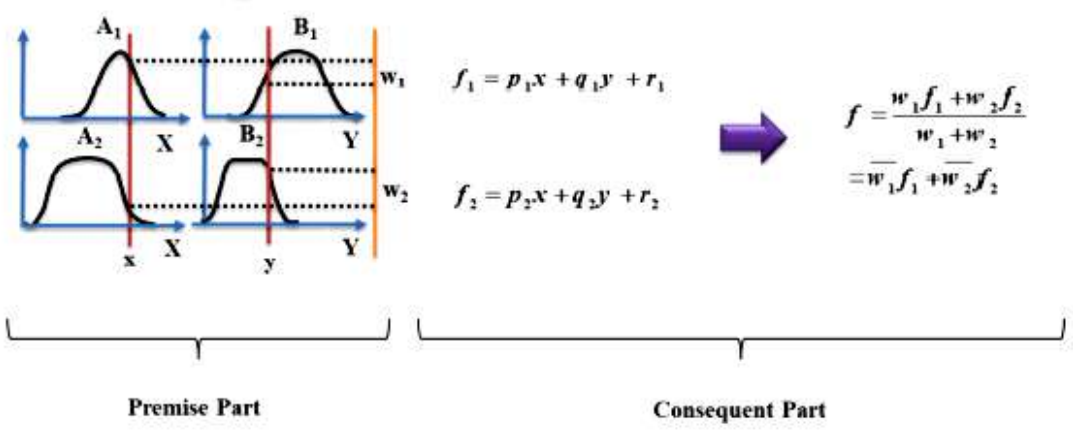


Figure II.12. Structure schématique du modèle flou TSK.

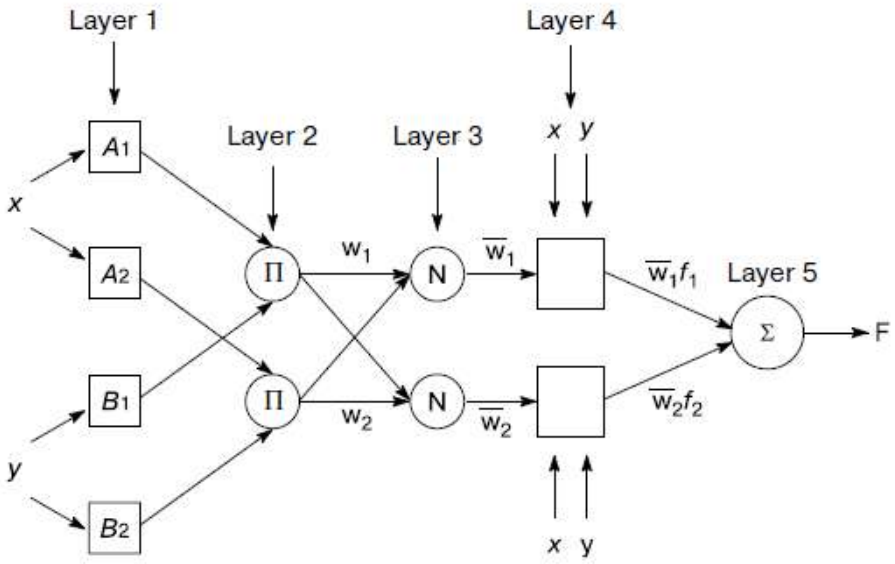


Figure II.13. L'architecture ANFIS avec deux paramètres d'entrée x et y.

Le mécanisme du modèle de Sugeno et l'ANFIS équivalent correspondant sont illustrés à la figure (II.13).

II.6. Machines à Vecteurs de Support (SVM)

Les Machines à Vecteurs de Support (Support Vector machine SVM) est actuellement un sujet brûlant dans le domaine de l'apprentissage statistique. À la fin des années 1990, les approches traditionnelles des réseaux de neurones ont souffert de graves difficultés de généralisation et de production de modèles qui sont la principale raison de la fondation des machines à vecteurs de support. Il a été développé en 1995 par Vladimir Vapnik et a rapidement gagné en popularité en raison de nombreuses fonctionnalités attrayantes.

Dans l'apprentissage statistique, les SVM ont supervisé des méthodes d'apprentissage avec des algorithmes d'apprentissage associés qui analysent l'ensemble de données. Il a d'abord été introduit comme méthode de résolution des problèmes de classification (SVC). Cependant, en raison de nombreuses caractéristiques intéressantes, il a été récemment étendu au domaine de l'analyse de régression. Connu sous le nom de régression à vecteurs de support (SVR) est la version de régression de SVM.

Les SVM appartiennent à une famille de classificateurs linéaires généralisés. Dans un autre terme, SVM est un outil de prédiction de classification et de régression qui utilise la théorie ML pour maximiser la précision prédictive tout en évitant automatiquement le sur-ajustement des données.

Les SVM peuvent être définis comme des systèmes qui utilisent l'espace d'hypothèses de fonctions linéaires dans un espace de caractéristiques de grande dimension, formés avec un algorithme d'apprentissage de la théorie d'optimisation qui implémente un biais d'apprentissage dérivé de la théorie de l'apprentissage statistique. Il est également utilisé pour de nombreuses applications, telles que l'analyse de l'écriture manuscrite, l'analyse des visages, etc., en particulier pour la classification de motifs et les applications basées sur la régression.

L'avantage de cette méthode est plus pratique que les autres modèles en phase de formation et a également une efficacité élevée. Cet algorithme désigne la meilleure ligne de séparation pour classer les données avec une plus grande marge de sécurité. Dans cette méthode, les vecteurs sont choisis comme critère de sélection et ces vecteurs sont appliqués aux meilleures limites et catégories de données. Ces vecteurs sont appelés vecteurs de support.

En résumé, nous identifions au moins trois raisons du succès des SVM: leur capacité à bien apprendre avec seulement un très petit nombre de paramètres libres, leur robustesse face à plusieurs types de violations de modèles et de valeurs aberrantes, et enfin et surtout leur efficacité de calcul par rapport à plusieurs autres méthodes.

II.6.1. Idée de base de la théorie de l'apprentissage statistique

La théorie de l'apprentissage statistique est un cadre pour les domaines ML liés aux domaines de la statistique et de l'analyse fonctionnelle. Le but de l'apprentissage statistique est de trouver la fonction la plus appropriée, représentant f les informations systématiques de la réponse de ses prédicteurs, qui seront utilisées pour la prédiction et l'inférence.

Pour estimer cette fonction inconnue, l'approche utilisée dans l'apprentissage statistique est appelée données d'apprentissage, dont le nom vient de l'usage qui forme notre méthode à l'estimation de f. Chaque point des données d'entraînement est une paire d'entrée-sortie, où l'entrée mappe une sortie. Le problème de la tendance statistique consiste à déduire la fonction qui mappe entre l'entrée et la sortie, de sorte que la fonction apprise puisse être utilisée pour prédire la sortie de l'entrée future.

La théorie de l'apprentissage statistique est largement utilisée dans la modélisation de régression et la classification des données. Dans le problème de régression, les réponses sont des valeurs quantitatives qui prennent une plage continue de valeurs, tandis que les réponses au problème de classification sont des valeurs qualitatives qui sont des éléments d'un

ensemble discret d'étiquettes. Dans le domaine de l'apprentissage statistique, l'approche pour trouver la fonction f est différente sur de nombreux aspects, par exemple le choix de la fonction de perte, cependant, ils partagent également certaines similitudes.

II.6.2. Apprentissage et généralisation

Les premiers algorithmes de ML visaient à apprendre les représentations de fonctions simples. Par conséquent, l'objectif de l'apprentissage était de produire une hypothèse qui effectuait la classification (ou la régression) correcte des données d'entraînement et les algorithmes d'apprentissage précoce ont été conçus pour trouver un ajustement aussi précis aux données.

La capacité d'une hypothèse à classer correctement les données, et non dans l'ensemble d'apprentissage, est appelée sa généralisation. SVM fonctionne mieux en termes de non sur-généralisation lorsque les réseaux de neurones pourraient finir par sur-généraliser facilement.

Une autre chose à observer est de trouver le meilleur compromis entre la complexité du modèle et l'erreur d'apprentissage minimale, figure (II.14).

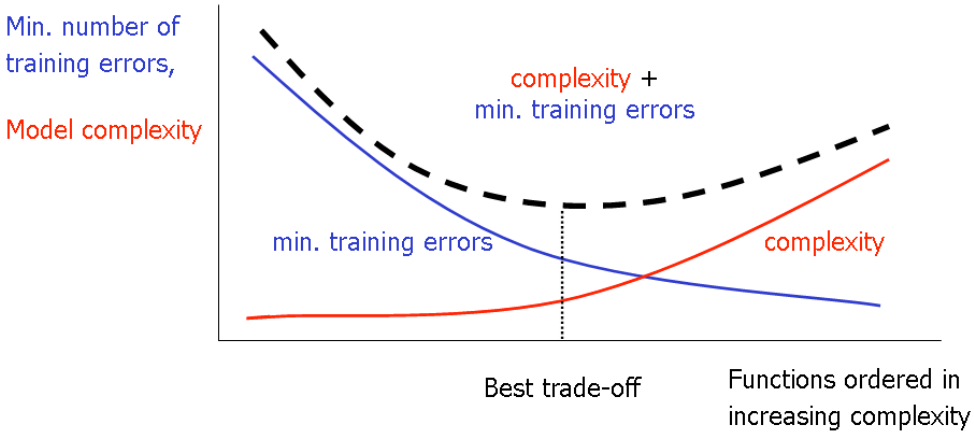


Figure II.14. Meilleur compromis entre la complexité du modèle et l'erreur l'apprentissage minimale.

II.6.3. Régression à vecteurs de Support (SVR)

Régression à Vecteurs de Support (SVR), la version étendue de SVM, est devenue un outil de calcul primordial en raison de ses applications efficaces dans les problèmes de régression et de prédiction. SVR a été développé pour la prédiction des fonctions de régression, et il est basé sur la conversion non linéaire implicite des données en un espace de caractéristiques de plus grande dimension.

Les principes fondamentaux de la SVR sont brièvement rapportés ici en considérant une fonction de régression *F* qui est estimée sur la base des données d'apprentissage sous la forme de:

$$F = \{(x_i, y_i) \qquad i = 1,2,3, \dots, n\}$$

Avec: x_i , y_i Les ensembles d'entrée et de sortie, respectivement,
n Le numéro global de l'ensemble de données.

SVR prédit les valeurs cibles en utilisant la fonction de régression comme suit:

$$F = w^T \varphi(x) + b \qquad \text{(II.9)}$$

Avec: F La sortie prévue,
w Le vecteur de poids,
b Le biais,
x Le vecteur d'entrée à haute dimension.

La planéité dans l'équation II.9 signifie que l'on cherche un petit *w*. Pour cela, il faut minimiser la norme euclidienne, i.e. $\|w\|^2$. Les coefficients *w* et *b* ont été calculés en minimisant la fonction de risque, comme indiqué ci-dessous.

$$R(F) = \frac{1}{2}\|w\|^2 + C\frac{1}{n}\sum L_\varepsilon (y_i, F_i) \qquad \text{(II.10)}$$

Où, $L_\varepsilon(y_i, F_i) = \begin{cases} 0 & si \quad |F_i - y_i| \le \varepsilon \\ |F_i - y_i| - \varepsilon & si\ non \end{cases}$

Le paramètre de pénalité C est utilisé pour identifier le compromis entre la complexité des fonctions et les pertes "représente l'équilibre entre la complexité du modèle et l'erreur approximative". Comme on peut le voir sur la figure (II.15), le paramètre ε est la fonction de perte qui représente la plage entre les valeurs réelles et la fonction de régression.

Cette plage peut être vue comme un tube autour de la fonction de régression. Tous les points situés à l'extérieur du tube sont considérés comme des erreurs d'entraînement. $L_\varepsilon(y_i,F_i)$ est appelée la fonction de perte insensible ε.

Équation II.10. peut être remplacé par la forme contrainte suivante.

$$minimize \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \qquad \text{(II.11)}$$

$$Subject\ to \begin{cases} w.\varphi(x) + b - y_i \le \varepsilon + \xi_i \\ y_i - (w.\varphi(x) + b) \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0, \qquad i = 1,2,3, \dots, n \end{cases}$$

Avec: ξ_i & ξ_i^* Deux variables d'écart positives, qui supposent des valeurs non nulles à l'extérieur du tube ε et nulles à l'intérieur, figure (II.15).

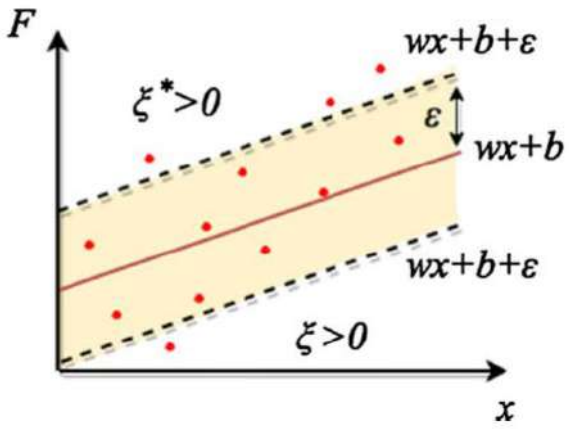


Figure II.15. Régression avec le tube insensible ε.

Le problème d'optimisation dans l'équation (II.11) peut être résolu plus facilement dans sa double formulation. Par conséquent, une technique de dualisation standard utilisant des multiplicateurs lagrangiens a été employée. En utilisant les multiplicateurs lagrangiens, ce problème peut être écrit dans la formulation duale comme suit:

$$Maximize\ \{-\tfrac{1}{2}\sum_{i,j=1}^{n}(\alpha_i,\alpha_i^*)(\alpha_j,\alpha_j^*)\left(\varphi(x_i).\varphi(x_j)\right)-\varepsilon\sum_{i=1}^{n}(\alpha_i,\alpha_i^*)+\sum_{i=1}^{n}y_i(\alpha_i,\alpha_i^*)\} \qquad \text{(II.12)}$$

$$Subject\ to\ \begin{cases}\sum_{i=1}^{n}(\alpha_i,\alpha_i^*)=0\\ 0\le\alpha_i\le C\,, \quad i=1,2,3,\dots,n\\ 0\le\alpha_i^*\le C\,, \quad i=1,2,3,\dots,n\end{cases}$$

Avec, α_i^* Multiplicateurs lagrangiens non linéaires.

La fonction SVR peut être obtenue en résolvant le problème de double maximisation dans l'équation (II.12) comme suit:

$$F(x,\alpha_i,\alpha_i^*)=\sum_{i=1}^{n}(\alpha_i,\alpha_i^*)\left(\varphi(x_i).\varphi(x_j)\right)+b \qquad \text{(II.13)}$$

Le vecteur intérieur-produit $((x_i)\cdot(x_j))$ représente la fonction de mappage de l'espace d'entrée à l'espace de caractéristiques, et il peut être remplacé par une fonction noyau $k\,(x_i, x_j)$. Par conséquent, équation (II.13) devient:

$$F(x,\alpha_i,\alpha_i^*)=\sum_{i=1}^{n}(\alpha_i-\alpha_i^*)K(x_i,x_j)+b \qquad \text{(II.14)}$$

II.6.4. Les fonctions de noyau

L'idée de base avec les SVM non linéaires (SVC, SVR) est de mapper les données d'entraînement dans un espace de caractéristiques de plus grande dimension via un mappage φ (x) et de construire un hyper-plan de séparation avec une marge maximale dans l'espace d'entrée. Comme représenté sur la figure (II.16) la fonction de décision linéaire dans l'espace

des caractéristiques correspond à une frontière de décision non linéaire dans l'espace d'entrée d'origine. En utilisant une fonction noyau, $K(x_i, x_j)=(\varphi(x_i)\cdot\varphi(x_j))$, il est possible de calculer l'hyper-plan de séparation sans effectuer explicitement le mappage dans l'espace des fonctionnalités. Pour s'assurer qu'une fonction du noyau correspond réellement à un certain espace de fonctionnalités, elle doit être symétrique, c'est-à-dire, $(x_i, x_j)=(x_j, x_i)$.

En règle générale, les noyaux doivent satisfaire le théorème de Mercer, qui stipule que la matrice du noyau, doit être semi-définie positive, c'est-à-dire qu'elle n'a pas de valeurs propres non négatives.

$$K = (K(x_i, x_j))_{i,j=1}^{n}$$

Les choix typiques pour les fonctions de noyau sont:

- **Linéaire**

$$K(x_i, x_j) = \varphi(x_i).\varphi(x_j)$$

- **Polynomiale**

$$K(x_i, x_j) = (xy + 1)^P$$

- **Fonction de Base Radiale (RBF)**

$$K(x_i, x_j) = e^{-\|x-y\|^2/2\sigma^2}$$

- **Sigmoid**

$$K(x_i, x_j) = \tanh(kxy + \delta)$$

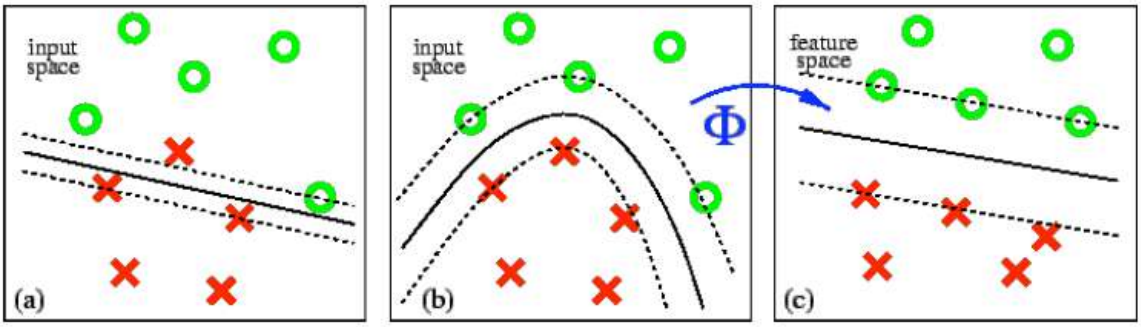


Figure II.16. Effet de la transformation du noyau.

Les données ne sont pas linéairement séparables dans l'espace d'entrée (a), et les données ne peuvent être séparées que par une surface non linéaire (b). Cette surface non linéaire dans l'espace d'entrée correspond à une surface linéaire dans un espace d'entités mappé par noyau. Chaque noyau correspond à un certain espace d'entités et comme aucun mappage explicite à cet espace d'entités ne se produit, des séparateurs linéaires optimaux peuvent être trouvés efficacement dans les espaces d'entités avec des millions de dimensions.

II.7. Conclusion

L'objectif de ce chapitre est de donner une vue générale de quelques concepts de base liés au thème et aux algorithmes sélectionnés pour modéliser les propriétés critiques et la solubilité de solutés solides dans le dioxyde de carbone supercritique. Notez que, l'existence de plusieurs autres types d'algorithmes non mentionnés et dans chaque algorithme, différents autres sous-types non mentionnés. D'autre part, en fonction de l'utilisation de chacun, il peut être appelé ou mieux se référer différemment. Ce dernier donne à ce domaine, l'informatique, sa puissance et sa diversité.

Comme décrit ci-dessus, chaque outil de programmation, technique, modèle ... à sa propre puissance qui ne se voit que s'il a été correctement utilisé par rapport à sa limite de fonctionnement, et en fonction des données utilisées chaque modèle peut montrer des résultats différents. C'est-à-dire quand il s'agit de quelle méthode est préférable de choisir pour représenter une propriété ou un phénomène, il n'y a pas de bonne réponse.

Cependant, pour un ensemble de données fixe, "essayer et voir" est l'une des méthodes efficaces pour comparer ces algorithmes et donner une décision critique sur l'algorithme qui peut représenter les résultats des expériences bien mieux que les autres.

Chapitre III.

Contribution à la modélisation des propriétés critiques par l'intelligence computationnelle.

Chapitre III
Contribution à la modélisation des propriétés critiques par l'intelligence computationnelle.

III.1.	Introduction	69
III.2.	Matériaux et méthodes	70
III.3.	Résultats et discussion	71
III.4.	Conclusion	84

III.1. Introduction

Les propriétés physiques des composés chimiques, telles que les propriétés critiques et le facteur acentrique, sont importants pour le scientifique pour concevoir et modéliser des processus chimiques tels que l'extraction supercritique et leur importance dans l'application de différentes équations d'états. Étant donné que le calcul expérimental de ces propriétés est coûteux, prend du temps et est parfois un peu impossible étant donné que certains composés plus gros ou de poids moléculaire plus élevé peuvent se dégrader chimiquement avant d'atteindre des conditions critiques [5]. Par conséquent, différentes méthodes de prédiction ont été publiées dans la littérature pour calculer ces propriétés critiques lorsque les valeurs expérimentales ne sont pas disponibles. Jusqu'à ce jour, deux approches ont été proposées pour estimer les propriétés critiques qui sont : la méthode de contribution de groupe et la relation quantitative propriété/structure (QSPR). Étant donné que les modèles basés sur les contributions de groupe souffrent de certaines limites [80]. Une autre approche basée sur QSPR a été utilisée pour estimer les propriétés physiques uniquement à partir des descripteurs moléculaires qui sont calculés en appliquant certains algorithmes mathématiques sur la structure moléculaire des composants [80–82]. Plusieurs approches ont été utilisées pour estimer ces propriétés, à savoir la contribution de groupe à partir des structures moléculaires [83,84], régression multilinéaire basée sur les approches QSPR à partir de descripteurs [5], en plus de différentes techniques de calcul comme les réseaux de neurones artificiels (RNA), le système d'inférence neuro-floue adaptatif (ANFIS) et la machine a vecteurs de support (SVM) ont été utilisés avec succès pour estimer ces propriétés de différents composants à partir de descripteurs [5,85]. Afin de le rendre claire, on a organisé ce chapitre en trois parties principales :

Dans la première partie nous expliquerons tous les matériels et les méthodes traitées, notamment les logiciels informatiques utilisées, la description de la base des données ; leur collection et leur préparation, le calcul des descripteurs moléculaire, et la procédure suivez

pour sélectionner juste les descripteurs pertinents, car elle revêt une grande importance dans la détermination de la qualité des modèles qui serons obtenus. Également nous expliquerons brièvement la méthodologie utilisée pour le développement de ces modèles, et la manière de choix de leurs paramètres.

Nous montrerons dans le seconde partie tous les résultats obtenus durant et après le développement des modèles de type RNA-QSPR, nous donnera l'évolution des nombres des descripteurs au cours le prétraitement, les architectures des meilleures modèles obtenus, nous ferons aussi des analyse tel que le calcul des erreurs et l'analyse de sensibilité qui mesure la contribution des variables d'entrée sur les propriétés à modéliser. Finalement, nous interprétant et nous discuterons ces résultats.

Dans la dernière partie de ce chapitre, nous présenterons une conclusion qui résume brièvement le travail et se tient sur les étapes les plus importantes dans cette contribution.

III.2. Matériels et méthodes

Dans ce travail, l'ensemble de données de 6700 composés chimiques ont été collectés à partir d'articles et de livres publiés antérieurement dans la littérature [5,86]. Les descripteurs ont été calculés à l'aide du logiciel alvaDesc basé sur le SMILE (Simplified Molecular-Input Line-Entry System) de chaque composé, qui calcule 5305 types de descripteurs (c.-à-d. Descripteurs 0D, 1D, 2D et 3D) [87].

La présélection des descripteurs pertinents s'est faite en deux étapes. La première sélection a été réalisée à l'aide du logiciel alvaDesc en réduisant tous les descripteurs comme suit:

- Les valeurs de descripteur constantes et quasi constantes ont été supprimées;
- Toutes les valeurs manquantes ou au moins une valeur manquante ont été supprimées;
- Tout descripteur avec un écart type relatif <0.001 a été supprimé;
- Les descripteurs avec une corrélation de paires supérieure ou égale à ≥ 0.95 ont été supprimés.

La deuxième étape consiste à utiliser la MLR par étapes comme sélection finale pour ne disposer que des descripteurs pertinents. 05 descripteurs sont restés après le criblage final qui sont {Sp: somme des polarisabilités atomiques (mise à l'échelle sur l'atome de carbone), Mp: polarisabilité atomique moyenne (mise à l'échelle sur l'atome de carbone), GD: graphique densité, H% pourcentage d'atomes H, C% : pourcentage d'atomes de C}. Le nombre d'entrées pour chaque propriété était de 5 descripteurs, températures d'ébullition (Tb) et poids moléculaires (Mw).

Les performances de la méthode SVM pour la régression dépendent principalement de plusieurs paramètres et de la fonction noyau. Étant donné que la relation entre les

descripteurs moléculaires sélectionnés et les propriétés sélectionnées est non linéaire et que pour les tâches de régression, une des fonctions de noyau couramment utilisées et en raison de ses bonnes performances générales est la fonction de base radiale gaussienne (RBF) [88].

III.2.1. Description de méthode hybride SVM-DA

Les machines à vecteurs de support (SVM) sont une méthode d'apprentissage supervisé présentée pour la première fois en 1992 par Boser et al. [89], quelques années plus tard, ils ont été introduits par Cortes et al. 1995 [90] pour les cas de classification et en 2013 Vapnik et al. [91] ont été étendus et ajustés aux problèmes de régression pour mapper la relation non linéaire entre les entrées et les sorties afin de construire la fonction de régression dans l'espace des caractéristiques à haute dimension $\Phi(x)$, mapper vers l'espace d'origine et définir la fonction de noyau (Kernel) appropriée $K(x_i, y_i)$, où x_i est la valeur d'entrée, y_i est la valeur de sortie et {i = 1,…, m} m est la taille de l'ensemble de données d'apprentissage $D = (x_i, y_i)$ [92]. La fonction de régression est conçue par $y = W^T.\Phi(x) + b$, où W est le vecteur de poids et b est le biais [93]. Les principaux avantages du modèle SVM sont à savoir, l'absence de minima locaux pendant la phase d'apprentissage, gérer la régression non linéaire, clairsemé, utiliser des fonctions de noyau [94]. Plus de détails sur l'équation d'optimisation du modèle SVM et les fonctions du noyau peuvent être trouvés dans le chapitre II [95–97]. Un nouvel algorithme d'optimisation appelé Dragonfly a été développé par Mirjalili et al. 2016 [98] basé sur le comportement des libellules. Cet algorithme a été couplé à la méthode SVM pour optimiser ses paramètres [99].

III.3. Résultats et discussion

III.3.1. Température critique

Modélisation par (MLR). Les valeurs de température critique prédites à l'aide du modèle MLR développé ont été comparées aux valeurs expérimentales en utilisant deux paramètres à savoir, l'erreur relative absolue moyenne (ERAM) et le coefficient de détermination (R^2). La régression entre les valeurs de la température critique prédites et expérimentales (figure (III.1.a)) montre les performances du modèle MLR en termes de {ERAM=3.376% et R^2=0.9104}. Sur cette figure, la ligne continue rouge indique l'ajustement exact entre les valeurs expérimentales et prédites, tandis que les points en cercles bleu montrent les points corrélés par le modèle proposé par rapport aux données expérimentales. Plus les points sont

proches de la ligne pleine, plus les données corrélées sont précises. L'équation de ce modèle s'écrit comme suit:

$$T_c = -18.93 - 0.0977M_w + 1.47T_b + 3.60S_p + 55.56M_p + 31.90GD + 0.28H\% + 0.57C\% \quad \text{(III.1)}$$

Modélisation par (RNA). L'objectif majeur de la deuxième partie et de savoir quelle est la meilleur architecture du réseau de neurones qui peux être capable de traiter la base de données disponibles.

La conception et l'optimisation du RNA ont étaient réalisées en cinq étapes. En bref :

- Au cours de la première étape, on a testé l'effet de séparation de la base de données sur l'optimisation du réseau.
- Au cours de la deuxième étape, les algorithmes d'apprentissages ont été testé pour trouver celles donnants la meilleure performance.
- Au cours de la troisième étape d'optimisation, on a testé les méthodes de prétraitement des données (méthodes de normalisations).
- La quatrième étape consistait a trouvé le nombre de couches cachées.
- Enfin, la dernière étape consiste à voir l'effet de nombre de neurones dans chaque couche cachée.

Toutes ces étapes ont été développées à l'aide du logiciel commerciale Matlab®. La propriété émergente d'un RNA est sa capacité d'apprentissage. A partir d'un état quelconque, il va, en s'appuyant sur un algorithme d'apprentissage, adapter la valeur des poids synaptiques pour fournir une réponse vraisemblable vis-à-vis du problème traité.

Les paramètres du meilleur modèle RNA ont été déterminés sur la base de la méthode d'essai et d'erreur. Les résultats montrent que, le meilleur réseau de neurones à rétro-propagation de perception multicouche formé avec l'algorithme de Levenberg-Marquardt (RNA-PM) a été obtenu avec l'architecture des neurones {7, 15, 1} dans la couche d'entrée, caché et de sortie respectivement. La fonction de transfert tangente hyperbolique sigmoïde *Tansig* (*f*) et linéaire *Purline* (*g*) a été utilisée respectivement dans la couche cachée et la couche de sortie. La courbe de régression est représentée sur la figure (III.1.b) où la performance du modèle RNA en termes de {ERAM = 2.639% et R^2 = 0.9450} et son expression peut s'écrire comme suit:

$$T_c = g\left(\sum_{j=1}^{15} W_{1,j}^{S}\, f\left(\sum_{i=1}^{7} W_{j,i}^{I}\, X_i + b_{hj}\right) + b_{O1}\right) \quad \text{(III.2)}$$

Où:

X_i	Vecteur d'entrée,
$W_{j,i}^{I}, W_{1,j}^{S}$	Matrice de pondération des couches entrée-cachées et des couches cachées-sortie,
b_{hj}, b_{O1}	Biais des couches cachées et de sortie respectivement.

Modélisation par (SVM). Un autre modèle a été développé basé sur l'approche SVM, le meilleur modèle est basé sur l'optimisation de ses paramètres à savoir, capacité (BoxConstraint) (C), le paramètre de largeur de noyau (Kernel scale) (γ), la quantité de vecteurs de support (QSV) et la fonction de noyau. Le modèle optimisé a été trouvé avec la fonction noyau (RBF) et {C = 10, γ = 1, et QSV = 4008} pour modéliser la température critique. Ce modèle a été trouvé avec {ERAM= 1.224%, et R^2=0.9375}.

Modélisation par (SVN-DA). Le dernier modèle est basé sur l'utilisation de SVM hybride avec un algorithme d'optimisation récent nommé Dragonfly (DA). La comparaison graphique entre les températures critiques prévues et expérimentales a été présentés dans la figure (III.1.d). Sur la base des résultats obtenus, le modèle SVM-DA proposé peut fortement modéliser la température critique en termes de {ERAM=0.7551% et R^2=0.9699} par rapport aux autres modèles développés. Le SVM-DA obtenu a été trouvé avec la fonction noyau (RBF) et {C = 10, γ = 0.5, et QSV = 3293}.

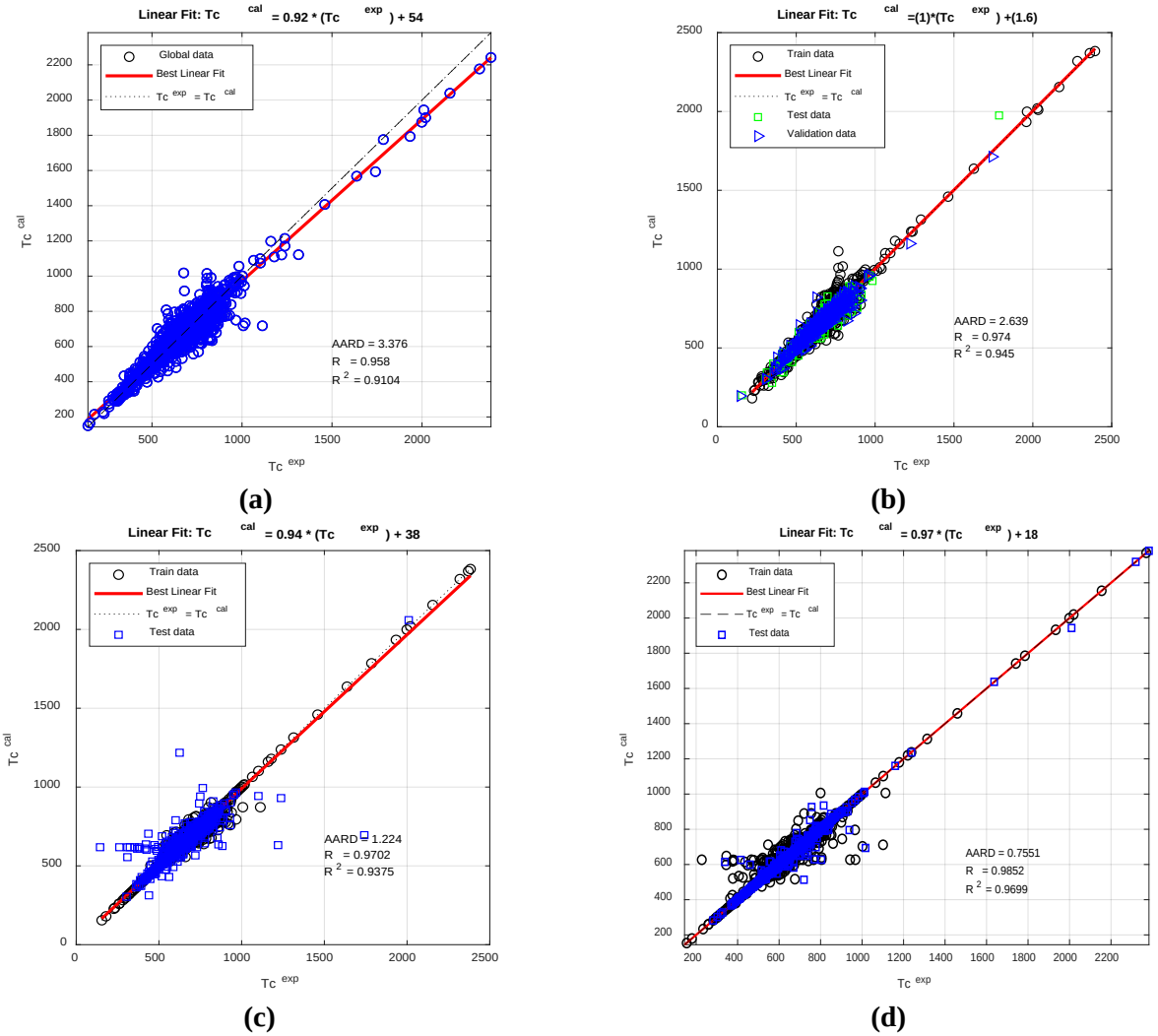


Figure III.1. Comparaison entre la température critique expérimentale et prévue de l'ensemble de la base de données. a: MLR, b: RNA, c: SVM, d: SVM-DA

III.3.2. Pression critique

Afin de modéliser la pression critique, en suivant les mêmes méthodologies de modélisation (MLR, RNA, SVM et SVM-DA) que celle adoptée pour modéliser la température critique:

Modélisation par (MLR). Pour analyser plus précisément les performances du modèle du MLR, le pourcentage de l'erreur relative absolue moyenne (ERAM *%)*, et le coefficient de détermination (R^2) sont calculés, les résultats montrent que le modèle MLR peut prédire les valeurs de pression critique avec {ERAM = 10.47% R^2 = 0.7821}. La régression entre les valeurs calculées par le modèle MLR et les valeurs expérimentales est représentée sur la figure (III.2.a). La pression critique peut être écrite en fonction des entrées sélectionnées via le modèle MLR comme suit:

$$P_c = 18.96 - 0.0751M_w + 0.0734T_b - 1.13S_p + 10.97M_p + 47.62GD - 0.124H\% - 0.1548C\% \quad \text{(III.3)}$$

Modélisation par (RNA). Pour le modèle RNA, en suivant la même méthodologie de conception et d'optimisation du réseau de neurones que celle adoptée pour modéliser la température critique, le réseau adapté après confirmation par une étude évaluative, est un *RNA optimal* de type *MLP* à une couche cachée avec 15 neurones et de fonction de transfert tangente sigmoïde (*Tansig*) et une couche de sortie avec la fonction linéaire (*Purline*). Un algorithme d'apprentissage le mieux adapté après confirmation par une étude évaluative, est celui de Levenberg-Marquardt (*Trainlm*). Les résultats montrent par la comparaison de régression est représentée sur la figure (III.2.b) où la performance du modèle RNA a été trouvée comme {ERAM = 5.897% et de coefficient de détermination R^2 = 0.911}.

Modélisation par (SVM). Le modèle SVM optimal a été obtenu avec la fonction du noyau RBF et {C = 10, γ = 1 et QSV = 4990} pour modéliser la pression critique. La précision de ce modèle était de {ERAM = 2.575% et de coefficient de détermination R^2 = 0.9491}. La représentation graphique de régression entre les valeurs calculées par le modèle SVM et les valeurs expérimentales est rapportée sur la figure (III.2.c).

Modélisation par (SVM-DA). La comparaison graphique présentée dans la figure (III.2.d) entre les valeurs prédites par SVM-DA et les valeurs expérimentales de la pression critique. Les paramètres statistiques calculés montre les performances de ce modèle en termes de {ERAM = 1.962% et R^2 = 0.9673. Ce modèle a été trouvé avec une fonction de noyau (RBF)

et {C = 10, γ = 0.5 et QSV = 4785} et montre une grande capacité de prédiction sur la pression critique par rapport aux autres modèles testés dans cette étude.

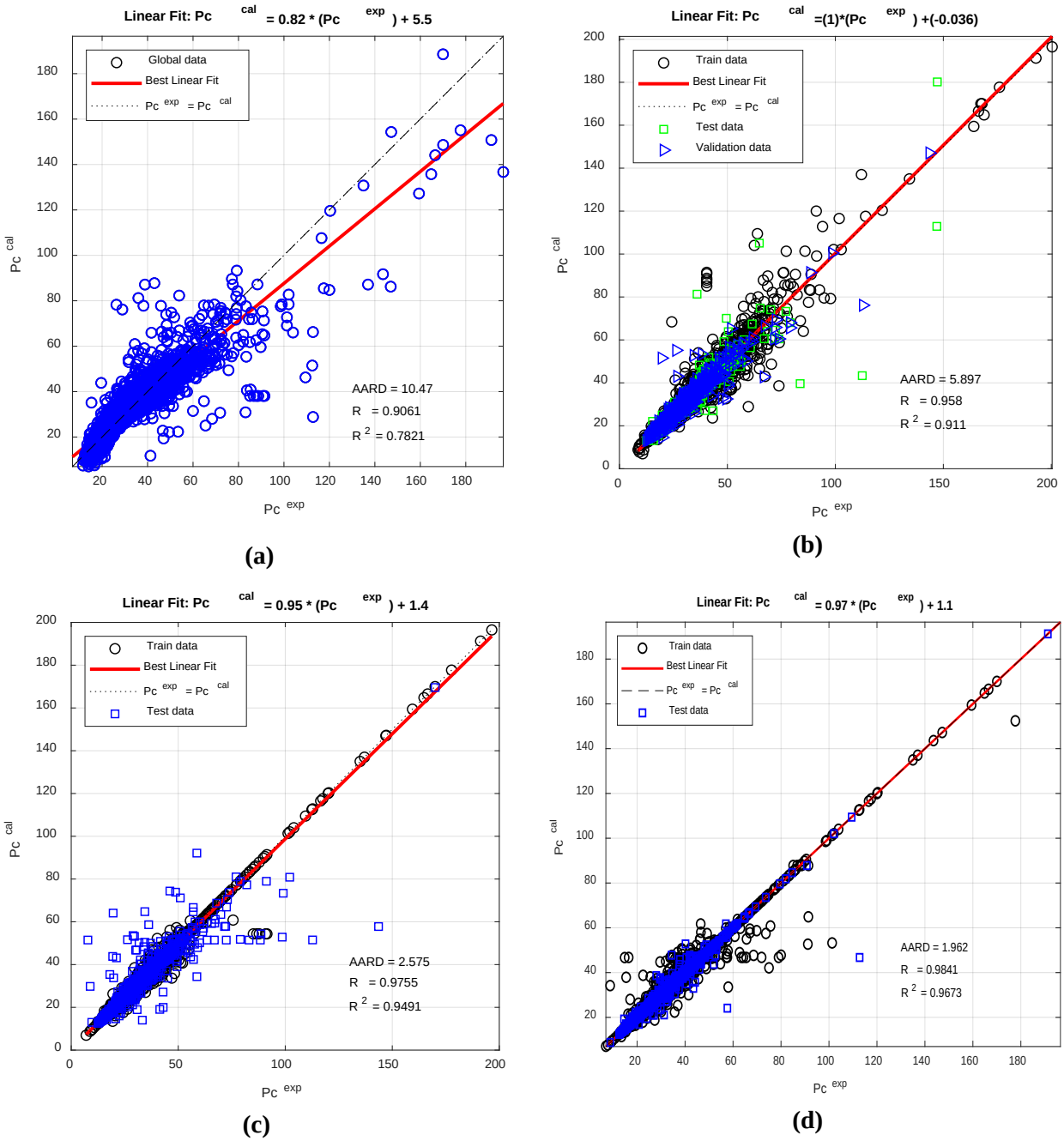


Figure III.2. Comparaison entre la pression critique expérimentale et prévue de l'ensemble de la base de données. a: MLR, b: RNA, c: SVM, d: SVM-DA.

III.3.3. Volume critique

Autre propriété critique à modéliser est le volume critique, en suivant les mêmes méthodologies de modélisation (MLR, RNA, SVM et SVM-DA) que celle adoptée pour modéliser la température critique et la pression critique:

Modélisation par (MLR). L'illustration graphique entre les valeurs expérimentales et les valeurs prédites par le modèle MLR est représenté sur la figure (III.3.a). Les résultats de modélisation du volume critique par la méthode MLR montre les performances du modèle MLR en termes de l'erreur relative absolue moyenne ERAM = 6.627% et de coefficient de détermination R^2 = 0.9339. L'équation de ce modèle s'écrit comme suit:

$$V_c = 127.14 + 0.4065M_w + 0.1988T_b + 28.169S_p - 129.90M_p - 11.84GD - 1.1247H - 1.3638\text{C}\% \quad \text{(III.4)}$$

Modélisation par (RNA). Le modèle RNA optimisé a été trouvé avec l'architecture de {7, 15, 1} neurones dans la couche d'entrée, cachée et de sortie respectivement. La fonction de transfert tangente sigmoïde (*Tansig*) a été utilisée dans la couche cachée et linéaire (*Pureline*) dans la couche de sortie. Les performances du modèle RNA représenté graphiquement par une régression entre les valeurs expérimentales et prédites (figure (III.3.b)). Les résultats montre la précision du modèle RNA optimisé avec {ERAM = 4.534% et de coefficient de détermination R^2 = 0.971.

Modélisation par (SVM). La figure (III.3.c) illustre graphiquement la régression entre les valeurs du volume critique calculés par le modèle SVM optimal et les valeurs expérimentales, cette figure montre la précision du modèle SVM avec l'erreur relative absolue moyenne ERAM = 2.338% et le coefficient de détermination R^2 = 0.9792. Le modèle SVM optimal a été trouvé avec la fonction de noyau (RBF) et {C = 10, γ = 0.5 et QSV = 4817} pour le volume critique.

Modélisation par (SVM-DA). La quatrième méthode développée afin de modéliser avec précision le volume critique est la modélisation par SVM-DA. Le SVM-DA obtenu a été trouvé avec la fonction noyau (RBF) et {C = 10, γ = 0.5 et QSV = 4795}. Les résultats montrent que SVM-DA peut très bien modéliser le volume critique avec {ERAM = 1.929% et de coefficient de détermination R^2 = 0.9856 par rapport aux autres modèles développés. La figure (III.3.d) montre le graphe de régression entre le volume critique expérimental et calculé à l'aide du modèle SVM-DA.

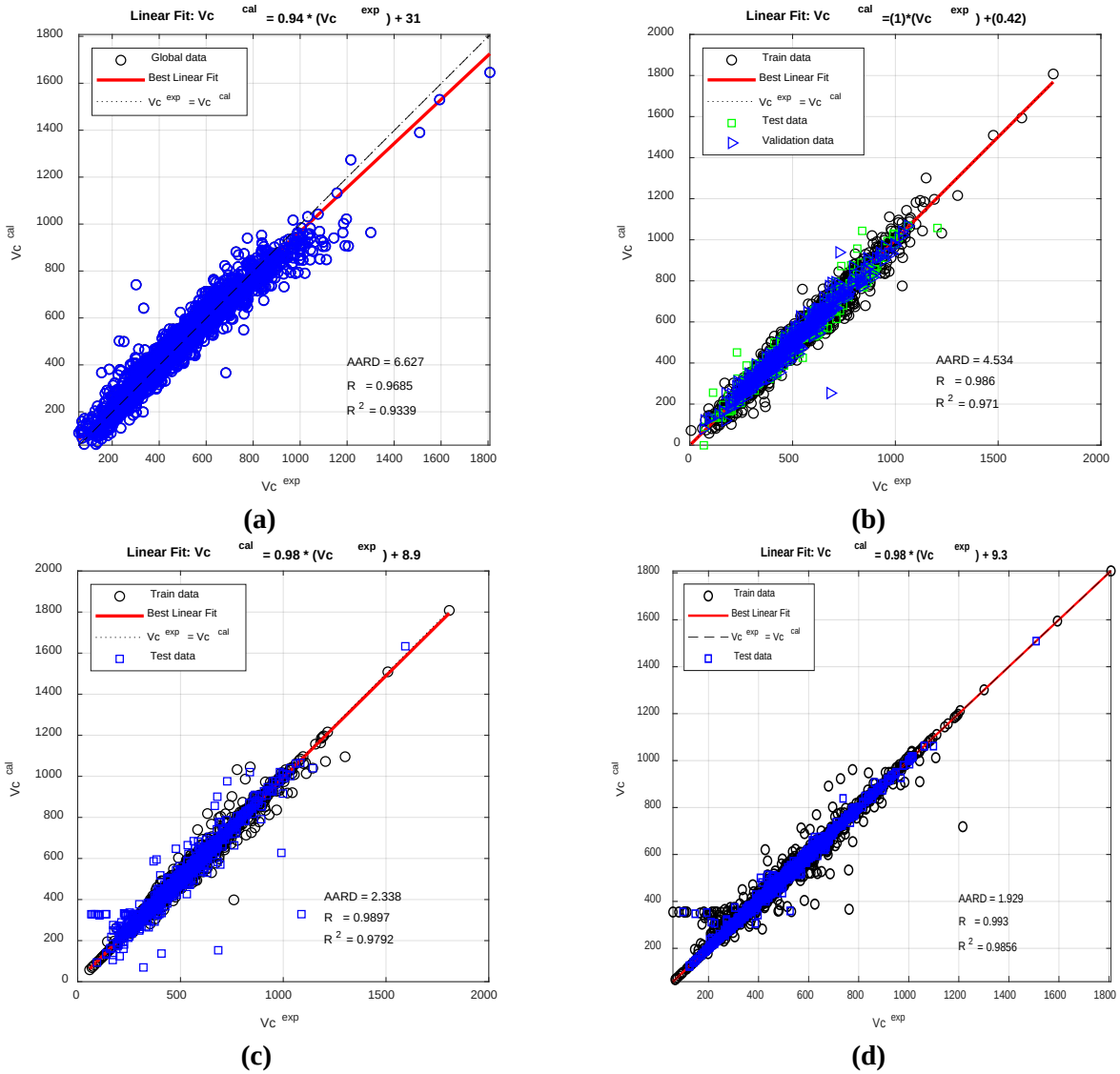


Figure III.3. Comparaison entre le volume critique expérimentale et prévue de l'ensemble de la base de données. a: MLR, b: RNA, c: SVM, d: SVM-DA

III.3.4. Facteur acentrique

Le quatrième paramètre (facteur acentrique) a été modélisé en utilisant les méthodes (MLR, RNA, SVM et SVM-DA) :

Modélisation par (MLR). La performance du modèle MLR pour modélisé le facteur acentrique a été trouvé avec {ERAM = 23.76% et R² = 0.232} comme présenté sur la figure (III.4.a). L'équation de ce modèle s'écrit comme suit:

$$\omega = 0.3369 - 6.86.10^{-5}M_w + 1.25.10^{-3}T_b + 9.858.10^{-3}S_p - 0.333M_p - 0.246GD - 1.448.10^{-3}H\% - 5.069.10^{-3}\% \quad \text{(III.5)}$$

Modélisation par (RNA). Le facteur acentrique a été modélisé en utilisant le modèle RNA optimisé avec la structure de {7, 20, 1} neurones dans la couche d'entrée, cachée et de sortie respectivement. La fonction de transfert tangente sigmoïde (*Tansig*) et linéaire (*Pureline*) a

été utilisée respectivement dans la couche cachée et la couche de sortie. Le modèle RNA optimal a été trouvé avec la performance de {ERAM = 9.733% et de coefficient de détermination R^2 = 0.890 comme présenté sur la figure (III.4.b).

Modélisation par (SVM). Le facteur acentrique a été modélisé en utilisant le modèle SVM optimisé avec la fonction noyau (RBF) et {C = 10, γ = 1 et QSV = 4349}. Le modèle précis a été trouvé avec la performance de {ERAM = 6.140% et de coefficient de détermination R^2 = 0.8725 comme présenté sur la figure (III.4.c).

Modélisation par (SVM-DA). Le facteur acentrique a été modélisé en utilisant le modèle SVM-DA optimisé avec la fonction de noyau (RBF) et {C = 100, γ = 0.5 et QSV = 3032}. Le SVM-DA obtenu a été trouvé avec la performance de {ERAM = 2.173% et de coefficient de détermination R^2 = 0.9766 comme présenté sur la figure (III.4.d).

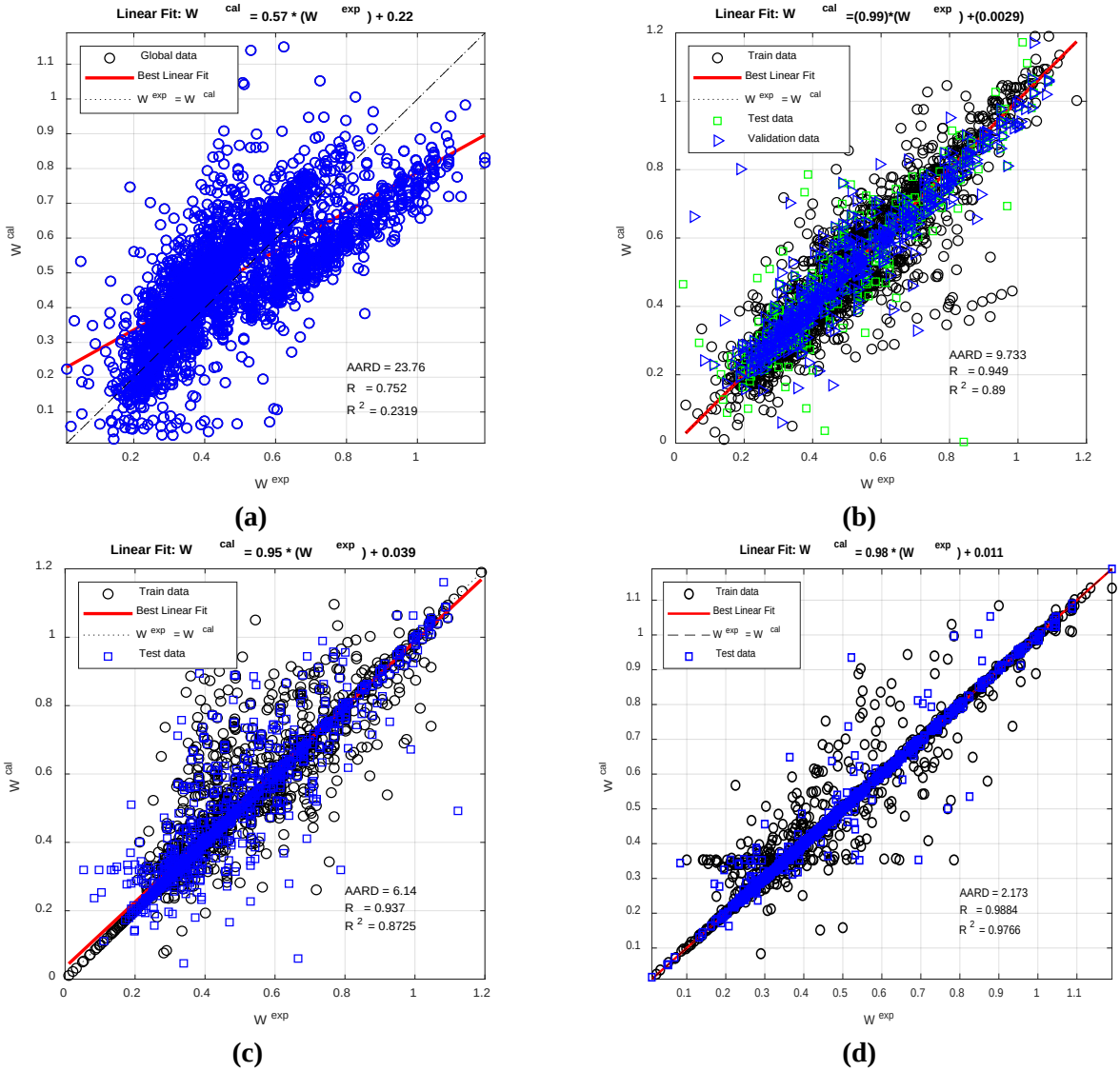


Figure III.4. Comparaison entre le acentrique facteur expérimentale et prévue de l'ensemble de la base de données. a: MLR, b: RNA, c: SVM, d: SVM-DA

Comme on peut le voir sur la figure (III.5), le modèle SVM-DA peut prédire les paramètres sélectionnés avec une grande précision par rapport aux autres modèles avec erreur relative absolue moyenne minimale par rapport les autres modèles développés. Pour cette raison, une analyse de sensibilité, le domaine d'applicabilité et la section d'évaluation statistique et de validation ont été effectuées sur la base des paramètres SVM-DA obtenus.

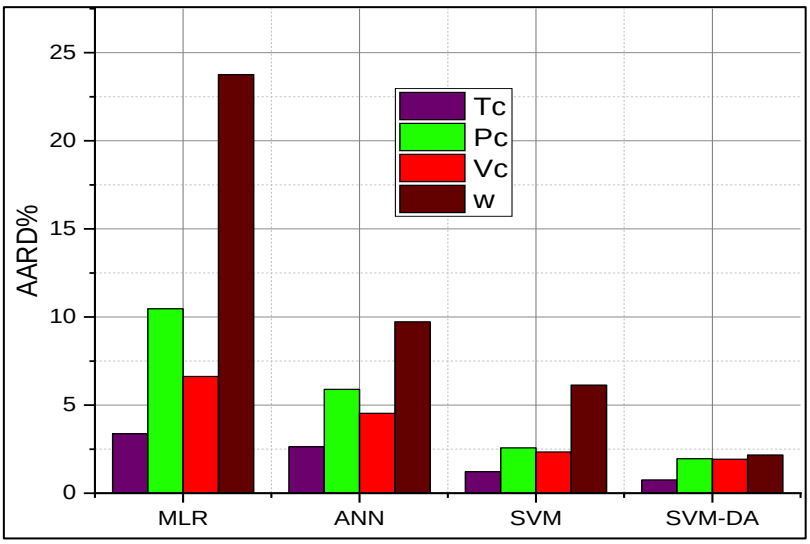


Figure III.5. Prédiction performance des 04 paramètres par les 04 modèles.

III.3.5. Domaine d'applicabilité des modèles développés

Dans cette étude, le domaine d'applicabilité a été présenté à l'aide de la méthode de William [100]. Dans cette méthode les résidus (δ) normalisés par rapport aux valeurs de levier (h) et l'effet de levier (h_i) du $i^{ème}$ composant est défini comme suit:

$$h_i = X_i(X^T X)^{-1} X_i^T \qquad i = 1, 2, 3 \dots n \qquad \text{(III.6)}$$

Où X est la matrice des entrées de l'ensemble d'apprentissage, X_i est le vecteur des entrées du composé souhaité, l'exposant signifie la transposée et n est le nombre de points d'apprentissage. Dans ce graphique, l'effet de levier critique (h^*) est calculé selon l'équation suivante [100]:

$$h^* = 3(p+1)/n \qquad \text{(III.7)}$$

Où p est le nombre d'entrées dans le modèle. Les résidus normalisés δ sont calculés par y:

$$\delta = \frac{y_i - \hat{y}_i}{\sqrt{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2/(n-p-1)}} \qquad \text{(III.8)}$$

Où y_i et $\hat{y}_i$ sont la valeur expérimentale et la valeur calculée pour le $i^{ème}$ composé, respectivement.

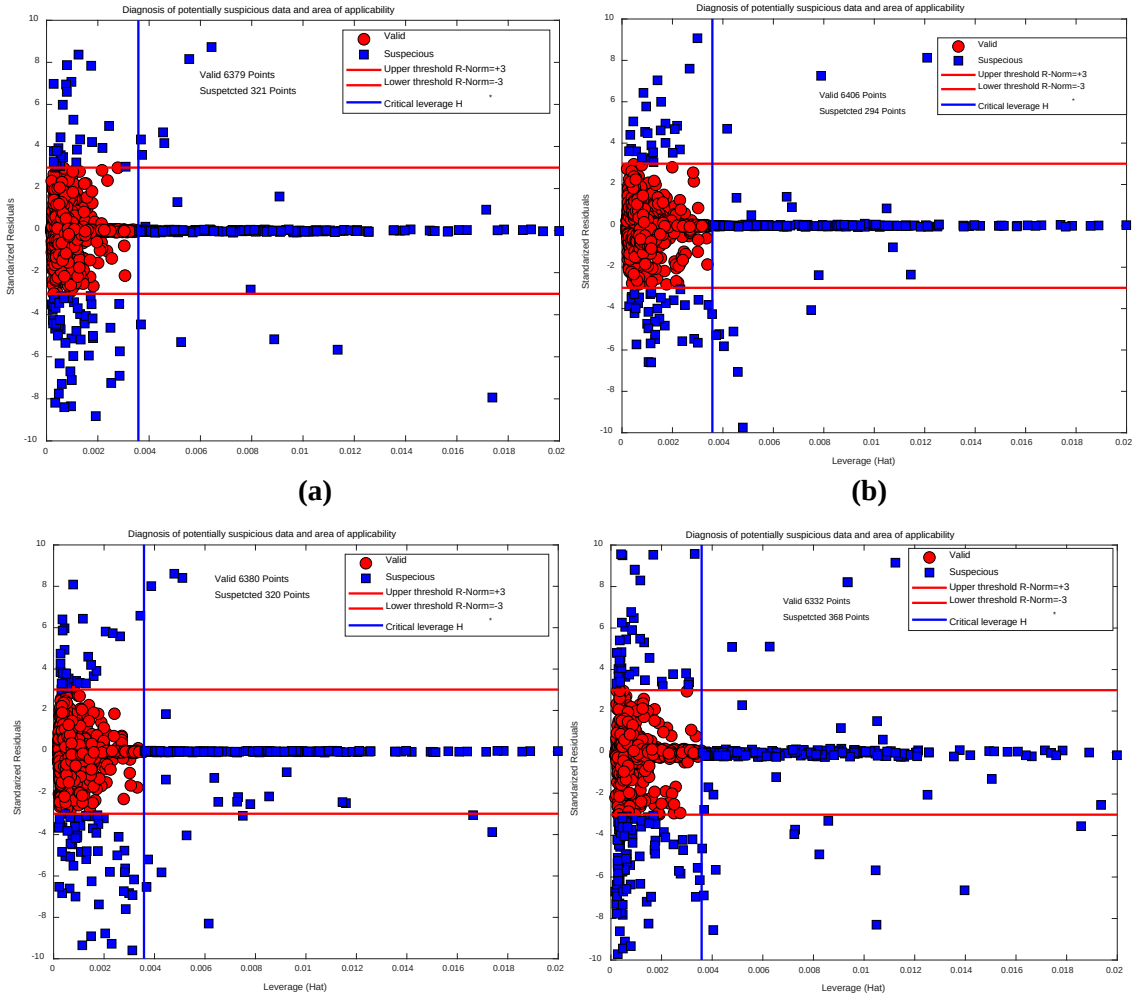


Figure III. 6. Domaine d'application (William) pour chaque propriété (a) T_C, (b) P_C, (c) V_C et (d) ω.

Les valeurs aberrantes sont des points de données lorsque le résidu normalisé se situe dans l'intervalle de ±3 unités d'écart-type et que la valeur de levier doit être supérieure à *h**. Comme on peut l'observer sur la figure (III.6), les valeurs aberrantes ne représentent qu'environ 4.791, 4.382, 4.772 et 5.492% du total des composés dans l'ensemble de données pour T_C, P_C, V_C et ω respectivement. Les modèles développés étaient capables de prédire la majorité des données expérimentales avec des valeurs de résidu normalisé se situe dans l'intervalle de ±3. Quelques points sont hors domaine d'application avec des valeurs plus élevées de *h* ou des résidus normalisés. Cependant, les valeurs aberrantes ne signifient pas que la capacité de prédiction du modèle n'est pas bonne, les ERAM et R sont le jugement de la capacité prédictive du modèle SVM-DA.

III.3.6. Évaluation statistique et validation de modèle élaboré

Évaluation statistique de modèle élaboré montre que plus de 90% des valeurs de l'erreur relative absolue ERA% des propriétés modélisées par le modèle SVM-DA sont inférieur à 2% (figure III.7). Les résultats prouvent que ce modèle a fait une bonne prédiction en produisant une très faible erreur de presque tout le jeu de données.

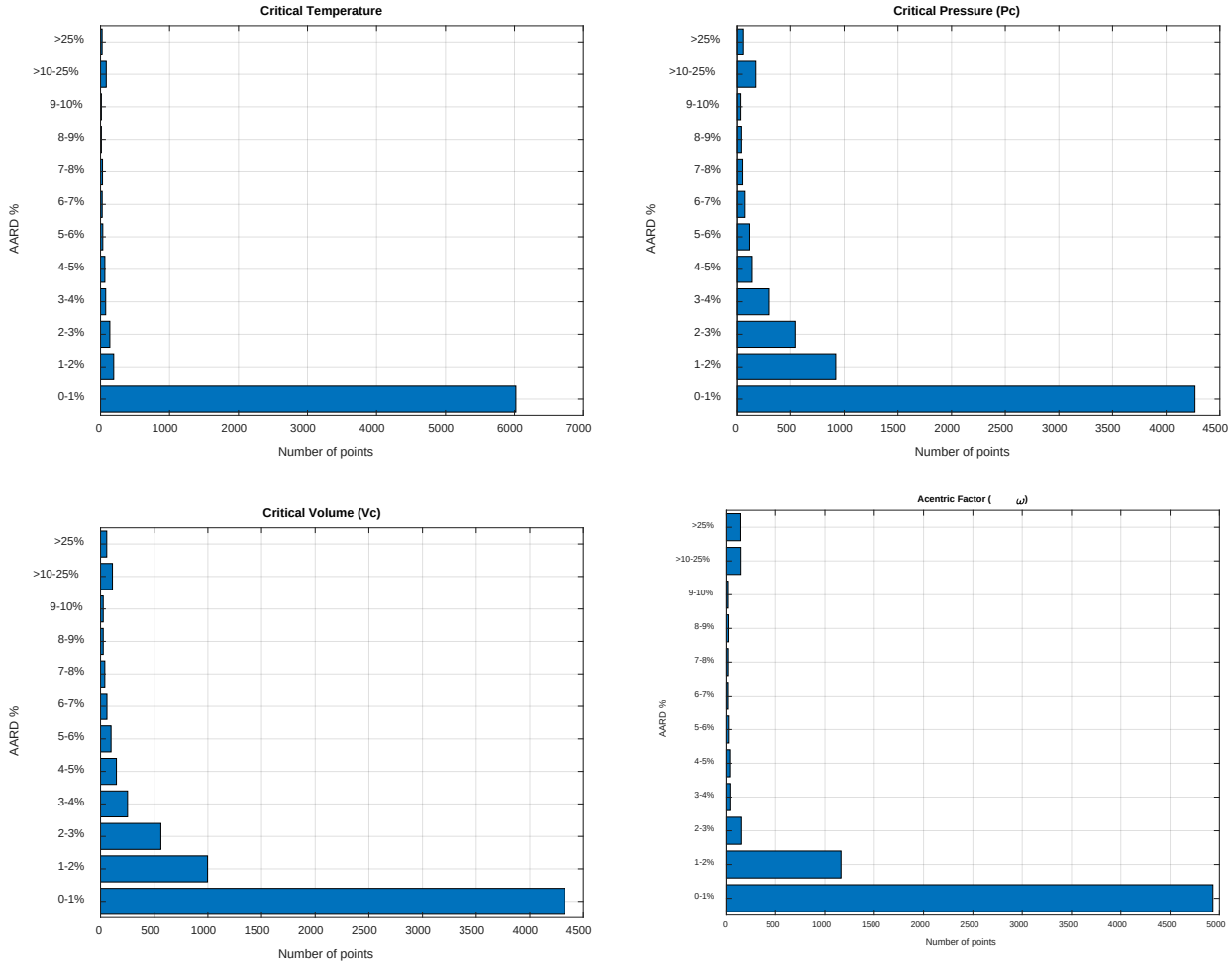


Figure III.7. Erreur Relative Absolu (ERA%) des modèles proposés pour la prédiction de T_C, P_C, V_C et ω d'environ 6700 points de données.

III.3.7. Analyse de sensibilité globale

L'objectif de l'analyse de sensibilité globale est de calculer l'effet des entrées sur chaque sortie en utilisant la méthode d'amplitude cosinus (CAM) [101–103]. Cette méthode est basée sur le calcul du facteur de l'importance (r_{ij}) qui mesure les relations entre les paramètres les paramètres d'entrées x_i et de sortie x_j, ce facteur peut être représenté par l'expression suivante [97]:

$$r_{ij} = \frac{\sum_{k=1}^{m} x_{ik} x_{jk}}{\sqrt{\sum_{k=1}^{m} x_{ij}^2 \sum_{k=1}^{m} x_{jk}^2}}$$

Les paires de données utilisées construisent un tableau de données X qui sont définis comme $X = \{X_1, X_2, X_3,...,X_m\}$

où X_i est un vecteur de longueurs m, exprimé comme: $X_i = \{x_{i1},x_{i2},x_{i3},...,x_{im}\}$;
i et k sont les dimensions de la matrice d'entrée (i = 1 à 6700, k = 1 à m) ;
j est les dimensions du vecteur de sortie (j = 1 à 6700) ;
m est le nombre de paramètres d'entrée (m = 7).

La figure (III.8) montre l'importance des entrées sélectionnées sur chaque sortie, il est montré clairement que tous les paramètres ont approximativement le même effet sur chaque sortie et confirment que tous les descripteurs sélectionnés pour chaque modèle n'ont pas été choisis par hasard

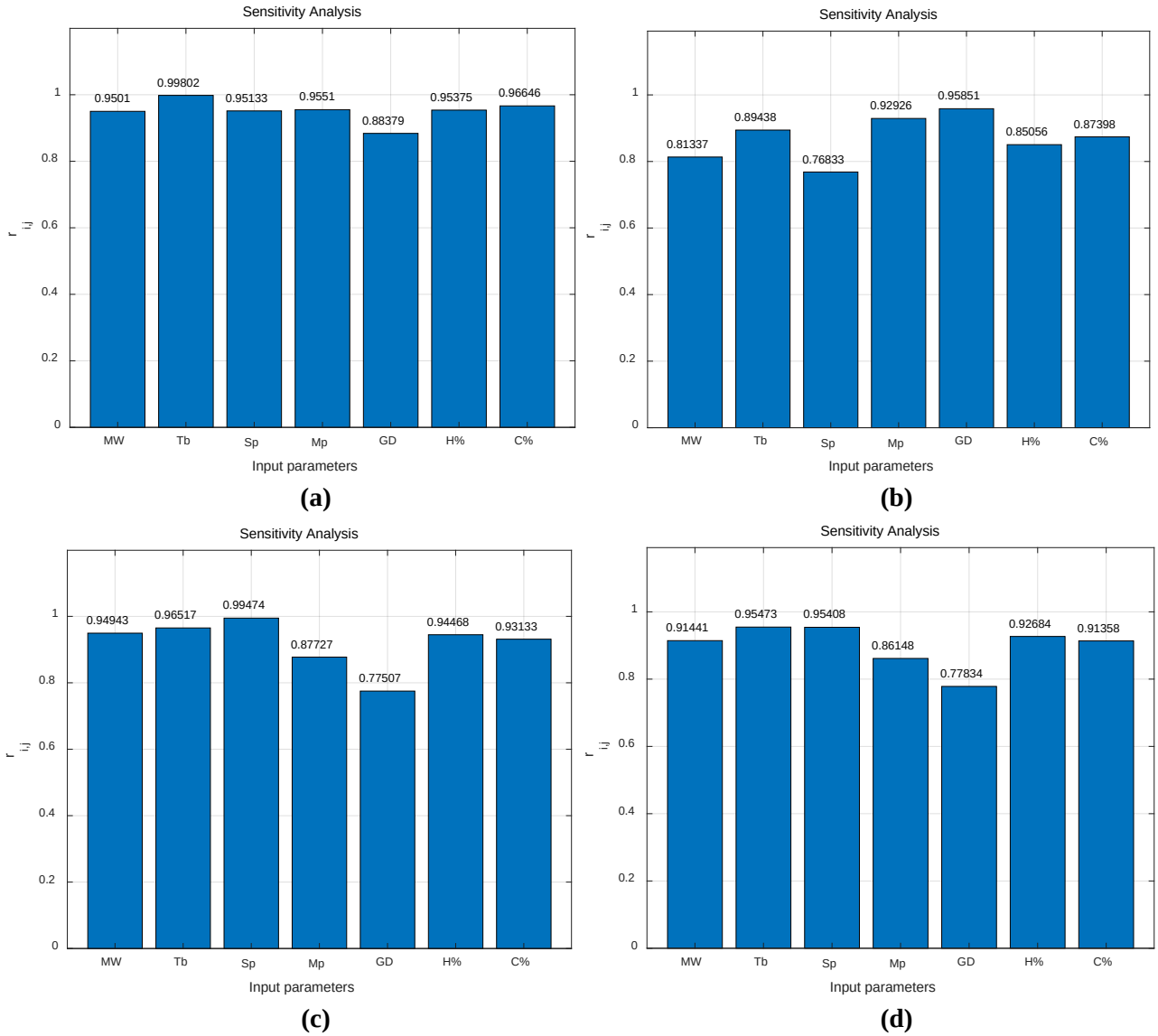


Figure III.8. Importance des entrées sur chaque sortie (a) T_C, (b) P_C, (c) V_C et (d) ω.

Dans le but de mesurer les performances du modèle conçu, nous avons une base de données constituée par un échantillon des propriétés expérimentales. La figure (III.9) montre le

comportement du modèle SVM-DA vis-à-vis de cette base, les propriétés expérimentales (représentées par des cercles bleus) et les propriétés prédites par SVM-DA (représentées par des signes (+) en rouge). Il est clair que les valeurs prédites par ce modèle sont en excellent accord avec les valeurs expérimentales. Avec un coefficient de régression R>0.98 et de erreur relative absolue moyenne ERAM<2.17% pour les quatre propriétés.

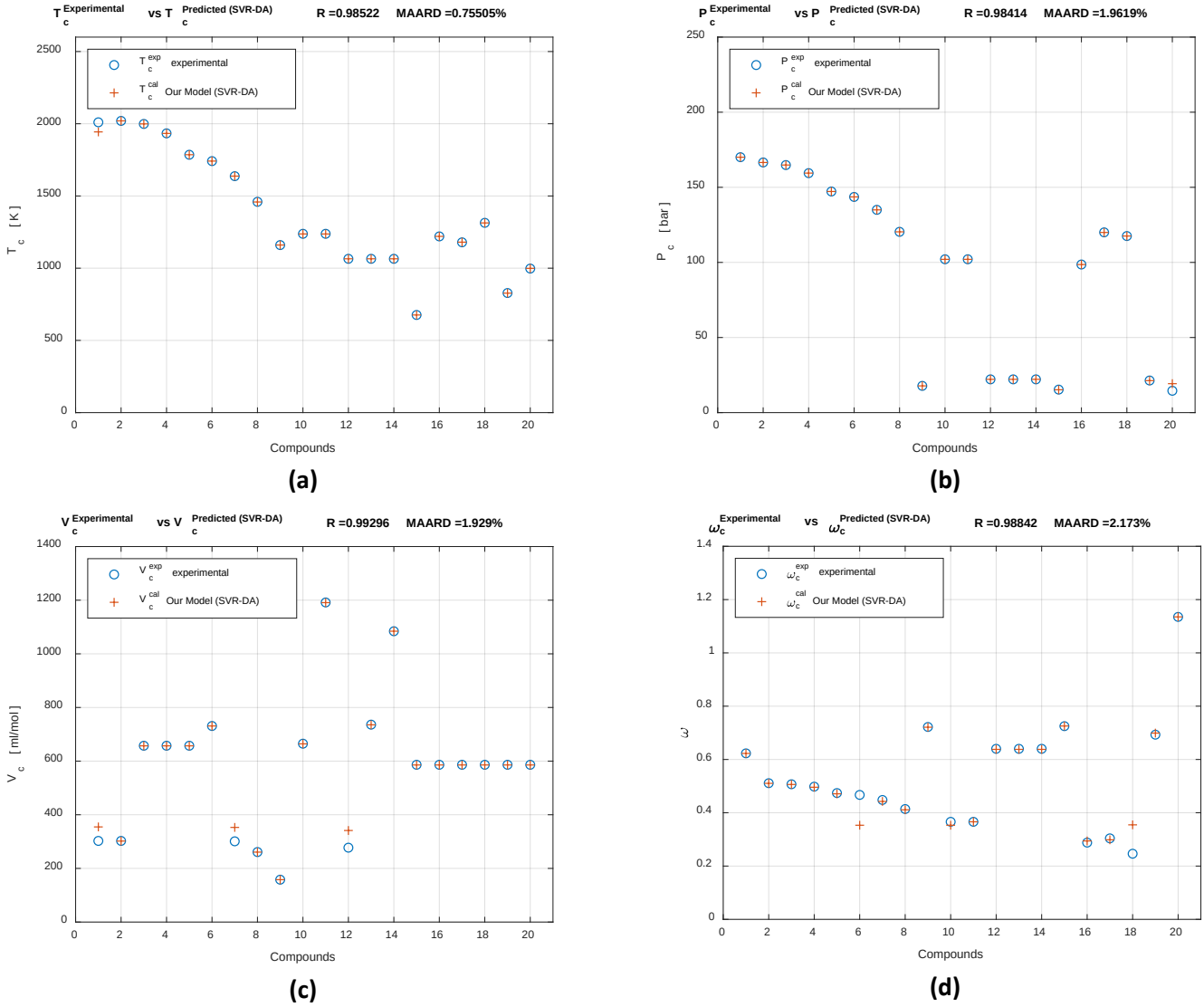


Figure III.9. Les propriétés expérimentales de certains composés vs prédits par SVM-DA.

Le meilleur modèle SVM-DA a été implémenté dans une interface utilisateur graphique conviviale conçue en utilisant le logiciel Matlab et présentée sur la figure (III.10). Cette interface peut calculer les sorties souhaitées en connaissant uniquement les entrées sélectionnées pour 6700 systèmes.

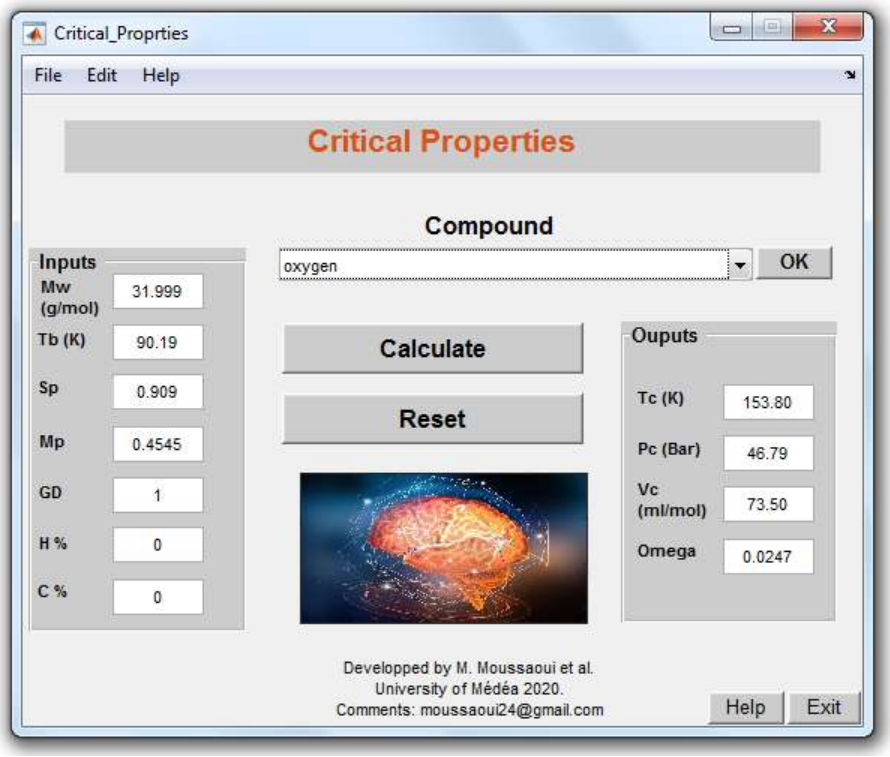


Figure III.10. Interface graphique.

III.4. Conclusion

Le but de cette étude était de tester la pertinence de quatre méthodes de calcul intelligentes dans la modélisation des propriétés critiques et de facteur acentrique de 6700 composés purs. Quatre modèles ont été conçus à savoir, MLR, RNA, SVM et SVM-DA. L'ensemble de données expérimentales des propriétés a été collecté à partir de la littérature tandis que la matrice d'entrée se compose d'un mélange de 05 descripteurs et 02 propriétés thermodynamiques. Tous les modèles ont été évalués statistiquement en termes de ERAM et R^2, les résultats obtenus montrent que le SVM-DA trouvé avec une très faible erreur (ERAM) de {0.7551%, 1.962%, 1.929% et 2.173%} et un coefficient de détermination élevé R^2 de {0.9699, 0.9673, 0.9856 et 0.9766} pour les quatre propriétés (*Tc*, *Pc*, *Vc* et *w*) respectivement. Il est conclu que le modèle SVM-DA est capable d'établir une relation non linéaire satisfaisante entre les descripteurs moléculaires et les propriétés critiques de composés purs avec une grande précision.

Sur la base du modèle SVM-DA, une analyse de sensibilité a été menée pour évaluer les effets des entrées sur chaque sortie, et les résultats montrent que toutes les entrées affectent fortement la sortie et leur choix était basé sur leur contribution à la phénoménologie étudiée.

Une autre section a été faite dans cette étude, qui est le domaine d'applicabilité du modèle par rapport à l'ensemble de données utilisé, et les résultats montrent que le modèle SVM-DA peut être appliqué pour s'adapter à près de 95% des données utilisées, ce qui montre la capacité du modèle lors de la modélisation des propriétés étudiées par rapport aux autres modèles.

Enfin, une interface utilisateur graphique conviviale a été conçue à l'aide du logiciel Matlab pour faciliter l'utilisation de ce modèle sans connaître aucun détail sur la phénoménologie ou le logiciel utilisé.

Chapitre IV.

Contribution à la modélisation de la solubilité des solutés solides dans le CO_2-SC par les modèles semi empiriques.

Chapitre IV
Contribution à la modélisation de la solubilité des solutés solides dans le CO_2-SC par les modèles semi empiriques.

IV.1. Introduction 86
IV.2. Matériels et Méthodes 86
IV.3. Résultats et discussions 91
IV.4. Conclusions 96

IV.1. Introduction

La détermination précise de la solubilité des composés solides dans les fluides supercritiques et sa corrélation sont essentielles pour le développement de la technologie supercritique dans les industries pharmaceutiques. Plusieurs corrélations empiriques ont été développées pour prédire et/ou estimer la solubilité des solutés solides dans le dioxyde de carbone supercritique (CO_2-SC). Ces modèles sont généralement des méthodes basées sur la densité du solvant supercritique dans lesquelles les effets de température et de pression sont tous deux pris en compte, ce qui conduit à une relation linéaire entre la densité du solvant et les données expérimentales. L'inconvénient principal de cette représentation simple est le caractère semi-empirique, et la nécessité de données expérimentales.

IV.2. Matériels et Méthodes

IV.2.1. Élaboration de la base de données

Les données de mesures expérimentales de la solubilité des médicaments solides dans les fluides supercritiques sont rapportées généralement dans la littérature par les variables intensives : la température, la pression et la solubilité exprimée soit en unité de masse du soluté solide soluble dans un volume de solution 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. Comme nous l'avons indiqué dans l'introduction générale les quatre dernières décennies ont été marquées par le nombre élevé de publications réservées à la mesure de la solubilité de solutés solides dans le CO_2-SC, les données utilisées pour la modélisation par les méthodes semi empiriques sont indiquées dans le tableau (IV.1). Ces données concernent donc 130 composés pour un total de 3269 points expérimentaux. Dans ce tableau, M_w : masse molaire, T : Température, P : pression, y_2 : Solubilité dans le CO_2-SC, N_i : Nombre de points expérimentaux.

Tableau IV.1. Données expérimentales utilisées pour la modélisation semi empirique de la solubilité des solutés solides dans le dioxyde de carbone supercritique.

N°	Solutés solides	MW	*T*(°K)	*P*(bar)	ρ_{CO2} (Kg/m^3)	-log (y_2)	N_i	Références
1	1,5-Naphthalenediamine (1,5-NDA)	158.204	313.15~333.15	110.00~200.00	357.88~840.66	4.7905~5.7212	27	[41]
2	4,4-diaminodiphenylmethane (DADPM)	198.269	313.15~333.15	110.00~200.00	357.88~840.66	3.7940~5.0269	27	[104]
3	4-Aminoantipyrine	203.240	308.20~328.20	100.30~220.40	338.62~879.62	3.4353~4.9830	21	[105]
4	Aspirin	180.157	308.15~328.15	120.00~250.00	504.51~901.23	3.4597~4.2007	24	[106]
5	Cholesteryl acetate	428.701	308.15~328.15	90.00~240.00	385.30~893.89	2.9830~5.3615	24	
6	Cholesteryl benzoate	490.772	308.15~328.15	120.00~270.00	120.00~270.00	4.2774~5.2765	20	
7	Cholesteryl butyrate	456.755	308.15~328.15	100.00~240.00	54.47~893.89	3.0491~4.6497	20	
8	Atorvastatin	558.640	308.00~348.00	121.60~354.60	327.00~955.00	2.8398~6.0000	45	[107]
9	Simvastatin	418.566	308.00~348.00	121.60~354.60	327.00~955.00	3.2716~5.6990	45	
10	Lovastatin	404.550	308.00~348.00	121.60~354.60	327.00~955.00	3.9431~4.9586	45	
11	Rosuvastatin	481.538	308.00~348.00	121.60~354.60	327.00~955.00	3.6126~5.5229	45	
12	Fluvastatin	411.470	308.00~348.00	121.60~354.60	327.00~955.00	3.2211~5.3010	45	
13	Atropine	289.369	308.00~348.00	122.00~355.00	327.00~955.00	2.7773~4.2218	45	[108]
14	Carbamazepine	236.269	308.00~348.00	122.00~355.00	477.00~955.00	4.0269~5.5229	39	
15	Codeine	299.364	308.00~348.00	122.00~355.00	327.00~955.00	2.9101~4.3979	45	
16	Diazepam	284.740	308.00~348.00	122.00~355.00	327.00~955.00	2.9547~3.7959	45	
17	Acenaphthene	154.212	308.15~348.15	121.60~354.60	327.00~955.00	1.8540~2.8938	45	
18	Fluoranthene	202.256	308.15~348.15	121.60~354.60	327.00~955.00	2.8195~3.9829	45	
19	Triphenylene	228.294	308.15~348.15	121.60~354.60	327.00~955.00	4.0222~5.6989	45	
20	Azelaic acid	188.221	313.15~333.15	100.00~300.00	289.95~909.89	4.9948~6.3768	14	[109]
21	Benzocaine	165.189	308.00~348.00	122.00~355.00	327.00~955.00	1.9165~3.5686	40	[110]
22	Benzoin	212.248	308.15~328.15	121.60~236.10	433.42~897.99	3.3872~4.3872	19	[111]
23	Mandelic acid	152.150	308.15~328.15	101.00~230.60	356.78~892.63	2.5370~4.5686	21	
24	Propyl 4-Hydroxybenzoate	180.203	308.15~328.15	94.10~220.90	361.93~885.02	3.2132~4.7212	21	
25	Bisacodyl	361.400	308.00~348.00	122.00~355.00	396.00~955.00	3.2343~5.0458	39	[112]
26	Caffeine	194.190	313.00~353.00	199.00~349.00	591.75~934.98	2.9469~3.5482	24	[113]
27	Naproxen	230.259	313.10~333.10	89.60~193.10	394.16~840.87	4.4976~5.7212	18	[30]
28	Ketoprofen	254.280	313.15~328.15	90.00~250.00	422.63~889.59	3.7258~5.4815	15	
29	Flurbiprofen	244.261	303.15~323.15	89.00~245.50	521.40~929.10	3.7059~4.7768	27	
30	Ibuprofen	206.281	308.15~318.15	80.00~220.00	290.96~890.04	2.1675~4.5229	29	
31	Gemfibrozil	250.340	308.20~328.20	100.10~220.20	400.28~879.40	2.3778~4.5317	21	[114]
32	Clofibric acid	214.645	308.2~328.2	100.10~220.20	337.35~879.40	3.0675~4.4934	21	
33	Irgacure® 2959 photoinitiator	224.256	308.20~328.20	101.00~254.00	326.90~901.80	3.5481~5.2840	21	[115]
34	Dichlone	227.040	313.00~ 333.00	70.70~325.80	160.80~905.80	3.6003~5.1549	23	[116]
35	Menadione	172.183	313.00~333.00	97.20~306.70	273.50~306.70	2.9974~3.1249	18	
36	Methimazole	114.166	308.00~348.00	122.00–355.00	327.00~955.00	2.7215~4.2676	40	[117]
37	Phenazopyridine	213.244	308.00~348.00	122.00–355.00	327.00~955.00	2.6944~4.3565	45	
38	Propranolol	259.349	308.00~348.00	122.00~355.00	327.00~955.00	1.6205~3.4461	45	
39	Nicotinamide	122.127	313.15~373.15	54.00~305.00	90.15~905.74	2.5031~5.0000	24	[118]
40	Nicotinic acid	123.110	313.15~373.15	45.00~302.00	78.13~908.18	4.9825~6.5086	22	
41	Oxymetholone	332.477	308.00~328.00	121.00~305.00	411.00~858.00	3.8268~4.7959	20	[119]
42	Cefixime trihydrate	507.489	308.00~328.00	183.00~335.00	634.00~875.00	6.5199~6.7958	18	
43	Progesterone (Steroids)	314.460	313.15~338.15	120.00~260.00	387.40~884.80	2.8386~4.8539	18	[120]
44	Sinapic acid	224.212	313.00~333.00	200.00~400.00	725.00~956.00	6.1524~8.1549	9	[121]

45	Protocatechuic acid	154.121	313.00~333.00	200.00~400.00	725.00~956.00	5.9731~ 6.8182	9	
46	Chrysin	254.241	313.15~333.15	200.00~400.00	725.00~956.00	6.9469~7.7447	9	
47	Pyrocatechol	110.112	308.15~363.15	100.00~400.00	203.00~975.00	2.3388~3.9136	33 32	[122][123] [124]
48	Phenol	94.113	333.15~363.15	100.00~350.00	208.00~864.00	1.0427~2.9431	33	[123]
49	Resorcinol	110.112	308.15~338.15	120.00~400.00	296.00~975.00	3.0119~3.9586	32	[124]
50	Quinine	324.424	308.15~328.15	80.00~240.00	210.62~906.51	5.0915~7.0809	27	[125]
51	Salicylic acid	138.120	308.15~318.15	92.60~157.90	463.00~824.00	3.4001~4.0132	20	[126]
52	Uracil	112.080	313.00~333.00	100.00~299.90	289.40~910.00	3.8861~5.6383	12	[127]
53	Tributylphosphate	266.318	303.00~363.00	150.00~250.00	372.00~923.00	1.0459~1.3152	24	[128]
54	N-phenylacetamide	135.166	308.00~328.00	104.40~225.00	356.70~889.20	3.3408~4.5331	24	[129]
55	o-Phthalic acid	166.132	308.00~328.00	80.00~210.00	206.80~875.60	5.2899~5.9957	15	[130]
56	1-aminoanthraquinone	223.231	323.15~383.15	125.00~250.00	236.98~834.01	4.4543~6.2596	18	[131]
57	1-nitroanthraquinone	253.213	323.15~383.15	125.00~250.00	236.98~834.01	4.0788~4.6253	18	
58	4-hydroxycinnamic acid (p-coumaric acid)	164.160	313.00~ 333.00	150.00~500.00	607.00~992.00	5.5930~6.8110	24	[132]
59	3,4-dihydroxycin-namic acid (caffeic acid)	180.159	313.00~ 333.00	150.00~500.00	607.00~992.00	7.3261~9.0969	24	
60	4-hydroxy-3-methoxycinnamic acid (ferulic acid)	194.186	313.00~333.00	150.00~500.00	607.00~992.00	4.3633~5.4356	24	
61	α-Tocopherol	430.717	313.00~ 353.00	199.00~349.00	591.60~935.00	2.4473~3.1938	24	[133]
62	β-Carotene	536.888	313.00~353.00	200.00~350.00	594.20~935.40	3.9281~7.0457	23	
63	2-Methylbenzoic acid	136.150	313.20~333.20	110.00~246.00	369.00~882.00	2.2441~3.6198	18	[134]
64	3-Methylbenzoic acid	136.150	313.20~333.20	110.00~246.00	369.00~882.00	2.2724~3.7352	18	
65	4-Methylbenzoic acid	136.150	313.20~333.20	110.00~246.00	369.00~882.00	3.1537~4.3665	18	
66	Erythromycin	733.937	308.00~348.00	122.00~355.00	327.00~955.00	3.5058~4.3665	45	[135]
67	Naphthalene	128.174	308.15~338.00	71.90~321.50	66.95~777.02	0.4898~3.2881	109	[136] [104]
68	Amiodarone Hydrochloride	681.778	313.42~343.20	120.00~300.00	347.04~909.94	2.9948~4.6003	28	[137]
69	1-amino-4-hydroxyanthraquinone	239.230	323.15~383.15	125.00~250.00	236.79~834.01	4.6117~6.3635	20	[138]
70	1-hydroxy-4-nitroanthraquinone	269.212	323.15~383.15	150.00~250.00	303.03~834.00	5.0634~6.2034	15	
71	Bisphenol A	228.291	308.00~328.00	110.00~210.00	417.06~874.40	5.5212~6.9281	15	[139]
72	p-Nitroaniline	138.126	308.00~328.00	110.00~210.00	414.90~873.67	4.3705~5.2055	15	[140]
73	Methyl Salicylate	152.149	343.15~423.15	90.00~310.00	129.10~613.47	1.1230~3.2924	44	[141]
74	Juglone	174.155	308.20~328.20	92.00~244.00	267.40~895.50	2.7986~4.6989	18	[142]
75	1-Eicosanol	298.555	308.20~328.20	36.20~412.30	71.30~976.40	2.6556~4.8268	24	[143]
76	Eicosanoic Acid	312.538	308.20~328.20	33.70~211.20	62.90~971.80	2.5654~5.2006	21	
77	1-Octadecanol	270.501	318.00~338.00	139.90~452.80	536.80~955.10	1.6402~2.9830	17	[144]
78	Stearic Acid	284.484	318.00~338.00	145.40~467.50	591.17~935.92	1.8327~3.2069	17	
79	Squalene	410.730	313.00~333.00	100.00~400.00	389.00~956.00	1.1124~3.3010	9	[145]
80	Carbazole	167.211	308.20	103.00~201.00	706.22~865.71	4.5591~4.8447	5	[146]
81	Anthracene Oil	178.234	308.20~333.15	80.00~470.00	206.49~981.24	3.6439~5.4828	5 8 23	[146] [147] [148]
82	Ascorbyl palmitate	414.539	308.10~313.10	130.00~200.00	720.15~865.50	5.2358~5.8041	8	[149]
83	butyl hydroxy anisole (BHA)	360.494	313.10~333.10	130.00~200.00	604.38~840.15	1.2218~1.8416	7	
84	Dodecyl gallate	338.444	313.50~333.10	150.00~250.00	604.38~878.02	4.6840~5.5086	8	
85	Propyl gallate	212.201	313.10~333.10	150.00~250.00	604.38~878.02	4.1586~5.2840	8	
86	Ascorbic acid	176.124	313.10	130.00~200.00	720.14~840.15	4.4962~4.7328	4	
87	Chrysene	228.294	308.15	84.00~251.00	539.86~901.61	5.0535~6.1373	11	[150]
88	Triphenylene	228.294	308.15~328.15	85.00~252.00	381.25~898.90	4.3747~5.9957	28	[150]
89	Lactic acid	90.078	313.00~328.00	58.20~197.10	143.07~836.03	2.7695~4.6197	29	[151]

90	2-hydroxyhexanoic acid	132.159	311.00~328.00	74.00~198.30	265.60~848.89	2.5834~4.6383	34	
91	Octacosane	394.772	308.15~318.15	80.00~275.00	485.60~894.80	3.4089~4.9101	28	[152]
92	Triacontane	422.826	308.15~318.15	90.00~250.00	529.90~250.00	3.5210~5.2924	18	
93	Hexadecanoic Acid	256.430	308.00~3018.00	128.00~226.00	639.60~884.56	2.9809~3.5884	10	[153]
94	Octadecanoic Acid	284.484	308.00~318.00	128.00~226.00	639.60~884.56	3.4894~4.1308	10	
95	Tetradecanoic acid	228.376	308.00~318.00	99.00~226.50	641.91~884.92	2.1278~2.9508	11	
96	Artemisinin	282.336	310.10~338.10	100.00~270.00	278.81~904.26	2.5753~4.0088	36	[154]
97	Ergosterol	396.659	318.15~333.15	120.00~240.00	421.84~848.03	4.5230~5.5324	15	[155]
98	Cholesterol	386.664	313.15~333.15	100.00~250.00	384.40~879.60	3.8386~5.6383	24	[106] [156]
99	(2,4-dichlorophenoxyacetic Acid) (2,4-D)	221.033	313.10~333.10	104.40~207.90	448.05~846.42	3.9508~4.7852	16	[157]
100	Methyl gallate	184.147	313.00~333.00	10.00~50.00	295.00~992.00	5.3727~7.7190	27	[158]
101	Protocatechualdehyde	138.122	313.00~333.00	10.00~50.00	295.00~992.00	5.4270~8.4948	24	
102	1,4-naphthoquinone	158.156	308.20~328.20	91.00~242.00	267.80~894.93	2.3071~4.3010	18	[159]
103	p-quinone (1,4-benzoquinone)	108.096	308.15~318.15	86.10~292.10	331.97~923.84	1.5028~2.9730	18	[160]
104	9,10-anthraquinone	208.216	273.15~318.15	84.10~306.30	379.83~1055.76	4.1308~6.0545	17	
105	Syringic acid	198.174	313.00~333.00	100.00~500.00	291.41~991.63	4.8959~9.0000	27	[161]
106	Vanillic acid	168.148	313.00~333.00	85.00~500.00	291.42~991.63	4.2167~6.0283	28	
107	Paracetamol (Acetaminophen)	151.165	313.00~353.00	11.00~250.00	19.63~880.24	5.0894~7.1427	10	[162]
108	Lycopenepe	536.885	323.15~353.15	200.00~400.00	595.40~923.80	5.7277~6.1864	20	[163]
109	Testosterone (Steroids)	288.431	308.15~328.15	87.02~242.82	243.85~895.47	4.1543~6.4089	39	[164]
110	O-Hydroxy benzoic acid	138.122	273.15~328.15	81.10~202.60	263.59~1182.34	3.2426~5.1549	48	[165]
111	Acetanilide	135.166	308.20~323.20	90.00~400.00	383.93~972.04	0.3003~4.8706	29	[166]
112	Propanamide	73.095	308.20~323.20	90.00~400.00	284.75~972.04	2.2453~3.6769	30	
113	Butanamide	87.122	308.20~323.20	90.00~400.00	284.75~972.04	0.9780~4.1026	30	
114	Cinnamic acid	148.161	308.20~328.20	12.30~23.61	546.67~889.50	3.3716~4.4559	19	[167]
115	Geranyl butyrate (10-Undecenoic acid)	184.279	308.00~333.00	100.00~180.00	291.42~845.47	1.7594~3.3979	18	[168]
116	10-Undecenoic acid	184.279	308.00~333.00	100.00~180.00	291.42~845.47	1.7594~3.3979	18	
117	2-Trifluoromethylbenzoic acid	190.121	308.20~323.20	93.40~226.00	370.04~883.63	2.0680~3.5850	21	[169]
118	3-Trifluoromethylbenzoic acid	190.121	308.20~323.20	94.10~225.40	345.25~883.06	1.3288~3.1487	21	
119	4-Trifluoromethylbenzoic acid	190.121	308.20~323.20	96.80~224.40	364.09~882.48	3.3410~4.5086	21	
120	Cannabinol	310.437	314.00~334.00	130.00~202.00	472.66~836.80	3.3458~3.8996	34	[170]
121	Ethanamide (acetamide)	59.068	308.20~323.20	90.00~400.00	284.75~972.04	2.5000~3.6228	30	[171]
122	2-propenamide (acrylamide)	71.079	308.20~323.20	90.00~400.00	284.75~972.04	2.7875~3.9233	28	
123	Hinokitiol	164.204	313.20~333.20	101.40~378.30	523.33~946.73	2.6038~23.2284	30	[172]
124	Iodopropynyl Butylcarbamate	281.093	313.15~333.15	87.60~341.50	417.44~930.67	2.3279~2.9208	27	[173]
125	Phenanthrene	178.234	308.15	113.00~332.00	729.62~943.89	2.5952~3.1249	11	
126	Levulinic Acid	116.116	313.00~342.40	84.10~188.30	268.20~824.70	2.0457~3.3089	34	[174]
127	m-Dinitrobenzene	168.108	308.00~328.00	95.00~145.00	299.78~803.41	2.2565~3.7212	18	[175]
128	1-chloro-2,4-dinitrobenzene	202.550	308.00~313.00	95.00~145.00	560.16~803.41	2.2083~2.7495	12	
129	Meloxicam sodium salt	391.392	303.00~323.00	149.00~255.00	704.40~921.60	4.89.41~5.3556	15	[176]
130	N-(4-Ethoxyphenyl) ethanamide	386.665	308.00~328.00	90.00~190.00	401.55~855.74	4.3685~5.0409	16	[177]

IV.2.2. Structure de la base de données

La base de données contienne la température T(K), la pression P (Bar), la densité du CO_2 (ρ_1) (kg/m^3) et la solubilité du soluté solide dans le CO_2 supercritique y_2 (en fraction molaire). Le tableau (IV.2) montre un échantillon de la structure de la base de données pour le composé Quinine saisie à partir du travail de Zabihi et al. (2017), [125].

Tableau IV.2. Les données expérimentales de Quinine [125].

Composé	T (K)	P (bar)	ρ (Kg/m^3)	$y_2\,10^6$
Quinine	308.15	80	414.11	0.23
	308.15	100	650.89	0.6
	308.15	120	718.75	0.78
	308.15	140	768.24	1.3
	308.15	160	803.34	1.7
	308.15	180	834.14	2.1
	308.15	200	866.64	2.6
	308.15	220	881.56	3.8
	308.15	240	906.51	4.3
	318.15	80	250.51	0.18
	318.15	100	453.14	0.55
	318.15	120	601.14	1.2
	318.15	140	678.94	1.6
	318.15	160	728.8	2.7
	318.15	180	765.65	3.4
	318.15	200	796.16	4.6
	318.15	220	821.1	5.8
	318.15	240	850.92	6.7
	328.15	80	210.62	0.083
	328.15	100	330.61	0.43
	328.15	120	425.39	1.7
	328.15	140	577.47	2.2
	328.15	160	640.69	3.6
	328.15	180	690.19	5.1
	328.15	200	728.75	7.2
	328.15	220	760.18	7.9
	328.15	240	794.56	8.1

IV.2.3. Proposition d'un nouveau modèle

Plusieurs modèles semi-empiriques ont été présentés par différents auteurs, utilisant de trois à neuf paramètres ajustables. Chacun d'entre eux est généralement une modification d'un modèle précédent. Bien que des études comparatives des différents modèles ont été réalisées, elles restent limitées en terme de systèmes étudiés et en terme de modèles comparés afin d'arriver à des conclusions convaincantes concernant le meilleur modèle ainsi que l'influence du nombre de paramètres ajustables sur la qualité de corrélation de ces modèles.

Le tableau (IV.3) récapitule les modèles fondés sur la densité proposés dans les dix 10 dernières années. Dans ce tableau, y_2 est la solubilité du soluté solide (fraction molaire), ρ_1 est la densité du SCF (en kg / m^3), T est la température (en K), P est la pression (en bar), a_i représenté les paramètres caractéristiques déterminés en ajustant les données expérimentales du soluté / solvant considéré.

Tableau IV. 3. Modèles semi-empiriques basée sur la densité pour la corrélation de la solubilité des solutés solide dans CO_2-SC.

Modèles	Equations	Références
Garlapati 2010	$\ln(y_2) = a_0 + (a_1 + a_2\rho_1)ln\rho_1 + \frac{a_3}{T} + a_4 ln(\rho_1 T)$	[153]
Keshmiri et al. 2014	$ln(y_2) = a_0 + \frac{a_1}{T} + a_2 P^2 + \left(a_3 + \frac{a_4}{T}\right) ln\rho_1$	[47]
Bian et al. 2016	$\ln(y_2) = a_0 + \frac{a_1}{T} + \frac{a_2\rho_1}{T} + (a_3 + a_4\rho_1)ln\rho_1$	[178]
Si Moussa et al. 2016	$ln(y_2) = a_0 + a_1\rho_1 + a_2{\rho_1}^2 + a_3\rho_1 T + a_4\frac{T}{\rho_1} + a_5 ln\rho_1$	[179]
Belghait et al. 2018	$\ln(y_2) = a_0 + a_1\rho_1 + a_2{\rho_1}^2 + a_3\rho_1 T + a_4 T + a_5 T^2 + a_6 ln\rho_1 + \frac{a_7}{T}$	[3]

Afin de contribuer à l'amélioration de la qualité de corrélation des données de solubilité des solutés solides dans le CO_2-SC. Nous avons choisi dans une 1^{ére} étape une base de données de 130 composés pour comparer la qualité de corrélation de cinq (05) modèles suggérés dans la littérature comme ayant les meilleures performance de corrélation. Sur la base de cette comparaison nous avons proposé un nouveau modèle semi-empirique basée sur la densité de solvant pour améliorer la qualité de corrélation.

Ce nouveau modèle dérivé d'une simple modification et combinaison du modèle de Si Moussa et al. (2016) [179] et le modèle de Belghait (2018) [3]. Cette modification donne le modèle **proposé** suivant :

$$\boldsymbol{ln(y_2) = a_0 + \left(a_1\rho_1 + a_2{\rho_1}^2 + a_3\rho_1 T\right) ln(T) + a_4\frac{T}{\rho_1} + \left(a_5 + a_6\rho_1 + a_7{\rho_1}^2\right) ln(\rho_1)} \quad \text{(IV.1)}$$

Les paramètres a_0 à a_7 sont obtenus par régression des données expérimentales.

IV.3. Résultats et discussions

IV.3.1. Ajustement des paramètres du modèle proposé

Les ajustements des paramètres des meilleures modèles de littérature sélectionnés (cinq modèles) ainsi que du modèle proposé (Eq. IV.1) ont été effectués à l'aide d'un

algorithme qui calcule la fonction objective c'est-à-dire la valeur de ERAM (l'Erreur Relative Absolue Moyenne) de chaque modèle et pour tous les composés. Cet algorithme a été implémenté dans un programme MATLAB qui comporte les étapes suivantes :

1. Lecture des données expérimentales qui sont présenté dans des fichiers Excel utilisant la fonction *xlsread* pour former une matrice qui contient les valeurs expérimentales, la température en (K), la pression en (bar), la densité en (Kg/m^3) et la solubilité en fraction molaire de chaque composé,
2. Approximations initiales des paramètres ajustés en minimisant la fonction objective (ERAM) par la fonction *ga* de MATLAB qui utilise les algorithmes génétiques.

$$\text{ERAM}\% = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{(y_2^{\text{exp}}(i)-y_2^{\text{cal}}(i))}{(y_2^{\text{exp}}(i))}\right| \qquad \text{(IV.2)}$$

3. Les approximations initiales sont utilisées ensuite pour affiner les paramètres ajustables en utilisant la fonction *lsqcurvefit* de MATLAB qui utilise la méthode des moindres carrés non linéaire en fixant en option du calcul l'utilisation de l'algorithme de *Levenberg-Marquardt*.
4. Calcul de la fonction objective, plus précisément la valeur de l'erreur relative absolue moyenne, utilisant une fonction qui comporte les solubilités expérimentales et les solubilités calculés déterminées 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 donné du composé étudié en utilisant la fonction *postreg* de MATLAB.

Les étapes 1 à 5 sont réalisées pour les 130 solutés solides de la base de données et représenté sur l'organigramme suivant.

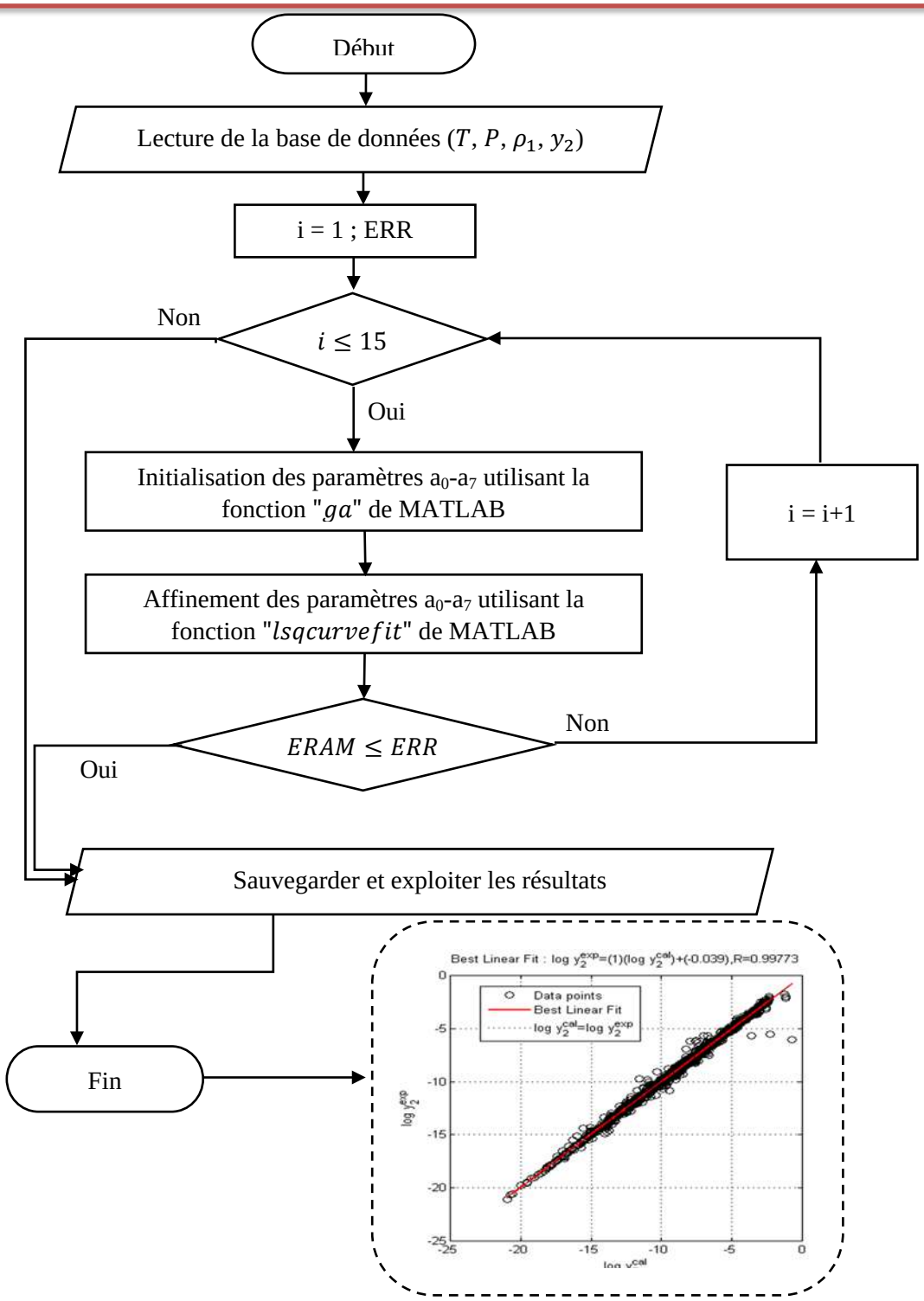


Figure IV.1. Organigramme de l'ajustement des paramètres

IV.3.2. Comparaison globale entre les performances de différents modèles

Les résultats de corrélation des données expérimentales de cette étude sont présentés en termes de tableau comparatif de l'erreur relative absolue moyenne ERAM (équation (IV.2)) pour chaque composé et globale pour chaque modèle. Des courbes de régressions

linéaires entre le logarithme décimal de la solubilité expérimentale et la solubilité calculée par le modèle semi-empirique pour l'ensemble des données (3269 points expérimentaux) ont été tracées.

Les résultats de calcul des ERAM (%) sont regroupés dans le tableau (IV.4). Les colonnes représentent les 05 modèles ainsi que le modèle que nous avons proposé dans cette étude. La comparaison globale entre les performances de différents modèles indique que le modèle proposé dans ce travail fournit la meilleure performance de corrélation, où l'ERAM varie entre 2.34 10^{-11}% et 65.09% avec une moyenne de 7.43%. La deuxième meilleure performance en corrélation est donnée par Si Moussa (2018) [3], où l'ERAM avec une moyenne de 8.59%.

Tableau IV.4. Comparaison globale de l'erreur de calcul par chacune des 06 modèles étudiées.

Erreurs	Garlapati (2010)[177]	Keshmiri (2014)[47]	Si Moussa (2016)[179]	Si Moussa (2018)[3]	Bian (2016)[180]	Notre Modèle
R	0.9954	0.9958	0.99675	0.9972	0.9962	**0.9978**
Min (ERA %)	$7.9590\ 10^{-14}$	$3.1473\ 10^{-13}$	$7.3498\ 10^{-12}$	8.4973210^{-1}	1.1052610^{-1}	**$2.34004\ 10^{-11}$**
Max (ERA %)	83.3420697	74.9072983	70.5027024	62.851034	79.0369891	**65.09721593**
Moyenne (ERA %)	13.1216563	12.6968104	9.84360819	8.59978512	11.4403401	**7.428115245**
Nbr. de systèmes les mieux corrélés	3	2	2	27	3	**93**

Ces résultats sont illustrés plus en détail sur la figure (IV.2), qui représente le nombre de systèmes les mieux corrélées. Il est clair que le modèle proposé dans ce travail est supérieur aux autres modèles puisqu'il corrèle au mieux 93 systèmes (72%) sur les 130 solides considérés. Le deuxième meilleur est le modèle à huit paramètres de Si Moussa et al. (2018) [3] avec 27 systèmes (20%).

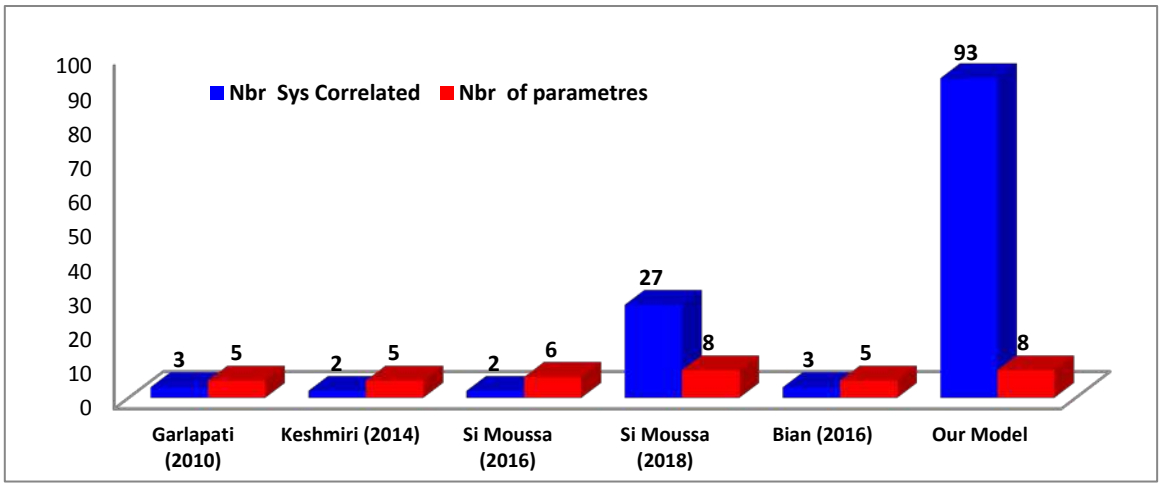


Figure IV.2. Nombre de systèmes les mieux corrélés par chaque modèle.

IV.3.3. Diagramme de dispersion et Vecteurs de validation (α, β, R):

La validation et les performances des modèles sélectionnés dans cette étude sont présentées sous forme de diagrammes de dispersion pour les 3269 points expérimentaux de la base de données. La figure (IV.3) montre le diagramme de dispersion du modèle proposé. Ces graphiques sont générés à l'aide de la fonction *postreg* de *MATLAB*, dans laquelle la solubilité calculée en échelle logarithmique est tracée en fonction de la solubilité expérimentale. Le diagramme de dispersion montre la dispersion du nuage de points de l'ensemble des données autour de la première bissectrice (log (y^{cal}) = log (y^{exp})), le degré de dispersion peut être évalué à partir de l'équation du meilleur ajustement linéaire (log (y^{cal}) = α. Log (y^{exp}) + β), où α est la pente et β est l'ordonnée à l'origine, avec le coefficient de corrélation R, sont utilisés comme vecteur de validation. La performance idéale est atteinte lorsque [α = 1, β = 0 et R=1].

La figure (IV.3) montre clairement que le modèle proposé a été trouvé avec un meilleur vecteur de validation (meilleures performances globales de corrélation).

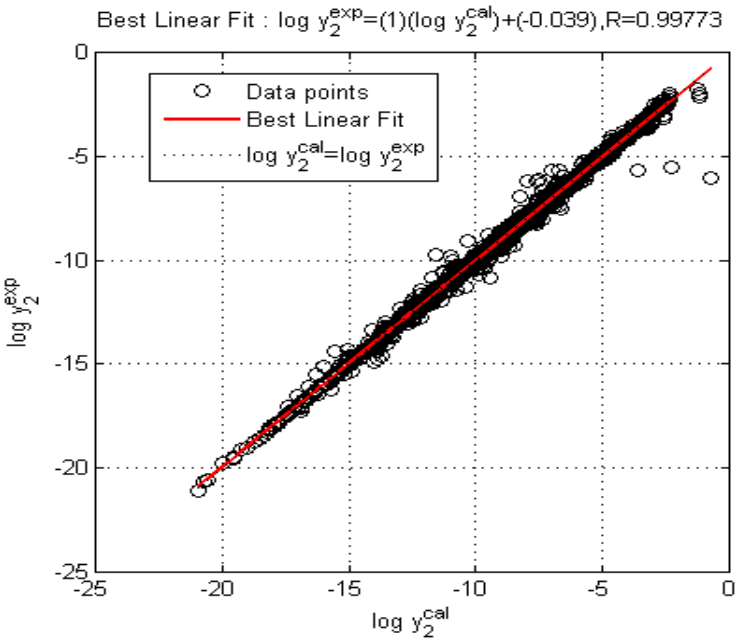


Figure IV. 3. Comparaison entre la solubilité expérimentale et prédite de 130 solutés dans CO_2-SC à l'aide d'un modèle semi-empirique.

IV.3.4. Discussion des résultats

Les interprétations qu'on peut formuler à partir des résultats obtenus dans cette étude sont les suivantes :

- Les résultats du tableau (IV.3). confirment d'une manière générale les résultats obtenus dans cette étude (les six modèles considérés) dans le sens qu'il n'existe pas de modèle qui donne la meilleure corrélation pour tous les composés étudiés.
- La base de données à une très grande importance lorsqu'il s'agit d'études comparatives.
- Les mêmes résultats obtenus lors de cette étude ont permis d'apporter une réponse à une conclusion souvent controversée dans la littérature relative à l'influence des nombre de paramètres ajustable sur la qualité de corrélation des modèles semi-empiriques. Cette étude a montré que les meilleures corrélations des données de solubilité expérimentale pour plus de 80% des composés ont été obtenues par les modèles à huit paramètres (Belghait 2018 [3], modèle proposé). Le performance des modèles de cinq et six paramètres (Garlapati 2010 [153], Keshmiri 2014 [47], Bian 2016 [178], Si Moussa 2016 [179]) sont plus faibles que celles des modèles à huit paramètres. Ceci signifie que le phénomène de solubilisation des solutés solides dans le CO_2-SC ne peut être représenté par une simple relation entre le logarithme de la solubilité et la densité.
- Le diagramme de dispersion permet de mieux apprécie la qualité de corrélation globale (3269 points expérimentaux) des modèles étudiés. En effet la figure (IV.3) montre la supériorité du modèle proposé dans cette étude. En se référant aux éléments de jugement sur la qualité de corrélation, cité plus haut, on constate que le modèle proposé possède le coefficient de corrélation le plus élevé (R=0.998) et une dispersion des points autour de la première bissectrice la plus faible

IV.4. Conclusions

Le nouveau modèle basé sur la densité proposé dans ce travail a été dérivé d'une simple modification du modèle de Belghait à huit paramètres afin de corréler la solubilité des solutés solides dans le CO_2-SC. La modification considère la règle de phase dans la mesure où la solubilité du soluté solide dans la phase supercritique n'est exprimée qu'en fonction de la densité du solvant et de la température d'équilibre. Les performances du modèle ont été testées sur une base de données de 130 médicaments représentant 3269 points expérimentaux collectés dans la littérature. La performance corrélative du modèle proposé a été comparée à celle de cinq modèles précédents qui ont été sélectionnés sur la base de propositions suggérées à partir des travaux antérieurs modèles ayant la meilleure performance.

Cette étude montre que, pour l'ensemble de données considéré, le modèle proposé dans ce travail présente le coefficient de corrélation le plus élevé (R = 0.9978) et fournit les meilleures performances de corrélation, où l'ERAM est de 7.42%.

Chapitre V.

Contribution à la modélisation de la solubilité des solutés solides dans le CO_2-SC par les techniques d'intelligence computationnelle.

Chapitre V
Contribution à la modélisation de la solubilité des solutés solides dans le CO_2-SC par les techniques d'intelligence computationnelle

V.1.	Introduction	97
V.2.	Méthodologie	98
V.3.	Résultats et discussion	105
V.4.	Conclusion	118

V.1. Introduction

La solubilité du soluté solide dans le CO_2-SC est principalement modélisée à l'aide des corrélations basées sur la densité, telles que le modèle de Belghait et al. 2018 [3] et le modèle de Hezave et al. 2013 [181].

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. De plus, RNA a également été utilisé avec succès pour modéliser la solubilité du soluté solide dans le CO_2-SC en se basant uniquement sur des paramètres thermodynamiques [4,182]. Le type le plus courant de modèle RNA qui est utilisé de nos jours est les réseaux de neurones de type MLP formé par l'algorithme de rétro-propagation de gradient (BP) [183]. Ce type de RNA a également été utilisé avec succès pour mapper des relations d'entrée/sortie complexes hautement non linéaires entre une variable dépendante et des variables plus indépendantes de tous les systèmes.

La machine à vecteurs de support SVM peut être considérée comme une alternative aux approches classiques telles que les corrélations semi-empiriques pour modéliser la solubilité du soluté solide dans CO_2-SC [58]. La méthode SVM a gagné un large éventail d'applications d'ingénierie dans la prévision et l'analyse de régression en raison de ses caractéristiques attrayantes et de ses remarquables performances de généralisation [184,185].

Par conséquent, la présente contribution a été réalisée pour évaluer les performances prédictives des techniques d'intelligence computationnelle, y compris RNA-MLP comme type de réseau neuronal le plus utilisé et SVM pour prédire la solubilité de 145 différents solutés solides dans CO_2-SC sur la base d'un mélange entre les propriétés thermodynamiques et les descripteurs moléculaires.

V.2. Méthodologie

Afin de développer un modèle RNA-QSPR ou SVM-QSPR, plusieurs étapes doivent être utilisées, elles sont résumées en détail dans la section suivante (Fig. V.1). Avant que des échantillons de données ne soient transmis à l'un des modèles d'apprentissage automatique, des descripteurs pertinents de chaque système ont été calculés et optimisés et ajoutés aux paramètres thermodynamiques de chaque système [186]. Après cela, les données ont été normalisées de sorte que les problèmes de bruit et les données corrélées soient résolus. Ensuite, diviser les données en deux échantillons d'apprentissage et de test respectivement. Après avoir réglé chaque architecture de modèle, l'optimisation de ses paramètres a été faite pendant la phase d'apprentissage. La phase suivante consiste à évaluer les performances des modèles optimisés en utilisant d'autres données non utilisées lors de la phase d'apprentissage. Dernière phase, des paramètres statistiques doivent être calculés pour évaluer la précision de chaque modèle et valider son utilisation pour la prévision de la solubilité des solutés solides dans la CO_2-SC. De plus, la procédure de modélisation (Fig. V.1) a été exécutée dans le logiciel MATLAB R2018a.

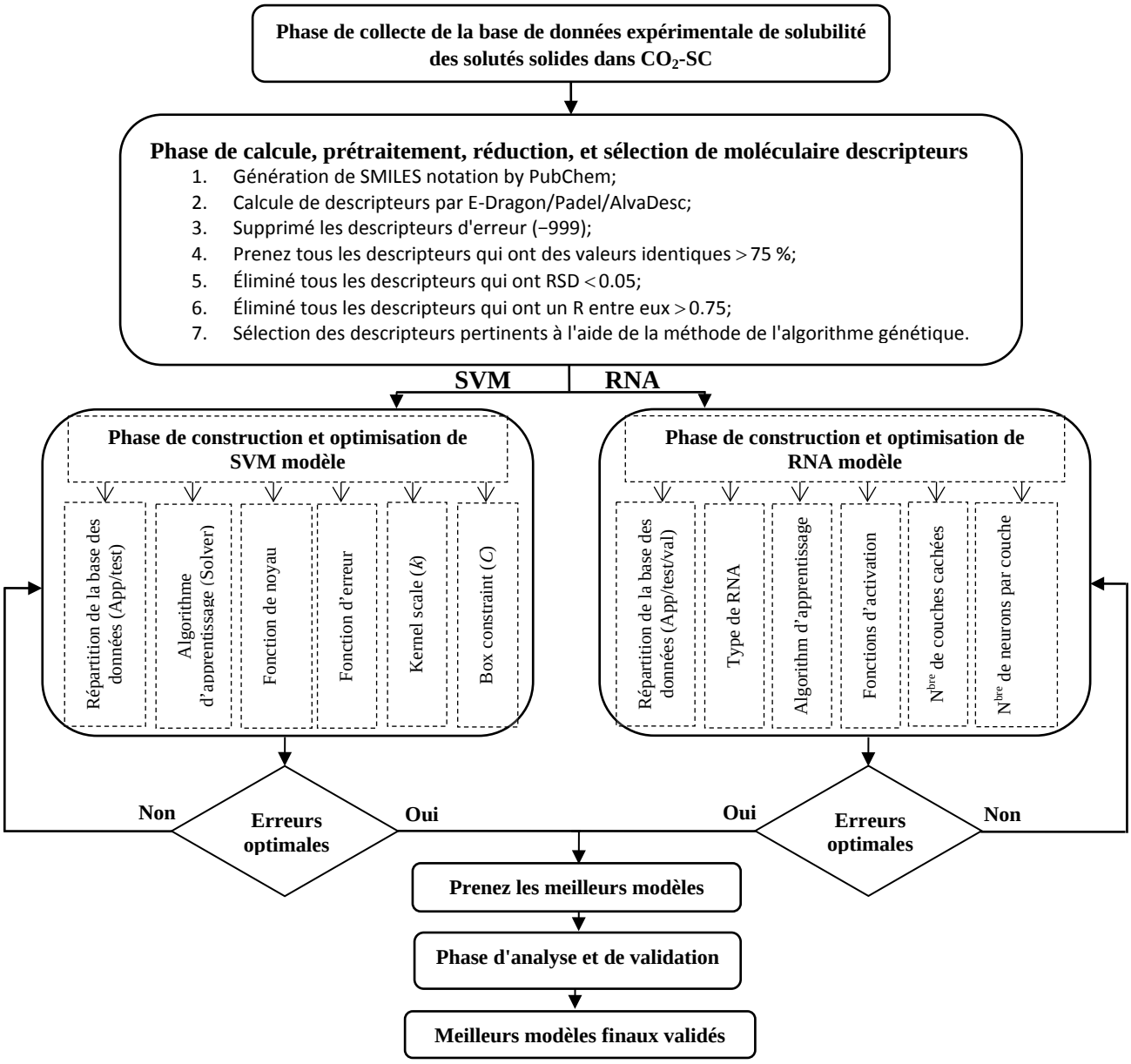


Figure V.1. Procédure de construction et d'optimisation des modèles RNA-QSPR et SVM-QSPR

V.2.1. Préparation des données

La fiabilité et l'efficacité de tout modèle dépendent fortement de la disponibilité de données expérimentales précises. Pour cela, les données expérimentales requises pour l'apprentissage et le test des techniques d'intelligence computationnelle ont été extraites à partir de la littérature. Le tableau (V.1) présente les propriétés des 145 systèmes binaires soluté-CO_2-SC sélectionnés comptant 3637 points de données expérimentale de solubilité dans CO_2-SC, parmi ces paramètres, le nom des médicaments correspondants, le poids moléculaire, la plage de température et de pression de travail, la solubilité expérimentale et les références des 145 systèmes sélectionnés.

Tableau V.1. Données expérimentales utilisées pour la modélisation de solubilité des solutés solides dans le dioxyde de carbone supercritique

N°	solutés solides	MW[g/mol]	T[K]	P [bar]	ρ_{CO2}[kg m^{-3}]	$-\log_{10}(y_2)$	N_i	Refs.
1	1,5-Naphthalenediamine (1,5-NDA)	158.204	313.15~333.15	110.00~200.00	357.88~840.66	4.7905~5.7212	27	[41]
2	4,4-diaminodiphenylmethane (DADPM)	198.269	313.15~333.15	110.00~200.00	357.88~840.66	3.7940~5.0269	27	[104]
3	4-Aminoantipyrine	203.240	308.20~328.20	100.30~220.40	338.62~879.62	3.4353~4.9830	21	[105]
4	Aspirin	180.157	308.15~328.15	120.00~250.00	504.51~901.23	3.4597~4.2007	24	[106]
5	Cholesteryl acetate	428.701	308.15~328.15	90.00~240.00	385.30~893.89	2.9830~5.3615	24	
6	Cholesteryl benzoate	490.772	308.15~328.15	120.00~270.00	120.00~270.00	4.2774~5.2765	20	
7	Cholesteryl butyrate	456.755	308.15~328.15	100.00~240.00	54.47~893.89	3.0491~4.6497	20	
8	Atorvastatin	558.640	308.00~348.00	121.60~354.60	327.00~955.00	2.8398~6.0000	45	[107]
9	Simvastatin	418.566	308.00~348.00	121.60~354.60	327.00~955.00	3.2716~5.6990	45	
10	Lovastatin	404.550	308.00~348.00	121.60~354.60	327.00~955.00	3.9431~4.9586	45	
11	Rosuvastatin	481.538	308.00~348.00	121.60~354.60	327.00~955.00	3.6126~5.5229	45	
12	Fluvastatin	411.470	308.00~348.00	121.60~354.60	327.00~955.00	3.2211~5.3010	45	
13	Atropine	289.369	308.00~348.00	122.00~355.00	327.00~955.00	2.7773~4.2218	45	[108]
14	Carbamazepine	236.269	308.00~348.00	122.00~355.00	477.00~955.00	4.0269~5.5229	39	
15	Codeine	299.364	308.00~348.00	122.00~355.00	327.00~955.00	2.9101~4.3979	45	
16	Diazepam	284.740	308.00~348.00	122.00~355.00	327.00~955.00	2.9547~3.7959	45	
17	Acenaphthene	154.212	308.15~348.15	121.60~354.60	327.00~955.00	1.8540~2.8938	45	
18	Fluoranthene	202.256	308.15~348.15	121.60~354.60	327.00~955.00	2.8195~3.9829	45	
19	Triphenylene	228.294	308.15~348.15	121.60~354.60	327.00~955.00	4.0222~5.6989	45	
20	Azelaic acid	188.221	313.15~333.15	100.00~300.00	289.95~909.89	4.9948~6.3768	14	[109]
21	Benzocaine	165.189	308.00~348.00	122.00~355.00	327.00~955.00	1.9165~3.5686	40	[110]
22	Benzoin	212.248	308.15~328.15	121.60~236.10	433.42~897.99	3.3872~4.3872	19	[111]
23	Mandelic acid	152.150	308.15~328.15	101.00~230.60	356.78~892.63	2.5370~4.5686	21	
24	Propyl 4-hydroxybenzoate	180.203	308.15~328.15	94.10~220.90	361.93~885.02	3.2132~4.7212	21	
25	Bisacodyl	361.400	308.00~348.00	122.00~355.00	396.00~955.00	3.2343~5.0458	39	[112]
26	Caffeine	194.190	313.00~353.00	199.00~349.00	591.75~934.98	2.9469~3.5482	24	[113]
27	Naproxen	230.259	313.10~333.10	89.60~193.10	394.16~840.87	4.4976~5.7212	18	[30]
28	Ketoprofen	254.280	313.15~328.15	90.00~250.00	422.63~889.59	3.7258~5.4815	15	
29	Flurbiprofen	244.261	303.15~323.15	89.00~245.50	521.40~929.10	3.7059~4.7768	27	
30	Ibuprofen	206.281	308.15~318.15	80.00~220.00	290.96~890.04	2.1675~4.5229	29	
31	Gemfibrozil	250.340	308.20~328.20	100.10~220.20	400.28~879.40	2.3778~4.5317	21	[114]
32	Clofibric acid	214.645	308.2~328.2	100.10~220.20	337.35~879.40	3.0675~4.4934	21	
33	Irgacure® 2959 photoinitiator	224.256	308.20~328.20	101.00~254.00	326.90~901.80	3.5481~5.2840	21	[115]
34	Dichlone	227.040	313.00~ 333.00	70.70~325.80	160.80~905.80	3.6003~5.1549	23	[116]
35	Menadione	172.183	313.00~333.00	97.20~306.70	273.50~306.70	2.9974~3.1249	18	
36	Methimazole	114.166	308.00~348.00	122.00–355.00	327.00~955.00	2.7215~4.2676	40	[117]
37	Phenazopyridine	213.244	308.00~348.00	122.00–355.00	327.00~955.00	2.6944~4.3565	45	
38	Propranolol	259.349	308.00~348.00	122.00~355.00	327.00~955.00	1.6205~3.4461	45	
39	Nicotinamide	122.127	313.15~373.15	54.00~305.00	90.15~905.74	2.5031~5.0000	24	[118]
40	Nicotinic acid	123.110	313.15~373.15	45.00~302.00	78.13~908.18	4.9825~6.5086	22	
41	Oxymetholone	332.477	308.00~328.00	121.00~305.00	411.00~858.00	3.8268~4.7959	20	[119]
42	Cefiximetrihydrate	507.489	308.00~328.00	183.00~335.00	634.00~875.00	6.5199~6.7958	18	
43	Progesterone (steroids)	314.460	313.15~338.15	120.00~260.00	387.40~884.80	2.8386~4.8539	18	[120]

44	Sinapic acid	224.212	313.00~333.00	200.00~400.00	725.00~956.00	6.1524~8.1549	9	[121]
45	Protocatechuic acid	154.121	313.00~333.00	200.00~400.00	725.00~956.00	5.9731~ 6.8182	9	
46	Chrysin	254.241	313.15~333.15	200.00~400.00	725.00~956.00	6.9469~7.7447	9	
47	Pyrocatechol	110.112	308.15~363.15	100.00~400.00	203.00~975.00	2.3388~3.9136	33 32	[122],[123] [124]
48	Phenol	94.113	333.15~363.15	100.00~350.00	208.00~864.00	1.0427~2.9431	33	[123]
49	Resorcinol	110.112	308.15~338.15	120.00~400.00	296.00~975.00	3.0119~3.9586	32	[124]
50	Quinine	324.424	308.15~328.15	80.00~240.00	210.62~906.51	5.0915~7.0809	27	[125]
51	Salicylic acid	138.120	308.15~318.15	92.60~157.90	463.00~824.00	3.4001~4.0132	20	[126]
52	Uracil	112.080	313.00~333.00	100.00~299.90	289.40~910.00	3.8861~5.6383	12	[127]
53	Tributylphosphate	266.318	303.00~363.00	150.00~250.00	372.00~923.00	1.0459~1.3152	24	[128]
54	N-phenylacetamide	135.166	308.00~328.00	104.40~225.00	356.70~889.20	3.3408~4.5331	24	[129]
55	o-Phthalic acid	166.132	308.00~328.00	80.00~210.00	206.80~875.60	5.2899~5.9957	15	[130]
56	1-aminoanthraquinone	223.231	323.15~383.15	125.00~250.00	236.98~834.01	4.4543~6.2596	18	[131]
57	1-nitroanthraquinone	253.213	323.15~383.15	125.00~250.00	236.98~834.01	4.0788~4.6253	18	
58	4-hydroxycinnamic acid (p-coumaric acid)	164.160	313.00~ 333.00	150.00~500.00	607.00~992.00	5.5930~6.8110	24	[132]
59	3,4-dihydroxycin-namic acid (caffeic acid)	180.159	313.00~ 333.00	150.00~500.00	607.00~992.00	7.3261~9.0969	24	
60	4-hydroxy-3-methoxycinnamic acid (ferulic acid)	194.186	313.00~333.00	150.00~500.00	607.00~992.00	4.3633~5.4356	24	
61	α-Tocopherol	430.717	313.00~ 353.00	199.00~349.00	591.60~935.00	2.4473~3.1938	24	[133]
62	β-Carotene	536.888	313.00~353.00	200.00~350.00	594.20~935.40	3.9281~7.0457	23	
63	2-Methylbenzoic acid	136.150	313.20~333.20	110.00~246.00	369.00~882.00	2.2441~3.6198	18	[134]
64	3-Methylbenzoic acid	136.150	313.20~333.20	110.00~246.00	369.00~882.00	2.2724~3.7352	18	
65	4-Methylbenzoic acid	136.150	313.20~333.20	110.00~246.00	369.00~882.00	3.1537~4.3665	18	
66	Erythromycin	733.937	308.00~348.00	122.00~355.00	327.00~955.00	3.5058~4.3665	45	[135]
67	Naphthalene	128.174	308.15~338.00	71.90~321.50	66.95~777.02	0.4898~3.2881	109	[136] [104]
68	Amiodaronehydrochloride	681.778	313.42~343.20	120.00~300.00	347.04~909.94	2.9948~4.6003	28	[137]
69	1-amino-4-hydroxyanthraquinone	239.230	323.15~383.15	125.00~250.00	236.79~834.01	4.6117~6.3635	20	[138]
70	1-hydroxy-4-nitroanthraquinone	269.212	323.15~383.15	150.00~250.00	303.03~834.00	5.0634~6.2034	15	
71	Bisphenol A	228.291	308.00~328.00	110.00~210.00	417.06~874.40	5.5212~6.9281	15	[139]
72	p-Nitroaniline	138.126	308.00~328.00	110.00~210.00	414.90~873.67	4.3705~5.2055	15	[140]
73	Methyl salicylate	152.149	343.15~423.15	90.00~310.00	129.10~613.47	1.1230~3.2924	44	[141]
74	Juglone	174.155	308.20~328.20	92.00~244.00	267.40~895.50	2.7986~4.6989	18	[142]
75	1-Eicosanol	298.555	308.20~328.20	36.20~412.30	71.30~976.40	2.6556~4.8268	24	[143]
76	Eicosanoicacid	312.538	308.20~328.20	33.70~211.20	62.90~971.80	2.5654~5.2006	21	
77	1-Octadecanol	270.501	318.00~338.00	139.90~452.80	536.80~955.10	1.6402~2.9830	17	[144]
78	Stearic acid	284.484	318.00~338.00	145.40~467.50	591.17~935.92	1.8327~3.2069	17	
79	Squalene	410.730	313.00~333.00	100.00~400.00	389.00~956.00	1.1124~3.3010	9	[145]
80	Carbazole	167.211	308.20	103.00~201.00	706.22~865.71	4.5591~4.8447	5	[146]
81	Anthraceneoil	178.234	308.20~333.15	80.00~470.00	206.49~981.24	3.6439~5.4828	5 8 23	[146] [147] [148]
82	Ascorbylpalmitate	414.539	308.10~313.10	130.00~200.00	720.15~865.50	5.2358~5.8041	8	[149]
83	butyl hydroxy anisole (BHA)	360.494	313.10~333.10	130.00~200.00	604.38~840.15	1.2218~1.8416	7	
84	Dodecyl gallate	338.444	313.50~333.10	150.00~250.00	604.38~878.02	4.6840~5.5086	8	
85	Propyl gallate	212.201	313.10~333.10	150.00~250.00	604.38~878.02	4.1586~5.2840	8	
86	Ascorbic acid	176.124	313.10	130.00~200.00	720.14~840.15	4.4962~4.7328	4	
87	Chrysene	228.294	308.15	84.00~251.00	539.86~901.61	5.0535~6.1373	11	[150]

88	Triphenylene	228.294	308.15~328.15	85.00~252.00	381.25~898.90	4.3747~5.9957	28	[150]
89	Lactic acid	90.078	313.00~328.00	58.20~197.10	143.07~836.03	2.7695~4.6197	29	[151]
90	2-hydroxyhexanoic acid	132.159	311.00~328.00	74.00~198.30	265.60~848.89	2.5834~4.6383	34	
91	Octacosane	394.772	308.15~318.15	80.00~275.00	485.60~894.80	3.4089~4.9101	28	[152]
92	Triacontane	422.826	308.15~318.15	90.00~250.00	529.90~250.00	3.5210~5.2924	18	
93	Hexadecanoicacid	256.430	308.00~3018.00	128.00~226.00	639.60~884.56	2.9809~3.5884	10	[153]
94	Octadecanoicacid	284.484	308.00~318.00	128.00~226.00	639.60~884.56	3.4894~4.1308	10	
95	Tetradecanoic acid	228.376	308.00~318.00	99.00~226.50	641.91~884.92	2.1278~2.9508	11	
96	Artemisinin	282.336	310.10~338.10	100.00~270.00	278.81~904.26	2.5753~4.0088	36	[154]
97	Ergosterol	396.659	318.15~333.15	120.00~240.00	421.84~848.03	4.5230~5.5324	15	[155]
98	Cholesterol	386.664	313.15~333.15	100.00~250.00	384.40~879.60	3.8386~5.6383	24	[106] [156]
99	(2,4-dichlorophenoxyacetic Acid) (2,4-D)	221.033	313.10~333.10	104.40~207.90	448.05~846.42	3.9508~4.7852	16	[157]
100	Methyl gallate	184.147	313.00~333.00	10.00~50.00	295.00~992.00	5.3727~7.7190	27	[158]
101	Protocatechualdehyde	138.122	313.00~333.00	10.00~50.00	295.00~992.00	5.4270~8.4948	24	
102	1,4-naphthoquinone	158.156	308.20~328.20	91.00~242.00	267.80~894.93	2.3071~4.3010	18	[159]
103	p-quinone (1,4-benzoquinone)	108.096	308.15~318.15	86.10~292.10	331.97~923.84	1.5028~2.9730	18	[160]
104	9,10-anthraquinone	208.216	273.15~318.15	84.10~306.30	379.83~1055.76	4.1308~6.0545	17	
105	Syringic acid	198.174	313.00~333.00	100.00~500.00	291.41~991.63	4.8959~9.0000	27	[161]
106	Vanillic acid	168.148	313.00~333.00	85.00~500.00	291.42~991.63	4.2167~6.0283	28	
107	Paracetamol (acetaminophen)	151.165	313.00~353.00	11.00~250.00	19.63~880.24	5.0894~7.1427	10	[162]
108	Lycopenepe	536.885	323.15~353.15	200.00~400.00	595.40~923.80	5.7277~6.1864	20	[163]
109	Testosterone (steroids)	288.431	308.15~328.15	87.02~242.82	243.85~895.47	4.1543~6.4089	39	[164]
110	o-Hydroxy benzoic acid	138.122	273.15~328.15	81.10~202.60	263.59~1182.34	3.2426~5.1549	48	[165]
111	Acetanilide	135.166	308.20~323.20	90.00~400.00	383.93~972.04	0.3003~4.8706	29	[166]
112	Propanamide	73.095	308.20~323.20	90.00~400.00	284.75~972.04	2.2453~3.6769	30	
113	Butanamide	87.122	308.20~323.20	90.00~400.00	284.75~972.04	0.9780~4.1026	30	
114	Cinnamic acid	148.161	308.20~328.20	12.30~23.61	546.67~889.50	3.3716~4.4559	19	[167]
115	Geranyl butyrate (10-undecenoic acid)	184.279	308.00~333.00	100.00~180.00	291.42~845.47	1.7594~3.3979	18	[168]
116	10-Undecenoic acid	184.279	308.00~333.00	100.00~180.00	291.42~845.47	1.7594~3.3979	18	
117	2-Trifluoromethylbenzoic acid	190.121	308.20~323.20	93.40~226.00	370.04~883.63	2.0680~3.5850	21	[169]
118	3-Trifluoromethylbenzoic acid	190.121	308.20~323.20	94.10~225.40	345.25~883.06	1.3288~3.1487	21	
119	4-Trifluoromethylbenzoic acid	190.121	308.20~323.20	96.80~224.40	364.09~882.48	3.3410~4.5086	21	
120	Cannabinol	310.437	314.00~334.00	130.00~202.00	472.66~836.80	3.3458~3.8996	34	[170]
121	Ethanamide (acetamide)	59.068	308.20~323.20	90.00~400.00	284.75~972.04	2.5000~3.6228	30	[171]
122	2-propenamide (acrylamide)	71.079	308.20~323.20	90.00~400.00	284.75~972.04	2.7875~3.9233	28	
123	Hinokitiol	164.204	313.20~333.20	101.40~378.30	523.33~946.73	2.6038~23.2284	30	[172]
124	Iodopropynylbutylcarbamate	281.093	313.15~333.15	87.60~341.50	417.44~930.67	2.3279~2.9208	27	[173]
125	Phenanthrene	178.234	308.15	113.00~332.00	729.62~943.89	2.5952~3.1249	11	
126	Levulinicacid	116.116	313.00~342.40	84.10~188.30	268.20~824.70	2.0457~3.3089	34	[174]
127	m-Dinitrobenzene	168.108	308.00~328.00	95.00~145.00	299.78~803.41	2.2565~3.7212	18	[175]
128	1-chloro-2,4-dinitrobenzene	202.550	308.00~313.00	95.00~145.00	560.16~803.41	2.2083~2.7495	12	
129	Meloxicam sodium salt	391.392	303.00~323.00	149.00~255.00	704.40~921.60	4.89.41~5.3556	15	[176]
130	N-(4-Ethoxyphenyl) ethanamide	386.665	308.00~328.00	90.00~190.00	401.55~855.74	4.3685~5.0409	16	[177]
131	Ketotifenfumarate (KTF)	425.50	308.2~338.2	120.00~300.00	384.17~929.68	2.9679~4.6757	28	[187]
132	Sertraline hydrochloride	342.69	308.00~338.00	120.00~300.00	347.04~909.94	4.0315~6.0000	28	[188]
133	Aprepitant	534.40	308.15~338.15	120.00~300.00	384.00~944.00	4.1152~5.3468	32	[189]

134	Coumarin-7	333.38	308.00~338.00	90.00~330.00	220.59~944.01	4.9961~5.3819	20	[190]
135	Letrozole	285.30	318.20~348.20	120.00~360.00	318.90~922.00	5.0701~5.7959	20	[191]
136	Imatinibmesylate	589.71	308.20~338.20	120.00~270.00	388.00~914.00	5.3559~7.0000	24	[192]
137	Esomeprazole	345.417	308.20~338.20	120.00~270.00	382.44~912.88	3.0409~4.9586	24	[193]
138	Loratadine	382.88	308.15~338.15	120.00~270.00	388.00~914.00	2.8855~5.3468	24	[194]
139	Lansoprazole	369.363	308.20~338.20	120.00~270.00	382.44~912.88	3.1331~4.9208	27	[195]
140	Sunitinib malate	532.56	308.00~338.00	120.00~270.00	388.00~914.00	4.0675~4.8894	24	[196]
141	Azathioprine	277.263	308.00~338.00	120.00~270.00	388.00~914.00	4.7375~5.5686	24	[197]
142	Sorafenibtosylate	637.03	308.00~338.00	120.00~270.00	388.00~914.00	4.9006~6.1675	24	[198]
143	Repaglinide	452.595	308.00~338.00	120.00~270.00	384.00~914.00	4.0214~5.5376	24	[199]
144	Oxcarbazepine	252.273	308.00~338.00	120.00~270.00	384.00~914.00	4.5726~6.9586	24	[200]
145	Pigment (Phthalocyanine green)	1127.154	308.00~338.00	120.00~270.00	384.00~914.00	3.9165~7.0000	24	[201]

V.2.2. Détermination des descripteurs

Les modèles peuvent également être utilisés pour effectuer une estimation des propriétés car ils établissent des corrélations quantitatives entre diverses propriétés physicochimique et la structure moléculaire [83]. Un modèle Relation Quantitative Structure-Propriété (QSPR) consiste en une relation mathématique entre la propriété d'intérêt et une variété de caractéristiques moléculaires (nommés descripteurs) dérivées de la structure moléculaire, qui vont des indices structurels et topologiques aux propriétés électroniques et quanto-chimiques.

L'une des étapes les plus importantes de la construction des modèles QSPR est la quantification des informations de structure des molécules étudiées,[202] qui sont appelés descripteurs moléculaires. Il existe plus de 11145 descripteurs moléculaires qui peuvent être utilisés pour résoudre plusieurs problèmes dans différents domaines tels que la chimie, le physique, la biologie et d'autres sciences [203].

Dans cette étude, 1666 descripteurs moléculaires ont été calculés en ligne par le logiciel E-Dragon 1.0 [204]. Le calcul de ces descripteurs a été précédé d'une étape très intéressante qui est la génération de la notation simplifiée du système d'entrée de ligne moléculaire "SMILES" (The simplified molecular-input line-entry system) avec l'utilisation du site PubChem (https://pubchem.ncbi.nlm.nih.gov/).

Après le calcul des descripteurs moléculaires, les descripteurs constants et quasi constants pour toutes les molécules ont été éliminés, et les paires avec un coefficient de corrélation supérieur à 0.95 ont été considérées comme intercorrélées, donc l'un d'entre eux dans chaque paire corrélée a été supprimé. Enfin, un algorithme génétique robuste (GA) est une méthode bien acceptée pour sélectionner le sous-ensemble optimal de descripteurs. GA

est une méthode d'optimisation puissante pour rechercher les optima globaux des solutions. Il est développé pour imiter certains des processus observés dans l'évolution naturelle [205].

Après le prétraitement, nous avons réduit le nombre de descripteurs possible, nous avons obtenu trois descripteurs pertinents pour chaque système pour construire les modèles RNA et SVM (Tableau V.2).

Tableau V.2. Les trois descripteurs moléculaires utilisés dans le modèle QSPR

Descripteurs	Définition	Type
MPC2	Nombre de chemins moléculaires d'ordre 2	topologique
WTPT-1	Moléculaire ID	topologique
AATS0i	Autocorrélation moreau-broto moyenne pondérée par le potentiel d'ionisation	autocorrélation

V.2.3. Critères d'évaluation de la performance

La capacité prédictive interne du modèle RNA et SVM a été évaluée par l'erreur quadratique moyenne (EQM), l'erreur relative absolue moyenne (ERAM), la validation croisée (LOO) Q^2_{LOO} et le coefficient de détermination (R^2). Les définitions de ces mesures sont données dans le tableau (V.3), où y_2^{exp} et y_2^{cal} sont respectivement les valeurs de sortie souhaitées (expérimentales) et prévues, $\widehat{y_2^{\mathrm{exp}}}$ est la moyenne des valeurs expérimentales de y_2^{exp}.

Tableau V.3. Définitions des mesures de la performance [206,207]

Mesure	Formule
coefficient de détermination (R^2) critère de validation > 0.8	$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(y_2^{\mathrm{exp}}(i) - y_2^{\mathrm{cal}}(i)\right)^2}{\sum_{i=1}^{N}\left(y_2^{\mathrm{exp}}(i) - \widehat{y_2^{\mathrm{exp}}}\right)^2}$
Erreur Relatif Absolu Moyen (ERAM)	$ERAM\% = \frac{100}{N}\sum_{i=1}^{N}\left\lvert\frac{\left(y_2^{\mathrm{exp}}(i) - y_2^{\mathrm{cal}}(i)\right)}{\left(y_2^{\mathrm{exp}}(i)\right)}\right\rvert$
Erreur Quadratique Moyenne (EQM)	$EQM = \left(\left(\frac{\sum_{i=1}^{N}\left(y_2^{\mathrm{exp}}(i) - y_2^{\mathrm{cal}}(i)\right)^2}{N}\right)\right)^{1/2}$
validation croisée Q^2_{LOO} critère de validation > 0.5	$Q^2_{\mathrm{LOO}} = 1 - \frac{\sum_{i=1}^{N}\left(y_2^{\mathrm{exp_train}}(i) - y_2^{\mathrm{cal_train}}(i)\right)^2}{\sum_{i=1}^{N}\left(y_2^{\mathrm{exp_train}}(i) - \bar{y}_2^{\mathrm{exp_train}}(i)\right)^2}$

La valeur d'EQM et ERAM% varie de 0 à ∞, où le R^2 est utilisé pour déterminer le degré de similitude entre la valeur prédite et sa valeur expérimentale. En général, la meilleure valeur pour R^2 et Q^2_{LOO} est 1 et la meilleure valeur de EQM et ERAM% est zéro, ce qui indique les hautes performances du modèle.

V.3. Résultats et discussion

V.3.1. Résultats du model RNA-QSPR

D'une manière générale, lorsqu'on modélise un processus physique, on détermine, par une analyse préalable du problème, les variables qui ont une influence sur le phénomène étudié. La sélection des entrées consiste à déterminer les variables pertinentes par rapport à la grandeur à modéliser. Elle vise deux points essentiels:

- Réduction de la dimension de l'espace de représentation des variables du modèle;
- Elimination des entrées qui n'ont pas ou peu d'influence sur la sortie.

Il est important de mentionner que chaque ajout d'une variable encombre le modèle et entraîne une multiplication des paramètres du réseau. Raison pour laquelle, la représentation des entrées doit contenir tous les facteurs pertinents par rapport à la grandeur à modéliser, mais seulement ceux-ci.

Dans notre cas on choisit comme des paramètres d'entrée pour notre modèle proposé sont la température, la pression, la densité de CO_2-SC, la température critique, la pression critique, le facteur acentrique et les trois descripteurs sélectionnés les plus pertinentes {MPC2, WTPT-1 et AATS0i}, tandis que la sortie était le logarithme de la solubilité dans CO_2-SC.

Comme les entrées ont des magnitudes différentes et afin de diminuer l'erreur, une normalisation doit être effectuée sur les entrées et les sorties du réseau de neurones artificiels (généralement entre -1 et +1) pour obtenir une convergence rapide.

L'ensemble de données a été divisé en trois sous-ensembles (66%) pour la phase d'apprentissage, (17%) pour le test et (17%) pour la validation, celui qui donne des résultats plus conformés en comparaison avec les autres séparations du modèle RNA.

L'étape cruciale au cours de l'optimisation des *RANs* évidemment c'est l'apprentissage; une propriété très importante dans les réseaux de neurones, après qu'on a opté pour la fonction *Newff*. Afin de découvrir quel algorithme exécute mieux dans notre cas on a testé les algorithmes probablement les plus importants. Après l'optimisation l'algorithme de rétropropagation de régularisation bayésienne semble celui qui donne des résultats plus conformés

Autre étape d'optimisation se base sur le test de différentes architectures du réseau de neurones par la modification du nombre de couches cachées (Une seule couche cachée, deux couches, trois couches et quatre couches). Après on a remarqué que le réseau avec deux couches cachées est le plus performant en terme d'erreur relative absolue.

Tandis que l'effet de la modification du nombre de neurones dans chaque couche cachée (le réseau de neurones avec deux couches cachées, le nombre de neurone a été varié de 10 à 30 neurones). Après on a remarqué que le réseau le plus performant avec architecturée de {9-25-20-1}. Cela signifie que a neuf 09 variables indépendantes (propriétés thermodynamiques et descripteurs), 25 neurones dans la première couche cachée, 20 neurones dans la deuxième couche cachée et une 01 sortie.

Les fonctions tangente sigmoïde (*Tansig*) et linéaire (*Pureline*) ont été utilisées comme fonction de transfert pour les couches cachées et de sortie, respectivement.

Le tableau (V.4) récapitule les différents éléments structurels du réseau de neurones optimal utilisé pour calculer la solubilité dans le CO_2-SC.

Tableau V.4. Architecture du modèle RNA optimisé

type RNA	**Couche d'entrée**	**couche cachée**				**Couche de sortie**		**Algorithme d'apprentissage**
		1st couche cachée		**2nd couche cachée**				
	nombres de neurones	nombre de neurones	Fonction de transfert	nombre de neurones	Fonction de transfert	nombre de neurones	Fonction de transfert	
MLP	09	25	TANSIG	20	TANSIG	1	PURELIN	Bayesian regularization backpropagation

La figure (V.2) affiche le diagramme de dispersion de tous les ensembles de données de la solubilité du soluté solide calculée en utilisant le modèle RNA-QSPR par rapport à la solubilité expérimentale (3637 points expérimentaux). Ces graphiques sont générés à l'aide de la fonction *postreg* de Matlab, retracent la solubilité calculée en fonction de la solubilité expérimentale.

Sur la figure (V.2), la première bissectrice montre l'ajustement exact entre les solubilités corrélées et les données expérimentales, tandis que les points de croisement ont démontré les données de solubilité corrélées par le modèle (RNA) proposé par rapport aux données expérimentales. En détail, plus les points sont proches de la ligne pleine, plus les données de solubilité corrélées sont précises.

L'échelle logarithmique de la solubilité expérimentale par rapport à la solubilité calculée est représentée sur la figure (V.2). Sur la base des résultats obtenus, on peut conclure que le modèle RNA-QSPR était capable de corréler la solubilité des solides dans CO_2-SC avec des erreurs acceptables (tableau V.5).

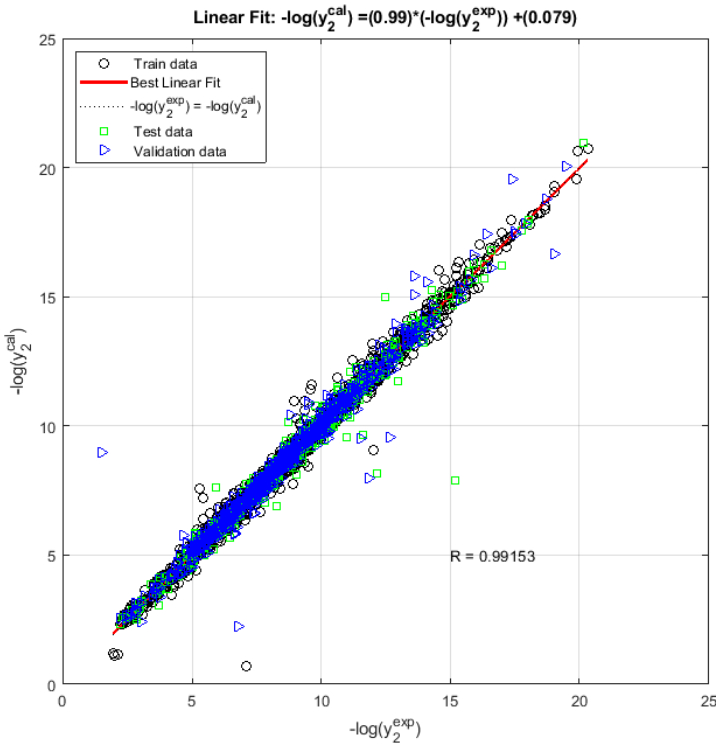


Figure V.2. La régression entre les valeurs expérimentales et prévues de la solubilité de 145 solutés dans CO_2-SC en utilisant RNA.

Pour analyser plus précisément les performances du modèle du *RNA* optimal, le pourcentage de l'erreur relative absolue moyenne (*ERAM %)*, le pourcentage de l'erreur quadratique moyenne (*EAM %)* sont calculés pour les trois sous phases et l'ensemble des données. Comme indiqué par la figure (V.2) et le tableau (V.5), le modèle *RNA* présente une bonne qualité de prédiction de calcul de la solubilité dans le CO_2-SC d'un point de vue global avec un pourcentage de l'*ERAM = 2.772%* et un pourcentage de l'*EQM = 0.3726 %* Pour l'ensemble des données. La comparaison des performances de test du *RN* montre la robustesse du modèle neuronal établi et la possibilité de prédiction de la solubilité dans le CO_2-SC en utilisant les réseaux de neurones statiques.

Tableau V.5. Résultats statistiques et de validation du modèle RNA-QSPR.

	Training	**Test**	**Validation**	**Globale**
R^2	0.9895	0.9733	0.9665	0.9830
ERAM %	2.637	2.900	3.189	2.772
EQM (%)	0.2932	0.4633	0.5241	0.3726
Q^2_{LOO}	0.9895	0.9731	0.9667	0.9830

V.3.2. Modélisation par les Machines à Vecteurs de Support SVM

Machines à Vecteurs de Support (SVM) trouve de nombreuses applications dans différents problèmes de régression [208]. Cette approche d'abord été proposée par *Vapnik* [209]. La qualité du modèle SVM dépend de l'optimisation appropriée des paramètres SVM pour l'ensemble de données. La sélection de ces paramètres (paramètre de largeur de noyau σ, paramètre pénalisant *C* et le paramètre de précision ε) est étudiée dans cette section.

Les mêmes entrées du modèle RNA ont été adoptées pour construire le modèle SVM. L'ensemble de données a été divisé en deux sous-ensembles (66%) pour la phase d'apprentissage, (33%) pour la phase de test du modèle SVM. La procédure d'optimisation a été répétée plusieurs fois afin de trouver l'optimum global le plus probable de la fonction de fitness. Dans l'étude, la fonction de base radiale (RBF) a été sélectionnée comme la fonction de noyau la plus couramment utilisée dans la littérature, les détails des autres paramètres sélectionnés après l'optimisation de modèle SVM sont résumés dans le tableau (V.6).

Tableau V.6. Détails du modèle SVM proposé.

Type	Valeur/Commentaire
Nombre de données utilisées pour l'apprentissage	2425
Nombre de données utilisées pour le test	1212
Algorithme d'optimisation (solveur)	Sequential Minimal Optimisation (SMO)
Fonction de noyau	Radial Basis Function (RBF)
Paramètre pénalisant "*C*"	92
Paramètre de largeur de noyau "σ"	1.8
Quantité de vecteurs de support	2568

Tableau V.7. Analyses statistiques pour le modèle SVM proposé.

	Train	Test	All data Set
R^2	0.9952	0.9666	0.9857
ERAM %	0.895	2.247	1.345
EQM	0.1968	0.5177	0.3393
Q^2_{LOO}	0.9953	0.9675	0.9859

Le tableau (V.7) montre une comparaison entre les valeurs expérimentales et prévues de solubilité pour l'apprentissage, le test et l'ensemble de données. Les résultats montrent que le modèle a capturé les caractéristiques assez précisément avec une régression relativement élevée {R^2} et faibles erreurs {ERAM, EQM et Q^2_{LOO}}.

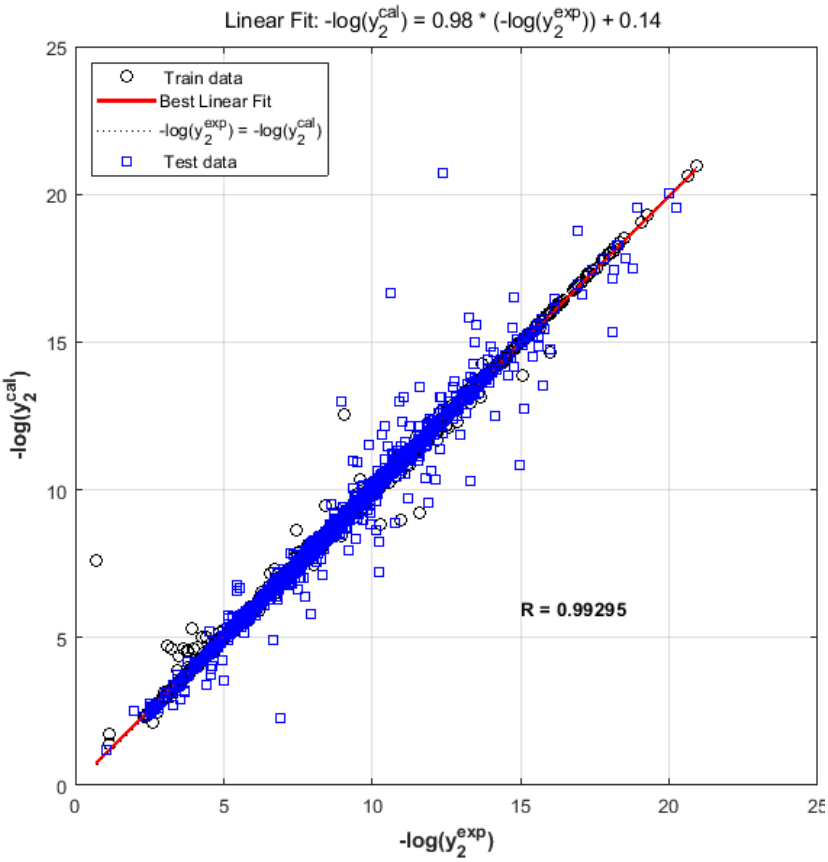


Figure V.3. Comparaison entre les valeurs expérimentales et prédites de la solubilité de 145 solutés dans CO_2-SC à l'aide du modèle SVM.

Sur la figure (V.3), la capacité du modèle a été évaluée en traçant les valeurs expérimentales de solubilité par rapport aux valeurs prédites par le modèle pour tous les ensembles de données. Le coefficient de corrélation élevé 0.9857 et une petite erreur relatif absolu moyen de 1.345 révèle la capacité du modèle à corréler la solubilité des solutés solides dans CO_2-SC en connaissant leurs propriétés physiques.

V.3.3. Comparaison

L'étude comparative entre les modèles RNA-QSPR et SVM-QSPR en termes de régression et de différentes erreurs en utilisant tous les ensembles de données a été menée dans cette section (Fig. V.4). Les résultats montrent que les données de solubilité des solutés solides dans CO_2-SC sont mieux corrélées par le modèle SVM-QSPR qu'avec le modèle RNA-QSPR. Les analyses statistiques montrent que la prédiction par le modèle machine à vecteurs de support (SVM) ont un excellent accord avec l'ensemble de données expérimentales.

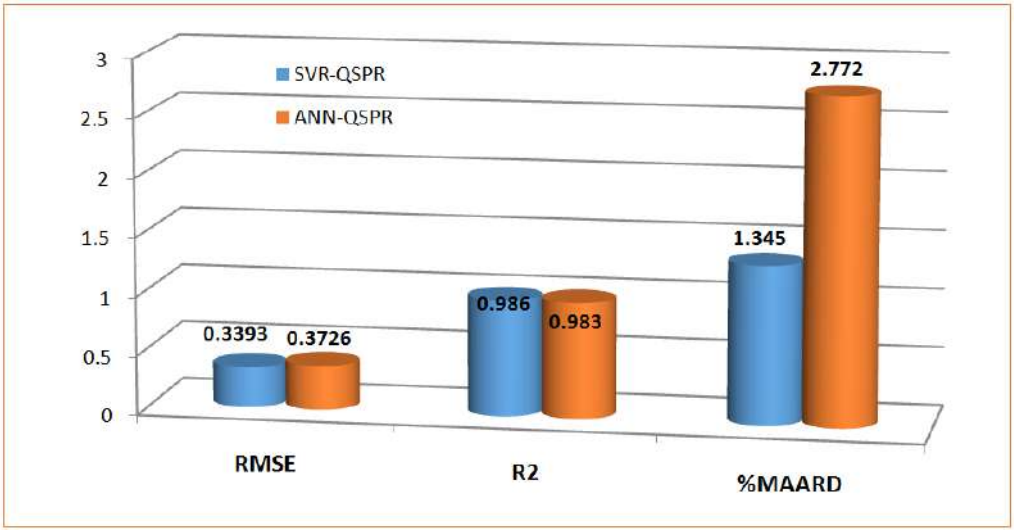


Figure V.4. Comparaison statistique entre le modèle RNA et SVM.

Tableau V.8. Comparaison entre les modèles RNA-SVM et SVM-QSPR développés et d'autres modèles intelligents rapportés dans la littérature.

Études	Méthodes	Nombre de composés	Données expérimentales	R^2	ERAM%
Ce Travail	**SVM-QSPR**	**145**	**3637**	**0.9857**	**1.3450**
Ce Travail	**RNA-QSPR**	**145**	**3637**	**0.9830**	**2.7720**
Baghban 2018 [7]	**LSSVR**	33	1162	0.9975	5.6100
Sodeifian 2018 [210]	**RNA**	6	155	0.91999	5.9300
Aminian 2017 [211]	**RNA**	20	439	0.9955	5.4200
Dadkhah 2017 [63]	**ANFIS**	29	795	0.9830	4.6617
Bian 2017 [6]	**GWO-SVR**	18	1148	0.9936	3.20
Vaferi 2013 [212]	**RNA**	8	198	0.99699	4.99
Eslamimanesh 2011 [213]	**FFNN**	21	795	0.9533	14.000

En plus de la validation statistique des modèles développés, une évaluation plus approfondie de la précision et de la régression obtenue avec les articles les plus pertinents développés par d'autres chercheurs (tableau V.8 et figure V.5). Malgré de nombreux articles publiés précédemment dans la littérature, les modèles développés dans cette étude ont montré de meilleures performances dans la corrélation de la solubilité des solutés solides dans CO_2-SC sur la base de certains descripteurs pertinents. Les modèles obtenus ont permis de corréler 145 systèmes (3637 points de données) avec les erreurs les plus faibles. Le modèle SVM-QSPR a été obtenu avec des paramètres statistiques très acceptables tels que R^2 = 0.9857, EQM = 0.3393 et ERAM = 0.9859% comme on peut le voir dans le tableau (V.8). Cependant, la comparaison globale en termes de systèmes corrélés indique que le modèle proposé fonctionne 145 systèmes avec de faibles erreurs et un coefficient de collecte élevé contrairement aux autres corrélations illustrées à la figure (V.5).

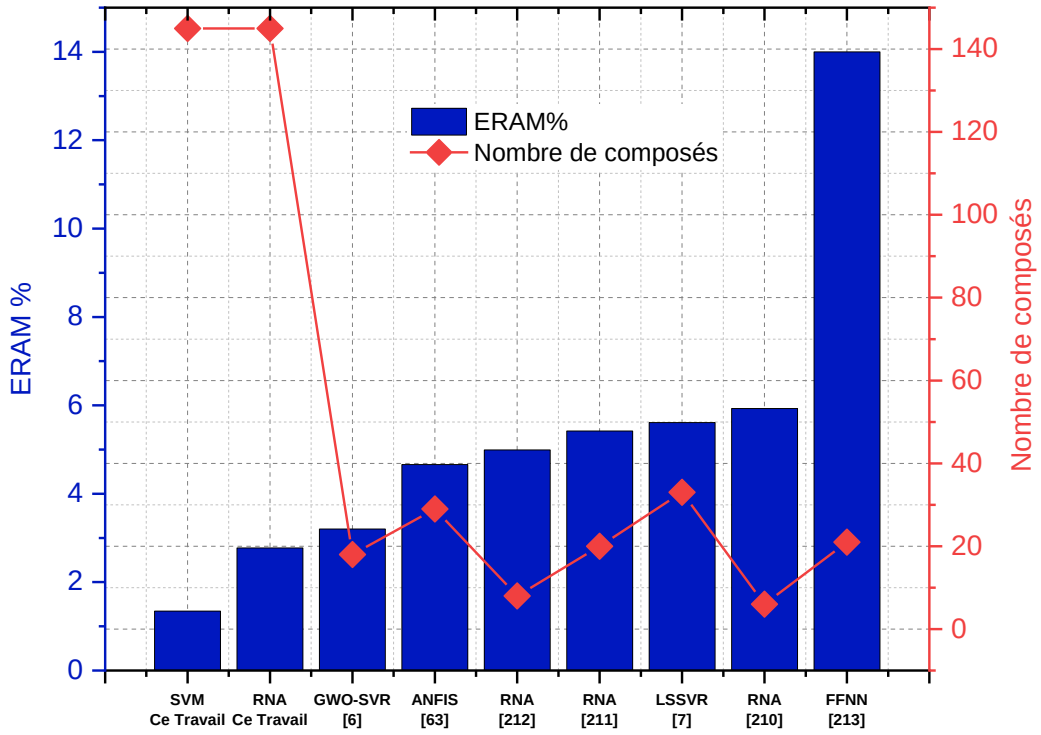


Figure V.5. Comparaison entre les modèles développés dans cette étude et ceux précédemment publiés dans la littérature en termes de (ERAM%) et nombre de composés.

La performance de notre modèle (SVM-QSPR) a de nouveau été comparée à six modèles basés sur la densité précédemment publiés (Tableau V.9.).

Tableau V.9. Modèles basés sur la densité pour la corrélation des solutés solides dans CO_2-SC.

Modèles	Equations	Références
Chrastil 1982	$ln(y_2) = a_0 + a_1 ln(\rho_1)$	[33]
Mèndez-Santiago and Teja (MST) 1999	$ln(y_2) = a_0 + \frac{a_1}{T} + a_2 P^2 + \left(a_3 + \frac{a_4}{T}\right) ln(\rho_1)$	[214]
Hezave and Lashkarbolooki 2013	$ln\left(\frac{P(y_2 - dxT)}{P^{ref}}\right) = a + \frac{b}{T} + c\,(\rho_1 - \rho_{ref})$	[181]
Keshmiri et al. 2014	$ln(y_2) = a_0 + \frac{a_1}{T} + a_2 P^2 + \left(a_3 + \frac{a_4}{T}\right) ln(\rho_1)$	[47]
Bian et al. 2016	$ln(y_2) = a_0 + \frac{a_1}{T} + \frac{a_2 \rho_1}{T} + (a_3 + a_4 \rho_1) ln(\rho_1)$	[178]
Si Moussa et al. 2018	$ln(y_2) = a_0 + a_1\rho_1 + a_2{\rho_1}^2 + a_3\rho_1 T + a_4 T + a_5 T^2 + a_6 \ln(\rho_1) + \frac{a_7}{T}$	[3]

Les paramètres des modèles de littérature sélectionnés ont été ajustés à l'aide d'algorithmes génétiques (fonction ga MATLAB). La comparaison globale de 145 composés calculée pour chacune des 07 corrélations étudiées.

Les résultats de corrélation des données expérimentales de cette étude sont présentés sous la forme d'un tableau comparatif de (ERAM%) et R^2 sont regroupés dans le tableau (V.10) et la figure (V.6).

Tableau V.10. Comparaison globale pour chacun des 06 modèles basés sur la densité et notre modèle SVM-QSPR.

Erreurs	Chrastil (1982) [33]	Mèndez-Santiago and Teja (1999) [214]	Hezave and Lashkarbolooki (2013) [181]	Keshmiri (2014) [47]	Bian (2016) [178]	Si Moussa (2018) [3]	Notre Modèle SVM-QSPR
ERAM %	19.3312	22.4011	13.0631	12.5889	11.4792	8.6026	**1.345**
R^2	0.9664	0.9727	0.9740	0.9820	0.9828	0.9848	**0.9857**

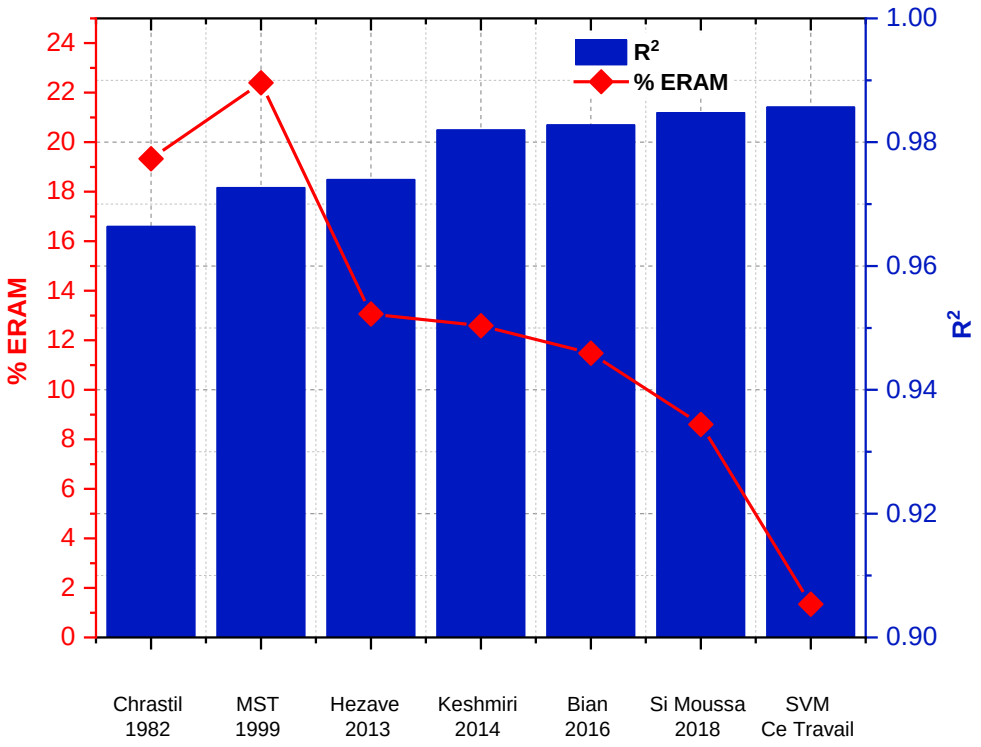


Figure V.6. Comparaison graphique globale pour chacun des 06 modèles basés sur la densité et notre modèle SVM-QSPR

Partager les résultats obtenus par le modèle SVM avec les utilisateurs peut être intéressant en utilisant une interface simple. La figure (V.7), une interface utilisateur graphique conviviale basée sur le logiciel Matlab et utilisant le modèle SVM formé a été développée pour évaluer la solubilité de 145 solutés solides dans CO_2-SC.

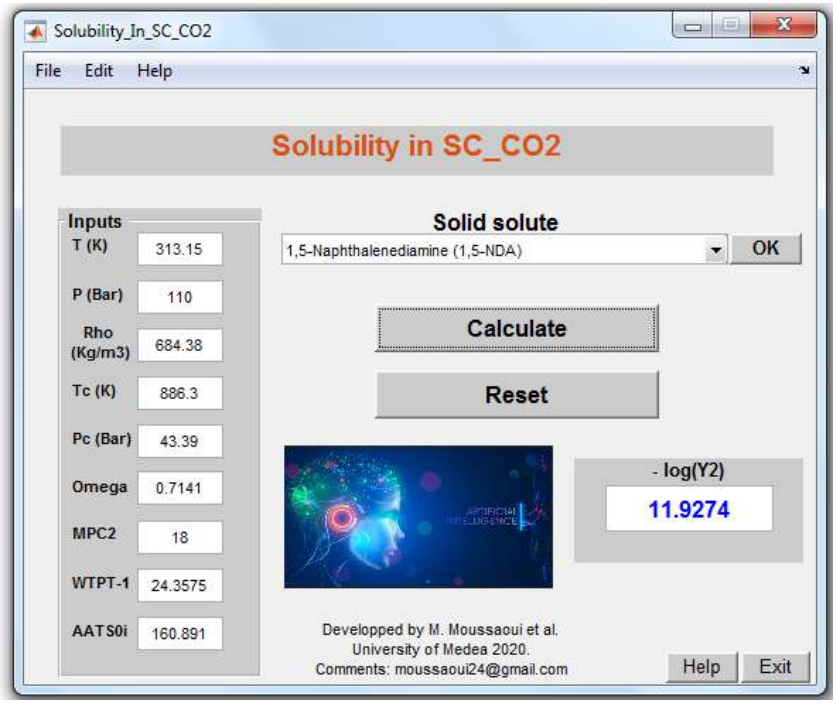


Figure V.7. Interface graphique pour évaluer la solubilité dans CO_2-SC à l'aide d'un modèle SVM.

V.3.4. Analyse de sensibilité globale

L'objectif de l'analyse de sensibilité globale est une technique mathématique appliquée pour trouver quelle variable d'entrée est considérée comme la plus importante pour la prédiction des valeurs de solubilité [102,103]. Dans ce travail, la méthode d'amplitude cosinus (CAM) a été utilisée pour évaluer l'effet de chaque entrée sur la sortie [101]. Les paires de données utilisées construisent un tableau de données X qui sont définis comme $X = \{X_1, X_2, X_3,...,X_m\}$ où X_i est un vecteur de longueurs m, exprimé comme: $X_i = \{x_{i1},x_{i2},x_{i3},...,x_{im}\}$. Les valeurs r_{ij} sont définies comme les relations de force entre les paramètres de sortie x_j et d'entrée x_i peuvent être calculées à l'aide de l'équation (V.1).

$$r_{ij} = \frac{\sum_{k=1}^{m} x_{ik}x_{jk}}{\sqrt{\sum_{k=1}^{m} x_{ij}^2 \sum_{k=1}^{m} x_{jk}^2}} \qquad \text{(V.1)}$$

Où ***i*** et **k** sont les dimensions de la matrice d'entrée (***i*** = 1 à 3637, ***k*** = 1 à ***m***), ***j*** est les dimensions du vecteur de sortie (***j*** = 1 à 3637) et ***m*** est le nombre de paramètres d'entrée (***m*** = 9).

L'importance des paramètres d'entrée dans SVM-QSPR pour la solubilité du soluté solide dans CO_2-SC est illustrée à la figure (V.8). Comme observé, tous les paramètres ont approximativement les mêmes effets. Les résultats prouvent que tous les paramètres d'entrée désignés dans cette étude ont des effets cruciaux sur la solubilité du soluté solide dans CO_2-SC, par conséquent, ils ont été sélectionnés de manière appropriée.

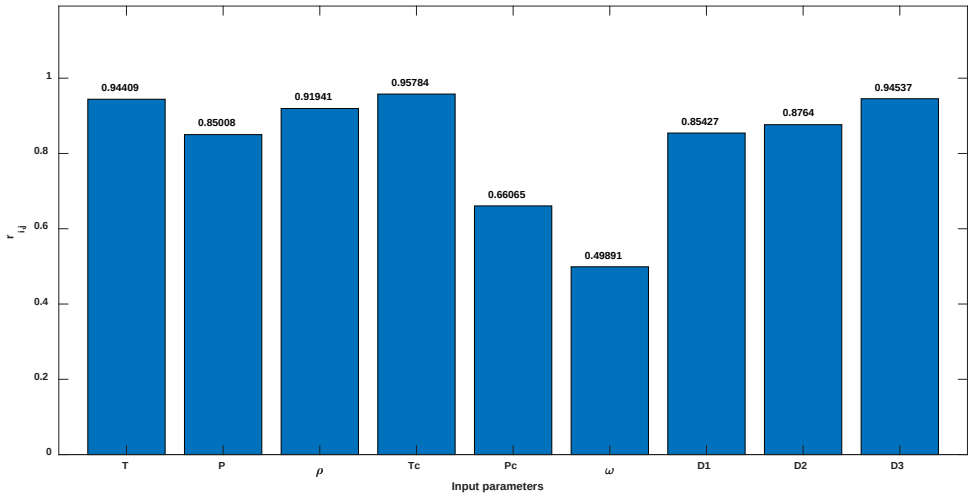
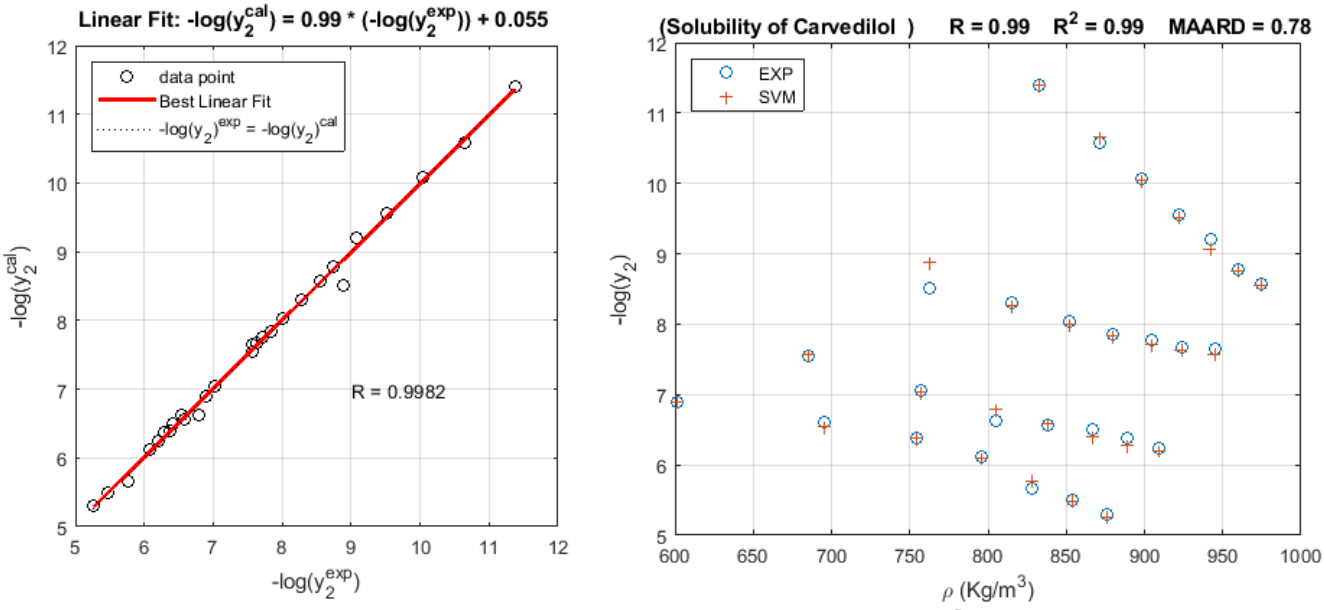


Figure V.8. Analyse de sensibilité entre les variables d'entrée et la sortie calculée.

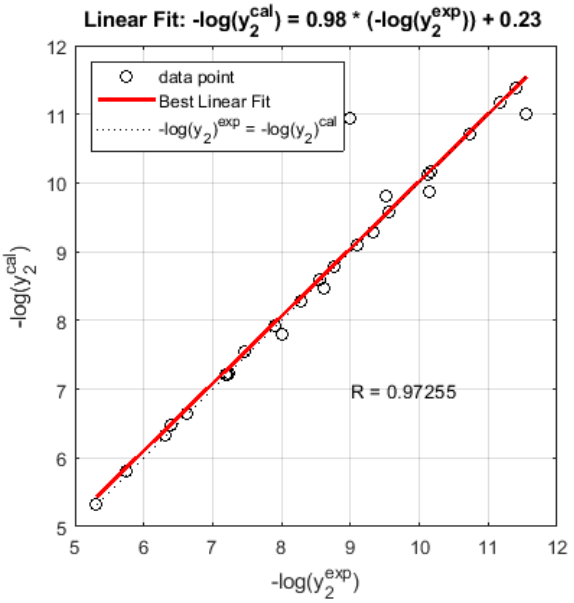
V.3.5. Capacité d'extrapolation

Dans cette section, la capacité d'extrapolation du modèle proposé a été discutée en utilisant les mesures expérimentales rapportées dans les articles suivants [215–222]. Les valeurs de solubilité prédites par le modèle SVM-QSPR ont été comparées aux valeurs expérimentales de huit systèmes. Les résultats de la comparaison illustrés dans les figures (V.9-V.16) de solubilité en fonction de la densité ont indiqué que le modèle fonctionnait bien avec une bonne qualité d'ajustement R. Ces résultats ont montré clairement que la performance du modèle de prévision basé sur SVM-QSPR est assez satisfaisante.

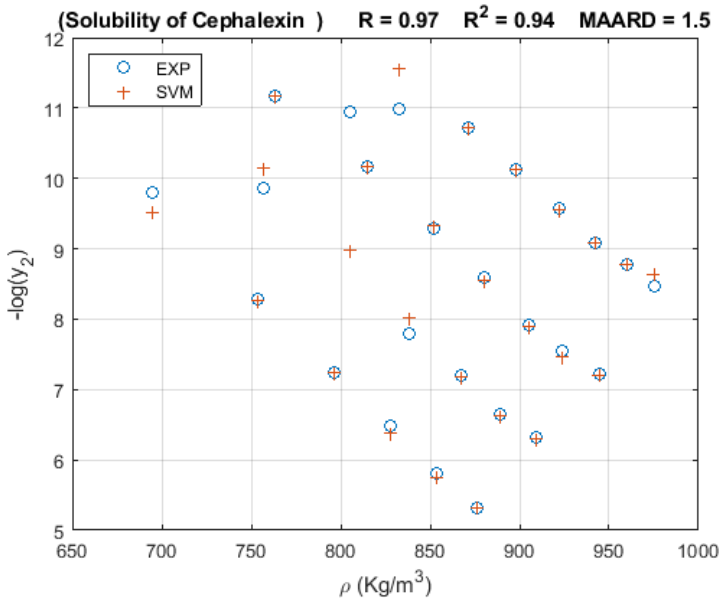


(a) **(b)**

Figure V.9. a: Les valeurs de solubilité prédite par rapport aux valeurs expérimentales pour le système Carvedilol, **b:** Comparaison entre données expérimentales [215] et SVM-QSPR résultats prévus pour le système Carvedilol par rapport à la densité de CO_2-SC.

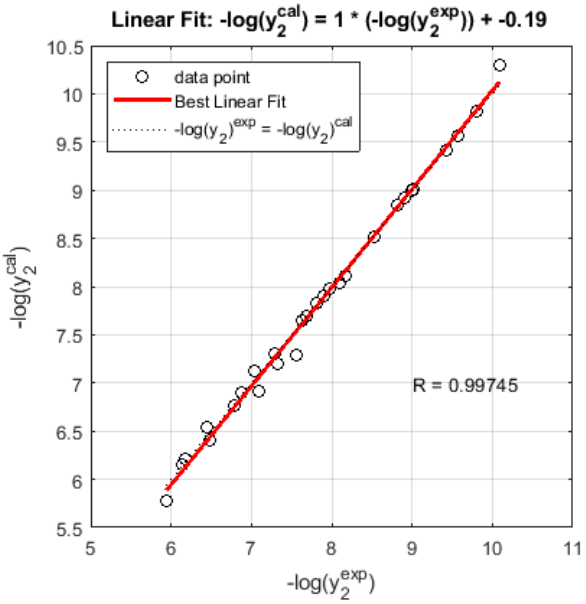


(a)

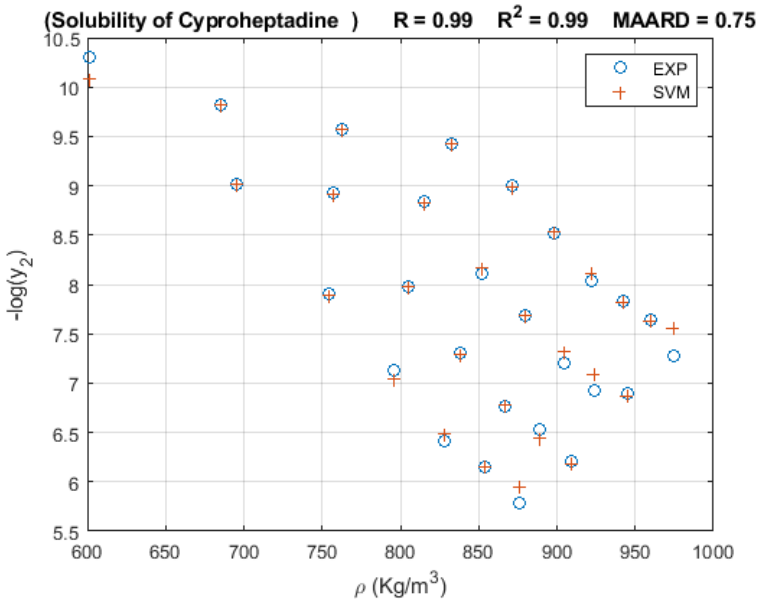


(b)

Figure V.10. a: Les valeurs de solubilité prédite par rapport aux valeurs expérimentales pour le système Cephalexin, **b:** Comparaison entre données expérimentales [216] et SVM-QSPR résultats prévus pour le système Cephalexin par rapport à la densité de CO_2-SC

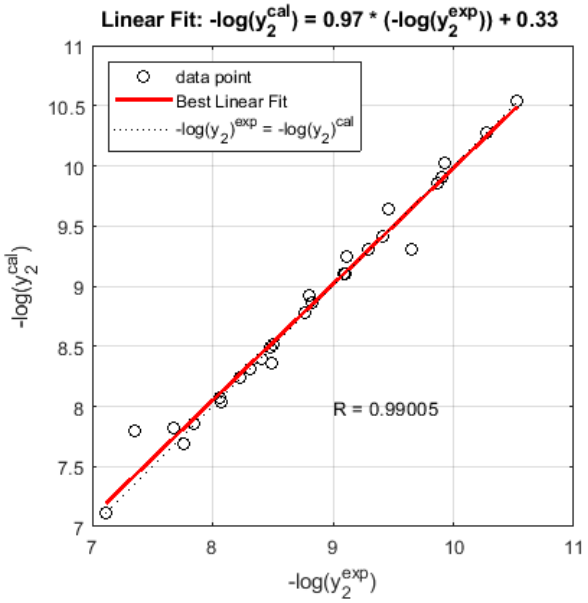


(a)

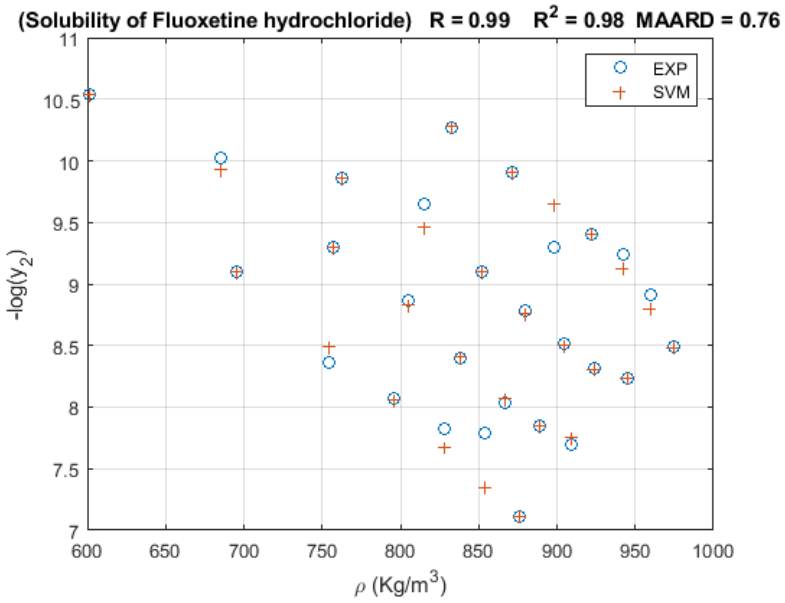


(b)

Figure V.11. a: Les valeurs de solubilité prédite par rapport aux valeurs expérimentales pour le système Cyproheptadine, **b:** Comparaison entre données expérimentales [217] et SVM-QSPR résultats prévus pour le système Cyproheptadine par rapport à la densité de CO_2-SC.

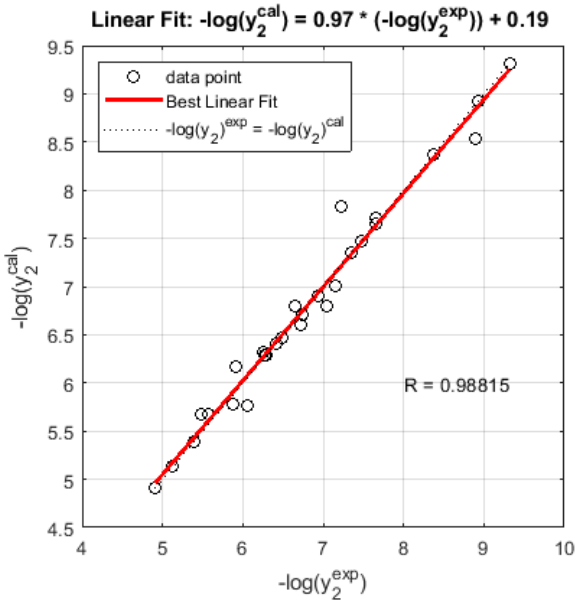


(a)

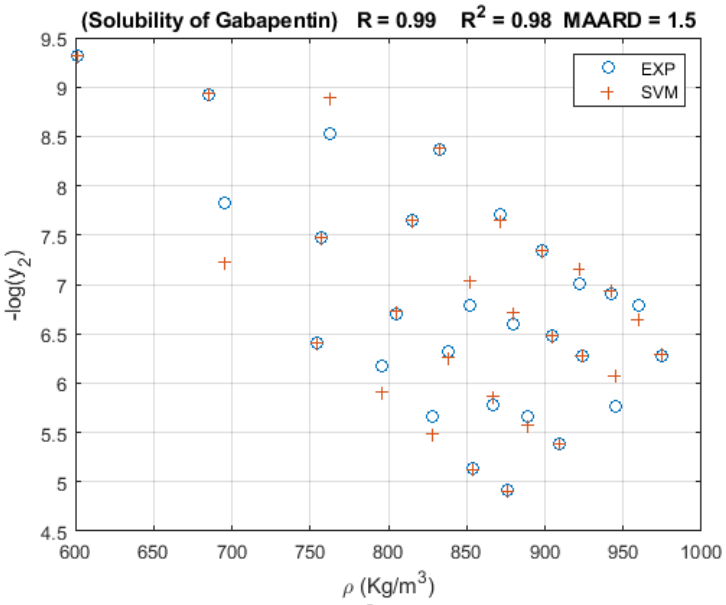


(b)

Figure V.12. a: Les valeurs de solubilité prédite par rapport aux valeurs expérimentales pour le système hydrochloride, **b:** Comparaison entre données expérimentales [218] et SVM-QSPR résultats prévus pour le système hydrochloride par rapport à la densité de CO_2-SC

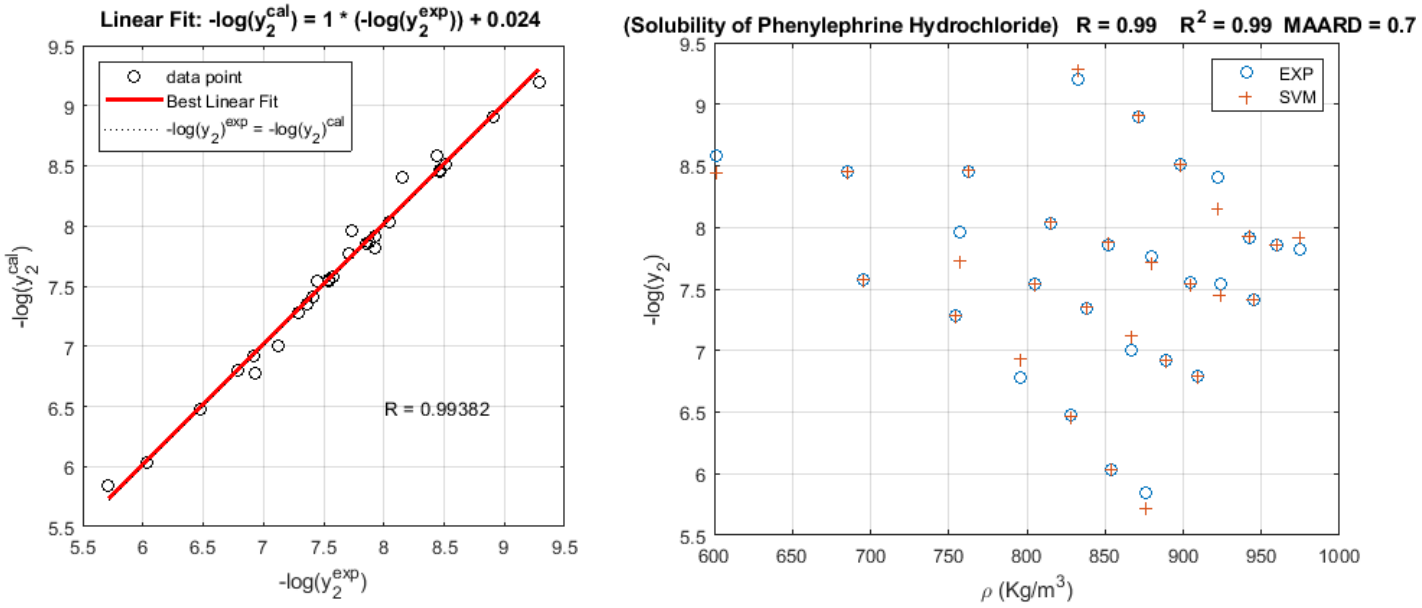


(a)



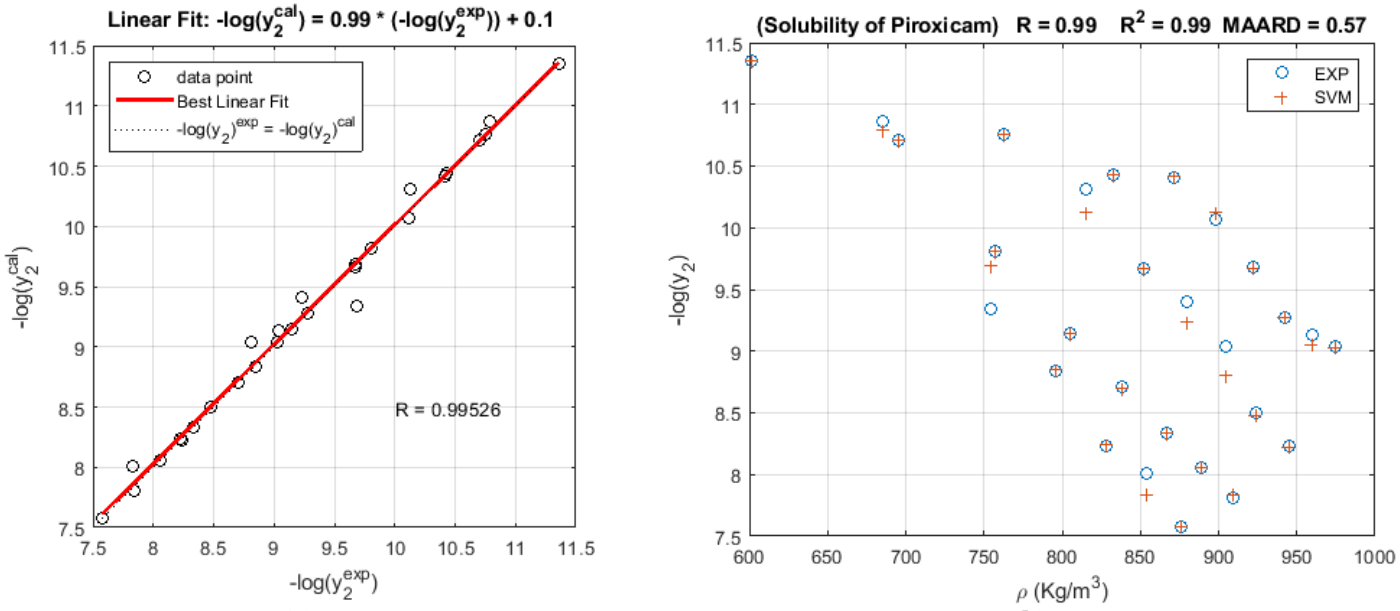
(b)

Figure V.13. a: Les valeurs de solubilité prédite par rapport aux valeurs expérimentales pour le système Gabapentin, **b:** Comparaison entre données expérimentales [219] et SVM-QSPR résultats prévus pour le système Gabapentin par rapport à la densité de CO_2-SC.



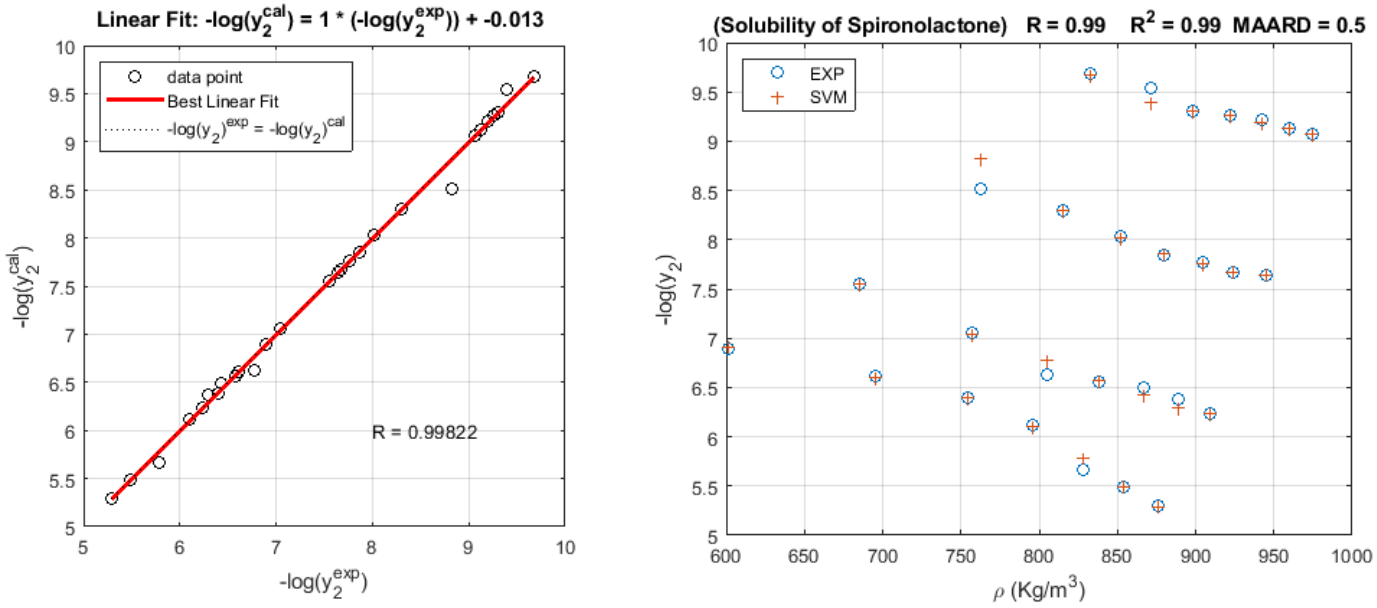
a b

Figure V. 14. a: Les valeurs de solubilité prédite par rapport aux valeurs expérimentales pour le système Phenylephrine Hydrochloride, **b:** Comparaison entre données expérimentales [220] et SVM-QSPR résultats prévus pour le système Phenylephrine Hydrochloride par rapport à la densité de CO_2-SC.



(a) **(b)**

Figure V.15. a: Les valeurs de solubilité prédite par rapport aux valeurs expérimentales pour le système Piroxicam, **b:** Comparaison entre données expérimentales [221] et SVM-QSPR résultats prévus pour le système Piroxicam par rapport à la densité de CO_2-SC.



(a) **(b)**

Figure V.16. a: Les valeurs de solubilité prédite par rapport aux valeurs expérimentales pour le système Spironolactone, **b:** Comparaison entre données expérimentales [222] et SVM-QSPR résultats prévus pour le système Spironolactone par rapport à la densité de CO_2-SC.

V.4. Conclusion

Dans cette contribution, une comparaison entre la modélisation par la machine à vecteurs de support (SVM) et le réseau neuronal artificiel (RNA) pour prédire la solubilité de 145 composés dans le dioxyde de carbone supercritique. Sur la base des paramètres statistiques calculés, les résultats montrent clairement que l'approche SVM-QSPR avec erreur relative absolue moyenne ERAM=1.345% et coefficient de déterminant R^2=0.9857 et est plus précis par rapport à RNA-QSPR avec ERAM=2.7720% et R^2=0.9830.

Les résultats de la prédiction de modèle SVM-QSPR ont été comparés, d'une part, avec les données expérimentales et, d'autre part, avec les résultats obtenus dans la littérature. Ces comparaisons montrent que le modèle SVM-QSPR proposé est de loin le meilleur et possède une très bonne tendance de prédiction, d'interpolation et parfois même d'extrapolation.

Conclusion générale

Conclusion générale

Depuis une dizaine d'années, l'intelligence computationnelle a développée de nombreuses applications dans des secteurs très variés. Le passage en revue de la littérature relative à la corrélation des données de solubilité des solutés solides dans le CO_2-SC montre qu'il existe trois approches : les équations d'état, les modèles semi-empiriques, et les méthodes d'intelligence computationnelle (Réseaux de Neurones Artificiels, les Machines à Vecteurs de Support, système d'inférence neuro-flou adaptatif, …).

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. Les méthodes d'intelligence computationnelle sont des puissants modèles informatiques capables de capturer et de représenter des relations d'entrée/sortie complexes. La motivation pour le développement de la technique d'intelligence computationnelle découle du désir de développer un système artificiel capable de réaliser des tâches "intelligentes" semblables à celles réalisées par le cerveau humain.

L'objectif principal de ce travail est la modélisation des propriétés thermochimiques telles que les propriétés critiques et la solubilité des solutés solides dans les fluides supercritiques.

Dans la première contribution nous avons modélisé les propriétés critiques telles que la pression critique, la température critique, le volumique critique et le facteur acentrique de 6700 composés purs à partir de 5 descripteurs moléculaires pertinents et de 2 propriétés thermodynamiques. Pour cela, quatre méthodes ont été utilisées à savoir la régression linéaire multiple (MLR), les réseaux de neurones artificiels (RNA), les machines à vecteurs de support (SVM) et les machines à vecteurs de support hybride avec l'algorithme d'optimisation Dragonfly (SVM-DA). Les résultats montrent que le SVM-DA a montré de meilleures performances de prédiction par rapport aux autres modèles en termes d'erreur relative absolue moyenne (ERAM%) de {0.7551, 1.962, 1.929 et 2.173} et le coefficient de détermination (R^2) de {0.9699, 0.9673, 0.9856 et 0.9766} pour la température critique, la pression critique, le volume critique et le facteur acentrique respectivement. Il est conclu que le modèle SVM-DA est capable de prédire les propriétés critiques de composés purs avec une grande précision.

Dans la deuxième 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 130 composés totalisant un ensemble de 3269 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 (ERAM = 7.42 %) et un coefficient de détermination (R^2 = 0.997). Les résultats du travail effectué dans cette contribution nous ont permis, en outre, de mieux comprendre quelques aspects de la modélisation semi-empirique de la solubilité des solutés solides dans le CO_2 -SC.

- Le choix de la base de données est primordial lorsqu'il s'agit de comparer les modèles semi-empiriques. Plus la base de données est large et complète plus les résultats de comparaison sont significatifs.
- La taille de la base de données est très importante dans une étude comparative, et aussi le choix de modèle qui aura des modifications est plus important pour la proposition d'un nouveau modèle
- D'une manière générale les résultats obtenus dans cette étude confirme les résultats des études antérieures en ce sens qu'il n'existe pas de modèle qui donne la meilleure corrélation pour tous les composés étudiés.

Dans la troisième contribution la solubilité de 145 solutés solides dans le CO_2 supercritique (CO_2-SC) a été modélisée à l'aide de techniques d'intelligence computationnelle basées sur des modèles de relation quantitative structure-propriété (QSPR). Une base de données de 3637 valeurs de solubilité a été recueillie à partir d'articles publiés antérieurement. Le logiciel Dragon a été utilisé pour calculer les descripteurs moléculaires de 145 systèmes solides. L'algorithme génétique (GA) a été mis en œuvre pour optimiser le sous-ensemble des descripteurs significativement contribués.

Cette contribution a montré que l'utilisation de combinaison entre les données thermodynamique et les descripteurs moléculaires comme des entrées des modèles a donné une performance de corrélation acceptable pour les deux techniques d'intelligence

computationnelle considérées. Avec erreur relative absolue moyenne globale (ERAM) d'environ 1.345% entre les valeurs expérimentales et calculées par le modèle de support vecteur régression (SVM-QSPR) qui est meilleure que celle obtenue à l'aide du modèle RNA-QSPR de 2.772%. Les résultats montrent que le modèle SVM-QSPR développé est plus précis et peut être utilisé comme un outil de modélisation puissant alternatif pour les études QSPR de la solubilité des solutés solides dans le dioxyde de carbone supercritique (CO_2-SC).

En perspective, tout en estimant avoir atteint notre objectif tracé pour cette étude, nous envisagerons :

- Appliquer d'autre méthodes d'intelligence computationnelle, telles que le "Deep Learning" dans le but de prédiction/corrélation de la solubilité des solutés solides dans les fluides supercritiques;
- Modéliser la solubilité dans les différents solvants tels que les liquides ionique;
- Appliquer la modélisation par l'intelligence computationnelle à d'autres phénomènes chimiques et thermiques dans l'industrie.

Références bibliographiques

Références bibliographiques

[1] J.A.P. Coelho, R.M. Filipe, G.P. Naydenova, D.S. Yankov, R.P. Stateva, Semi-empirical models and a cubic equation of state for correlation of solids solubility in scCO2: Dyes and calix [4] arenes as illustrative examples, Fluid Phase Equilib. 426 (2016) 37–46.

[2] C. Si Moussa, Modélisation des équilibres de phases à pressions élevées par les réseaux de neurones artificiels, Thèse de Doctorat, Ecole Nationale Polytechnique, Algerie, 2010.

[3] A. Belghait, C. Si-Moussa, M. Laidi, S. Hanini, Corrélation semi-empirique de solubilité des solutés solides dans le dioxyde de carbone supercritique: étude comparative et proposition d'un nouveau modèle basé sur la densité, Comptes Rendus Chim. 21 (2018) 494–513. doi:10.1016/j.crci.2018.02.006.

[4] A. Abdallah el hadj, M. Laidi, C. Si-Moussa, S. Hanini, Novel approach for estimating solubility of solid drugs in supercritical carbon dioxide and critical properties using direct and inverse artificial neural network (ANN), Neural Comput. Appl. 28 (2017) 87–99. doi:10.1007/s00521-015-2038-1.

[5] F. Gharagheizi, A. Eslamimanesh, A.H. Mohammadi, D. Richon, Determination of critical properties and acentric factors of pure compounds using the artificial neural network group contribution algorithm, J. Chem. Eng. Data. 56 (2011) 2460–2476. doi:10.1021/je200019g.

[6] X.Q. Bian, Q. Zhang, L. Zhang, J. Chen, A grey wolf optimizer-based support vector machine for the solubility of aromatic compounds in supercritical carbon dioxide, Chem. Eng. Res. Des. 123 (2017) 284–294. doi:10.1016/j.cherd.2017.05.008.

[7] A. Baghban, A. Jalali, A.H. Mohammadi, S. Habibzadeh, Efficient modeling of drug solubility in supercritical carbon dioxide, J. Supercrit. Fluids. 133 (2018) 466–478. doi:10.1016/j.supflu.2017.10.032.

[8] C.C. de la Tour, Nouvelle note sur les effets qu'on obtient par l'application simultanée de la chaleur et de la compression a certains liquides, Ann. Chim. Phys. 22 (1823) 410–415.

[9] A. Idrissi, E. Cinar, S. Longelin, P. Damay, The effect of temperature on urea--urea interactions in water: a molecular dynamics simulation, J. Mol. Liq. 110 (2004) 201–208.

[10] C. Mathieu, Le dioxyde de carbone supercritique: un fluide prometteur dans la formulation pharmaceutique, These de doctorat, Université de Lorraine, 2012.

[11] R.S. Oakes, A.A. Clifford, C.M. Rayner, The use of supercritical fluids in synthetic organic chemistry, J. Chem. Soc. Perkin Trans. 1. (2001) 917–941.

[12] D.L. Tomasko, H. Li, D. Liu, X. Han, M.J. Wingert, L.J. Lee, K.W. Koelling, A review of CO2 applications in the processing of polymers, Ind. Eng. Chem. Res. 42 (2003) 6431–6456.

[13] M.F. Kemmere, T. Meyer, Supercritical carbon dioxide: in polymer reaction engineering, John Wiley & Sons, 2006.

[14] W. Francis, W. Francis, Ternary systems of liquid carbon dioxide, J. Phys. Chem. 58 (1954).

[15] H.-D. Sung, J.-J. Shim, Solubility of CI disperse red 60 and CI disperse blue 60 in supercritical carbon dioxide, J. Chem. Eng. Data. 44 (1999) 985–989.

[16] C. Crampon, G. Charbit, E. Neau, High-pressure apparatus for phase equilibria studies: solubility of fatty acid esters in supercritical CO2, J. Supercrit. Fluids. 16 (1999) 11–20.

[17] G. Sherman, S. Shenoy, R.A. Weiss, C. Erkey, A static method coupled with gravimetric analysis for the determination of solubilities of solids in supercritical carbon dioxide, Ind. Eng. Chem. Res. 39 (2000) 846–848.

[18] A.A. Isse, A. Galia, C. Belfiore, G. Silvestri, A. Gennaro, Electrochemical reduction and carboxylation of halobenzophenones, J. Electroanal. Chem. 526 (2002) 41–52.

[19] S. Bristow, B.Y. Shekunov, P. York, Solubility analysis of drug compounds in supercritical carbon dioxide using static and dynamic extraction systems, Ind. Eng. Chem. Res. 40 (2001) 1732–1739.

[20] H. Bezaze, Modélisation thermodynamique des équilibres de phase à haute préssion, Thèse de doctorat, Université Mentouri -Constantine, 2012.

[21] J. Vidal, Thermodynamique: application au génie chimique et à l'industrie pétrolière, Editions Technip,

1997.

[22] B.E. Poling, J.M. Prausnitz, The Properties Of Gases And Liquids, The McGraw, 2001.

[23] S.I. Sandler, Models for thermodynamic and phase equilibria calculations, Dekker, 1994.

[24] B.E. Poling, J.M. Prausnitz, J.P. O'connell, The properties of gases and liquids, The McGraw, 2001.

[25] F. Troadec, Simulation numérique directe d'un écoulement supercritique pour validation des approches RANS et LES, Rouen, 2010.

[26] Y.S. Wei, R.J. Sadus, Equations of state for the calculation of fluid-phase equilibria, AIChE J. 46 (2000) 169–196.

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

[28] G.M. Kontogeorgis, E.C. Voutsas, I. V Yakoumis, D.P. Tassios, An equation of state for associating fluids, Ind. Eng. Chem. Res. 35 (1996) 4310–4318.

[29] J. Gross, G. Sadowski, Perturbed-chain SAFT: An equation of state based on a perturbation theory for chain molecules, Ind. Eng. Chem. Res. 40 (2001) 1244–1260.

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

[31] A. Abdallah El Hadj, C. Si-Moussa, S. Hanini, M. Laidi, Application of PC-SAFT and cubic equations of state for the correlation of solubility of some pharmaceutical and statin drugs in SC-CO2, Chem. Ind. Chem. Eng. Quarterly/CICEQ. 19 (2013) 449–460.

[32] E. Stahl, W. Schilz, E. Schütz, E. Willing, A quick method for the microanalytical evaluation of the dissolving power of supercritical gases, Angew. Chemie Int. Ed. English. 17 (1978) 731–738.

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

[34] Y. Adachi, B.C.-Y. Lu, Supercritical fluid extraction with carbon dioxide and ethylene, Fluid Phase Equilib. 14 (1983) 147–156.

[35] J.M. Del Valle, J.M. Aguilera, An improved equation for predicting the solubility of vegetable oils in supercritical carbon dioxide, Ind. Eng. Chem. Res. 27 (1988) 1551–1553.

[36] S.K. Kumar, K.P. Johnston, Modelling the solubility of solids in supercritical fluids with density as the independent variable, J. Supercrit. Fluids. 1 (1988) 15–22.

[37] K.D. Bartle, A.A. Clifford, S.A. Jafar, G.F. Shilstone, Solubilities of solids and liquids of low volatility in supercritical carbon dioxide, J. Phys. Chem. Ref. Data. 20 (1991) 713–756.

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

[39] J. Méndez-Santiago, A.S. Teja, The solubility of solids in supercritical fluids, Fluid Phase Equilib. 158 (1999) 501–510.

[40] A. Jouyban, H.-K. Chan, N.R. Foster, Mathematical representation of solute solubility in supercritical carbon dioxide using empirical expressions, J. Supercrit. Fluids. 24 (2002) 19–35.

[41] D.L. Sparks, R. Hernandez, L.A. Estévez, Evaluation of density-based models for the solubility of solids in supercritical carbon dioxide and formulation of a new model, Chem. Eng. Sci. 63 (2008) 4292–4301. doi:10.1016/j.ces.2008.05.031.

[42] C. Garlapati, G. Madras, Solubilities of solids in supercritical fluids using dimensionally consistent modified solvate complex models, Fluid Phase Equilib. 283 (2009) 97–101.

[43] C. Garlapati, G. Madras, New empirical expressions to correlate solubilities of solids in supercritical carbon dioxide, Thermochim. Acta. 500 (2010) 123–127.

[44] S.J. Nejad, H. Abolghasemi, M.A. Moosavian, M.G. Maragheh, Prediction of solute solubility in supercritical carbon dioxide: A novel semi-empirical model, Chem. Eng. Res. Des. 88 (2010) 893–898.

[45] R. Ch, G. Madras, An association model for the solubilities of pharmaceuticals in supercritical carbon dioxide, Thermochim. Acta. 507 (2010) 99–105.

[46] X. Bian, Z. Du, Y. Tang, An improved density-based model for the solubility of some compounds in supercritical carbon dioxide, Thermochim. Acta. 519 (2011) 16–21.

[47] K. Keshmiri, A. Vatanara, Y. Yamini, Development and evaluation of a new semi-empirical model for correlation of drug solubility in supercritical CO2, Fluid Phase Equilib. 363 (2014) 18–26. doi:10.1016/j.fluid.2013.11.013.

[48] A.A. Amooey, A simple correlation to predict drug solubility in supercritical carbon dioxide, Fluid Phase Equilib. 375 (2014) 332–339.

[49] S.B. Hozhabr, S.H. Mazloumi, J. Sargolzaei, Correlation of solute solubility in supercritical carbon dioxide using a new empirical equation, Chem. Eng. Res. Des. 92 (2014) 2734–2739.

[50] M.A. Khansary, F. Amiri, A. Hosseini, A.H. Sani, H. Shahbeig, Representing solute solubility in supercritical carbon dioxide: A novel empirical model, Chem. Eng. Res. Des. 93 (2015) 355–365.

[51] X.-Q. Bian, Q. Zhang, Z.-M. Du, J. Chen, J.-N. Jaubert, A five-parameter empirical model for correlating the solubility of solid compounds in supercritical carbon dioxide, Fluid Phase Equilib. 411 (2016) 74–80.

[52] A. KhazaiePoul, M. Soleimani, S. Salahi, Solubility prediction of disperse dyes in supercritical carbon dioxide and ethanol as co-solvent using neural network, Chinese J. Chem. Eng. 24 (2016) 491–498.

[53] A. Aminian, Estimating the solubility of different solutes in supercritical CO2 covering a wide range of operating conditions by using neural network models, J. Supercrit. Fluids. 125 (2017) 79–87.

[54] V. Vapnik, The support vector method of function estimation, in: Nonlinear Model., Springer, 1998: pp. 55–85.

[55] R. Li, H.-N. Wang, H. He, Y.-M. Cui, Z.-L. Du, Support vector machine combined with k-nearest neighbors for solar flare forecasting, Chinese J. Astron. Astrophys. 7 (2007) 441.

[56] J.A.K. Suykens, J. Vandewalle, Least squares support vector machine classifiers, Neural Process. Lett. 9 (1999) 293–300.

[57] J. De Brabanter, B. De Moor, J.A.K. Suykens, T. Van Gestel, J.P.L. Vandewalle, Least squares support vector machines, World scientific, 2002.

[58] B. Mehdizadeh, K. Movagharnejad, A comparative study between LS-SVM method and semi empirical equations for modeling the solubility of different solutes in supercritical carbon dioxide, Chem. Eng. Res. Des. 89 (2011) 2420–2427. doi:10.1016/j.cherd.2011.03.017.

[59] M. Duan, Z. Liu, D. Yan, W. Peng, A. Baghban, Application of lssvm algorithm for estimating higher heating value of biomass based on ultimate analysis, Energy Sources, Part A Recover. Util. Environ. Eff. 40 (2018) 709–715.

[60] F.-J. Chang, Y.-T. Chang, Adaptive neuro-fuzzy inference system for prediction of water level in reservoir, Adv. Water Resour. 29 (2006) 1–10.

[61] Y.-T. Chang, L.-C. Chang, F.-J. Chang, Intelligent control for modeling of real-time reservoir operation, part II: artificial neural network with operating rule curves, Hydrol. Process. An Int. J. 19 (2005) 1431–1444.

[62] H. Safari, S. Nekoeian, M.R. Shirdel, H. Ahmadi, A. Bahadori, S. Zendehboudi, Assessing the dynamic viscosity of Na--K--Ca--Cl--H2O aqueous solutions at high-pressure and high-temperature conditions, Ind. Eng. Chem. Res. 53 (2014) 11488–11500.

[63] M.R. Dadkhah, A. Tatar, A. Mohebbi, A. Barati-Harooni, A. Najafi-Marghmaleki, M.M. Ghiasi, A.H. Mohammadi, F. Pourfayaz, Prediction of solubility of solid compounds in supercritical CO2 using a connectionist smart technique, J. Supercrit. Fluids. 120 (2017) 181–190. doi:10.1016/j.supflu.2016.06.006.

[64] P. Willett, Chemoinformatics: a history, Wiley Interdiscip. Rev. Comput. Mol. Sci. 1 (2011) 46–56.

[65] J. Devillers, Methods for building QSARs, in: Comput. Toxicol., Springer, 2013: pp. 3–27.

[66] G. Poezevara, Fouille de graphes pour la découverte de contrastes entre classes: application à

l'estimation de la toxicité des molécules, Thèse de doctorat, Université de Caen, France, 2011.

[67] R. Todeschini, V. Consonni, Molecular descriptors for chemoinformatics: volume I: alphabetical listing/volume II: appendices, references, John Wiley & Sons, 2009.

[68] S. Haykin, Neural networks: a comprehensive foundation, Prentice-Hall, Inc., 2007.

[69] G. Dreyfus, Neural networks: methodology and applications, Springer Science & Business Media, 2005.

[70] A.P. Engelbrecht, Computational intelligence: an introduction, John Wiley & Sons, 2007.

[71] B. Widrow, M.E. Hoff, Adaptive switching circuits, Stanford U, 1960.

[72] F. Piéraut, Optimisation des réseaux de neurones de grande capacité, IEEE Trans. NNs. 20 (2002).

[73] S. Samarasinghe, Neural networks for applied sciences and engineering: from fundamentals to complex pattern recognition, Crc Press, 2016.

[74] J.A. Hertz, Introduction to the theory of neural computation, CRC Press, 2018.

[75] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning internal representations by error propagation, California, 1985.

[76] P. Gallinari, O. Gascuel, Statistique, apprentissage et généralisation. Application aux réseaux de neurones, Rev. d'intelligence Artif. 10 (1996) 285–343.

[77] B. Curry, P.H. Morgan, Model selection in neural networks: some difficulties, Eur. J. Oper. Res. 170 (2006) 567–577.

[78] A.P. Plumb, R.C. Rowe, P. York, M. Brown, Optimisation of the predictive ability of artificial neural network (ANN) models: a comparison of three ANN programs and four classes of training algorithm, Eur. J. Pharm. Sci. 25 (2005) 395–405.

[79] M.-J. Lee, C.-F. Kou, J.-W. Cheng, H. Lin, Vapor--liquid equilibria for binary mixtures of carbon dioxide with 1, 2-dimethoxybenzene, 2-methoxyphenol, or p-cresol at elevated pressures, Fluid Phase Equilib. 162 (1999) 211–224.

[80] D. Abooali, M.A. Sobati, Novel method for prediction of normal boiling point and enthalpy of vaporization at normal boiling point of pure refrigerants: A QSPR approach, Int. J. Refrig. 40 (2014) 282–293. doi:10.1016/j.ijrefrig.2013.12.007.

[81] B. Hemmateenejad, P. Ilani-kashkouli, Quantitative structure-property relationship study to predict speed of sound in diverse organic solvents from solvent structural information, Ind. Eng. Chem. Res. 51 (2012) 14884–14891. doi:10.1021/ie3016297.

[82] A. Khajeh, H. Modarress, QSPR prediction of surface tension of refrigerants from their molecular structures, Int. J. Refrig. 35 (2012) 150–159. doi:10.1016/j.ijrefrig.2011.08.007.

[83] A.R. Katritzky, M. Kuanar, S. Slavov, C.D. Hall, M. Karelson, I. Kahn, D.A. Dobchev, Quantitative correlation of physical and chemical properties with chemical structure: utility for prediction, Chem. Rev. 110 (2010) 5714–5789.

[84] J. Marrero, R. Gani, Group-contribution based estimation of pure component properties, Fluid Phase Equilib. 183–184 (2001) 183–208. doi:10.1016/S0378-3812(01)00431-9.

[85] M.A.S. Zahra Ghomisheh, Ali Ebrahimpoor Gorji, Prediction of critical properties of sulfur-containing compounds: New QSPR models, J. Mol. Graph. Model. (2019) 135577. doi:10.1016/j.scitotenv.2019.135577.

[86] C.L. Yaws, The yaws handbook of physical properties for hydrocarbons and chemicals: physical properties for more than 54,000 organic and inorganic chemical compounds, Coverage for C1 to C100 Organics and Ac to Zr Inorganics, Gulf Professional Publishing, 2015.

[87] A. Mauri, alvaDesc: A Tool to Calculate and Analyze Molecular Descriptors and Fingerprints, in: Ecotoxicological QSARs, Springer, 2020: pp. 801–820.

[88] S. Dong, Y. Zhang, Z. He, N. Deng, X. Yu, S. Yao, Investigation of support vector machine and back propagation artificial neural network for performance prediction of the organic rankine cycle system, Energy. 144 (2018) 851–864. doi:10.1016/j.energy.2017.12.094.

[89] B.E. Boser, I.M. Guyon, V.N. Vapnik, A training algorithm for optimal margin classifiers, in: Proc.

Fifth Annu. Work. Comput. Learn. Theory, 1992: pp. 144–152.

[90] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (1995) 273–297. doi:10.1007/BF00994018.

[91] V. Vapnik, The nature of statistical learning theory, Springer science & business media, 2013.

[92] S. Abdollahi, H.R. Pourghasemi, G.A. Ghanbarian, R. Safaeian, Prioritization of effective factors in the occurrence of land subsidence and its susceptibility mapping using an SVM model and their different kernel functions, Bull. Eng. Geol. Environ. 78 (2019) 4017–4034. doi:10.1007/s10064-018-1403-6.

[93] K. Yan, C. Shi, Prediction of elastic modulus of normal and high strength concrete by support vector machine, Constr. Build. Mater. 24 (2010) 1479–1485. doi:10.1016/j.conbuildmat.2010.01.006.

[94] J. Tinoco, A.G. Correia, P. Cortez, Support vector machines applied to uniaxial compressive strength prediction of jet grouting columns, Comput. Geotech. 55 (2014) 132–140. doi:10.1016/j.compgeo.2013.08.010.

[95] H. Yan, J. Zhang, S.S. Rahman, N. Zhou, Y. Suo, Predicting permeability changes with injecting CO2 in coal seams during CO2 geological sequestration: A comparative study among six SVM-based hybrid models, Sci. Total Environ. 705 (2020) 135941. doi:doi.org/10.1016/j.scitotenv.2019.135941.

[96] H. Benimam, C.S. Moussa, M. Hentabli, S. Hanini, M. Laidi, Dragon fl y-Support Vector Machine for Regression Modeling of the Activity Coefficient at Infinite Dilution of Solutes in Imidazolium Ionic Liquids Using σ⁻Profile Descriptors, (2020). doi:10.1021/acs.jced.0c00168.

[97] M. Moussaoui, M. Laidi, S. Hanini, M. Hentabli, Artificial neural network and support vector regression applied in quantitative structure-property relationship modelling of solubility of solid solutes in supercritical CO2, Kem. u Ind,. 69 (2020) 611–630. doi:10.15255/KUI.2020.004.

[98] S. Mirjalili, Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems, Neural Comput. Appl. 27 (2016) 1053–1073. doi:10.1007/s00521-015-1920-1.

[99] A. Ibrir, Y. Kerchich, N. Hadidi, H. Merabet, M. Hentabli, Prediction of the concentrations of PM1, PM2. 5, PM4, and PM10 by using the hybrid dragonfly-SVM algorithm, Air Qual. Atmos. Heal. (2020) 1–11. doi:10.1007/s11869-020-00936-1.

[100] P.J. Rousseeuw, A.M. Leroy, Robust regression and outlier detection, John wiley & sons, 2005.

[101] E. Uzlu, A. Akpinar, H.T. Özturk, S. Nacar, M. Kankal, Estimates of hydroelectric generation using neural networks with the artificial bee colony algorithm for Turkey, Energy. 69 (2014) 638–647. doi:10.1016/j.energy.2014.03.059.

[102] H. Ahmadi, M. Rodehutscord, Application of Artificial Neural Network and Support Vector Machines in Predicting Metabolizable Energy in Compound Feeds for Pigs, Front. Nutr. 4 (2017) 1–8. doi:10.3389/fnut.2017.00027.

[103] S. Andrea, T. Stefano, C. Francesca, R. Marco, Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models, John Wiley & Sons, Ltd, 2004. doi:10.1002/0470870958.

[104] K. Khimeche, P. Alessi, I. Kikic, A. Dahmani, Solubility of diamines in supercritical carbon dioxide. Experimental determination and correlation, J. Supercrit. Fluids. 41 (2007) 10–19. doi:10.1016/j.supflu.2006.09.004.

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

[106] Z. Huang, S. Kawi, Y.C. Chiew, Solubility of cholesterol and its esters in supercritical carbon dioxide with and without cosolvents, J. Supercrit. Fluids. 30 (2004) 25–39. doi:10.1016/S0896-8446(03)00116-5.

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

[108] Y. Yamini, N. Bahramifar, Solubility of polycyclic aromatic hydrocarbons in supercritical carbon dioxide, J. Chem. Eng. Data. 45 (2000) 53–56. doi:10.1021/je990129s.

[109] D.L. Sparks, R. Hernandez, L.A. Estévez, N. Meyer, T. French, Solubility of Azelaic Acid in Supercritical Carbon Dioxide, J. Chem. Eng. Data. 52 (2007) 1246–1249. doi:10.1021/je600572z.

[110] A. Garmroodi, J. Hassan, Y. Yamini, Solubilities of the drugs benzocaine, metronidazole benzoate, and naproxen in supercritical carbon dioxide, J. Chem. Eng. Data. 49 (2004) 709–712. doi:10.1021/je020218w.

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

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

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

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

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

[116] A.G. Reveco-Chilla, A.L. Cabrera, J.C. de la Fuente, F.C. Zacconi, J.M. del Valle, L.M. Valenzuela, Solubility of menadione and dichlone in supercritical carbon dioxide, Fluid Phase Equilib. 423 (2016) 84–92. doi:10.1016/j.fluid.2016.04.001.

[117] Y. Yamini, J. Arab, M. Asghari-khiavi, Solubilities of phenazopyridine, propranolol, and methimazole in supercritical carbon dioxide, J. Pharm. Biomed. Anal. 32 (2003) 181–187. doi:10.1016/S0731-7085(03)00016-5.

[118] P. Kotnik, M. Škerget, Z. Knez, 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 (2011) 338–343. doi:10.1021/je100697a.

[119] M. Khamda, M.H. Hosseini, M. Rezaee, Measurement and correlation solubility of cefixime trihydrate and oxymetholone in supercritical carbon dioxide (CO 2), J. Supercrit. Fluids. 73 (2013) 130–137. doi:10.1016/j.supflu.2012.09.006.

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

[121] J.T. Paula, I.M.O. Sousa, M.A. Foglio, F.A. Cabral, Solubility of protocatechuic acid, sinapic acid and chrysin in supercritical carbon dioxide, J. Supercrit. Fluids. 112 (2016) 89–94. doi:10.1016/j.supflu.2016.02.014.

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

[123] J. García-González, M.J. Molina, F. Rodríguez, F. Mirada, Solubilities of phenol and pyrocatechol in supercritical carbon dioxide, J. Chem. Eng. Data. 46 (2001) 918–921. doi:10.1021/je0003795.

[124] Y. Yamini, M.R. Fat'Hi, N. Alizadeh, M. Shamsipur, Solubility of dihydroxybenzene isomers in supercritical carbon dioxide, Fluid Phase Equilib. 152 (1998) 299–305. doi:10.1016/S0378-3812(98)00385-9.

[125] F. Zabihi, M. Mirzajanzadeh, J. Jia, Y. Zhao, Measurement and calculation of solubility of quinine in supercritical carbon dioxide, Chinese J. Chem. Eng. 25 (2017) 641–645. doi:10.1016/j.cjche.2016.10.003.

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

[127] G.I. Burgos-Solórzano, J.F. Brennecke, M.A. Stadtherr, Solubility measurements and modeling of

molecules of biological and pharmaceutical interest with supercritical CO2, Fluid Phase Equilib. 220 (2004) 57–69. doi:10.1016/j.fluid.2004.01.036.

[128] S.J. Nejad, R. Mohammadikhah, H. Abolghasemi, M.A. Moosavian, M.G. Maragheh, A Novel equation of state (EOS) for prediction of solute solubility in supercritical carbon dioxide: Experimental determination and correlation, Can. J. Chem. Eng. 87 (2009) 930–938. doi:10.1002/cjce.20232.

[129] S.Y. Huang, M. Tang, S.L. Ho, Y.P. Chen, Solubilities of N-phenylacetamide, 2-methyl-N-phenylacetamide and 4-methyl-N-phenylacetamide in supercritical carbon dioxide, J. Supercrit. Fluids. 42 (2007) 165–171. doi:10.1016/j.supflu.2007.04.001.

[130] G. Tian, J. Jin, Z. Zhang, J. Guo, Solubility of mixed solids in supercritical carbon dioxide, Fluid Phase Equilib. 251 (2007) 47–51. doi:10.1016/j.fluid.2006.11.002.

[131] K. Tamura, R.S. Alwi, T. Tanaka, K. Shimizu, Solubility of 1-aminoanthraquinone and 1-nitroanthraquinone in supercritical carbon dioxide, J. Chem. Thermodyn. 104 (2017) 162–168. doi:10.1016/j.jct.2016.09.032.

[132] R. Murga, M.T. Sanz, S. Beltrán, J.L. Cabezas, Solubility of three hydroxycinnamic acids in supercritical carbon dioxide, J. Supercrit. Fluids. 27 (2003) 239–245. doi:10.1016/S0896-8446(02)00265-6.

[133] M. Johannsen, G. Brunner, Solubilities of the Fat-Soluble Vitamins A, D, E, and K in Supercritical Carbon Dioxide, J. Chem. Eng. Data. 42 (1997) 106–111. doi:10.1021/je960219m.

[134] K.-L. Tsai, F.-N. Tsai, Solubilities of Methylbenzoic Acid Isomers in Supercritical Carbon Dioxide, J. Chem. Eng. Data. 40 (1995) 264–266. doi:10.1021/je00017a057.

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

[136] V. Pauchon, Z. Cissé, M. Chavret, J. Jose, A new apparatus for the dynamic determination of solid compounds solubility in supercritical carbon dioxide: Solubility determination of triphenylmethane, J. Supercrit. Fluids. 32 (2004) 115–121. doi:10.1016/j.supflu.2004.03.003.

[137] G. Sodeifian, S.A. Sajadian, F. Razmimanesh, Solubility of an antiarrhythmic drug (amiodarone hydrochloride) in supercritical carbon dioxide: Experimental and modeling, Fluid Phase Equilib. 450 (2017) 149–159. doi:10.1016/j.fluid.2017.07.015.

[138] K. Tamura, R.S. Alwi, Solubility of anthraquinone derivatives in supercritical carbon dioxide, Dye. Pigment. 113 (2015) 351–356. doi:10.1016/j.dyepig.2014.09.003.

[139] J. Jin, Y. Wang, H. Liu, Z. Zhang, Determination and calculation of solubility of bisphenol A in supercritical carbon dioxide, Chem. Eng. Res. Des. 91 (2013) 158–164. doi:10.1016/j.cherd.2012.06.013.

[140] J. Jin, Y. Ning, K. Hu, H. Wu, Z. Zhang, Solubility of p -Nitroaniline in Supercritical Carbon Dioxide with and without Mixed Cosolvents, J. Chem. Eng. Data. 58 (2013) 1464–1469. doi:10.1021/je300987d.

[141] S. Ismadji, Solubility of methyl salicylate in supercritical carbon dioxide at several temperatures, J. Chem. Eng. Data. 53 (2008) 2207–2210. doi:10.1021/je800476n.

[142] S. Marceneiro, P. Coimbra, M.E.M. Braga, A.M.A. Dias, H.C. De Sousa, Measurement and correlation of the solubility of juglone in supercritical carbon dioxide, Fluid Phase Equilib. 311 (2011) 1–8. doi:10.1016/j.fluid.2011.08.024.

[143] J. Yau, F. Tsai, Solubilities of 1-Eicosanol and Eicosanoic Acid in Supercritical Carbon Dioxide from 308.2 to 328.2 K at Pressures to 21.26 MPa, J. Chem. Eng. Data. 39 (1994) 827–829.

[144] K. Anatoly, T. George, Solubility of I-Octadecanol and Stearic Acid in Supercritical Carbon Dioxide, 187 (1989) 184–187.

[145] H.A. Martinez-Correa, D.C.A. Gomes, S.L. Kanehisa, F.A. Cabral, Measurements and thermodynamic modeling of the solubility of squalene in supercritical carbon dioxide, J. Food Eng. 96 (2010) 43–50. doi:10.1016/j.jfoodeng.2009.06.041.

[146] I. Goodarznia, F. Esmaeilzadeh, Solubility of an anthracene, phenanthrene, and carbazole mixture in supercritical carbon dioxide, J. Chem. Eng. Data. 47 (2002) 333–338. doi:10.1021/je010093f.

[147] M. Ashraf-Khorassani, M.T. Combs, L.T. Taylor, F.K. Schweighardt, P.S. Mathias, Solubility study of sulfamethazine and sulfadimethoxine in supercritical carbon dioxide, fluoroform, and subcritical freon 134A, J. Chem. Eng. Data. 42 (1997) 636–640. doi:10.1021/je960402f.

[148] J.W. Hampson, A Recirculating Equilibrium Procedure for Determining Organic Compound Solubility in Supercritical Fluids. Anthracene in Carbon Dioxide, J. Chem. Eng. Data. 41 (1996) 97–100. doi:10.1021/je950167l.

[149] A. Cortesi, I. Kikic, P. Alessi, G. Turtoi, S. Garnier, Effect of chemical structure on the solubility of antioxidants in supercritical carbon dioxide : experimental data and correlation, 14 (1999) 139–144.

[150] L. Barna, J. Blanchard, E. Rauzy, C. Berro, D. Lyon, Solubility of Flouranthene , Chrysene , and Triphenylene in Supercritical Carbon Dioxide, 9568 (1996) 1466–1469.

[151] J. Gregorowicz, Solubilities of lactic acid and 2-hydroxyhexanoic acid in supercritical CO2, Fluid Phase Equilib. 166 (1999) 39–46. doi:10.1016/S0378-3812(99)00283-6.

[152] E. Reverchon, P. Russo, A. Stassi, Solubilities of Solid Octacosane and Triacontane in Supercritical Carbon Dioxide, J. Chem. Eng. Data. 38 (1993) 458–460. doi:10.1021/je00011a034.

[153] C. Garlapati, G. Madras, Solubilities of hexadecanoic and octadecanoic acids in supercritical CO with and without cosolvents, J. Chem. Eng. Data. 53 (2008) 2913–2917. doi:10.1021/je8007149.

[154] H. Xing, Y. Yang, B. Su, M. Huang, Q. Ren, Solubility of Artemisinin in Supercritical Carbon Dioxide, J. Chem. Eng. Data. 48 (2003) 330–332.

[155] S.H. Cheng, F.C. Yang, Y.H. Yang, C.C. Hu, W.T. Chang, Measurements and modeling of the solubility of ergosterol in supercritical carbon dioxide, J. Taiwan Inst. Chem. Eng. 44 (2013) 19–26. doi:10.1016/j.jtice.2012.09.001.

[156] S.L.J. Yun, K.K. Liong, G.S. Gurdial, N.R. Foster, Solubility of cholesterol in supercritical carbon dioxide, Ind. Eng. Chem. Res. 30 (1991) 2476–2482. doi:10.1021/ie00059a018.

[157] S.J. Macnaughton, N.R. Foster, Solubility of DDT and 2,4-D in Supercritical Carbon Dioxide and Supercritical Carbon Dioxide Saturated with Water, Ind. Eng. Chem. Res. 33 (1994) 2757–2763. doi:10.1021/ie00035a027.

[158] R. Murga, M.T. Sanz, S. Beltrán, J.L. Cabezas, Solubility of some phenolic compounds contained in grape seeds, in supercritical carbon dioxide, J. Supercrit. Fluids. 23 (2002) 113–121. doi:10.1016/S0896-8446(02)00033-5.

[159] S. Marceneiro, M.E.M. Braga, A.M.A. Dias, H.C. De Sousa, Measurement and Correlation of 1 , 4-Naphthoquinone and of Plumbagin Solubilities in Supercritical Carbon Dioxide, J. Chem. Eng. Data. (2011) 4173–4182.

[160] P. Coutsikos, K. Magoulas, D. Tassios, Solubilities of p-Quinone and 9 , 10-Anthraquinone in Supercritical Carbon Dioxide, J. Chem. Eng. Data. 42 (1997) 463–466. doi:10.1021/je960309r.

[161] R. Murga, M.T. Sanz, S. Beltrán, J.L. Cabezas, Solubility of syringic and vanillic acids in supercritical carbon dioxide, J. Chem. Eng. Data. 49 (2004) 779–782. doi:10.1021/je034129a.

[162] S. Bristow, B.Y. Shekunov, P. York, Solubility Analysis of Drug Compounds in Supercritical Carbon Dioxide Using Static and Dynamic Extraction Systems, Ind. Eng. Chem. Res. 40 (2001) 1732–1739. doi:10.1021/ie0002834.

[163] J. Shi, M. Khatri, S.J. Xue, G.S. Mittal, Y. Ma, D. Li, Solubility of lycopene in supercritical CO2 fluid as affected by temperature and pressure, Sep. Purif. Technol. 66 (2009) 322–328. doi:10.1016/j.seppur.2008.12.012.

[164] E. Kosal, C.H. Lee, G.D. Holder, Solubility of progesterone, testosterone, and cholesterol in supercritical fluids, J. Supercrit. Fluids. 5 (1992) 169–179. doi:10.1016/0896-8446(92)90004-4.

[165] N.R. Foster, S.L.J. Yun, S.S.T. Ting, Solubility of oleic acid in supercritical carbon dioxide, J. Supercrit. Fluids. 4 (1991) 127–130. doi:10.1016/0896-8446(91)90041-4.

[166] J.P. Coelho, G.P. Naydenov, D.S. Yankov, R.P. Stateva, Experimental measurements and correlation of the solubility of three primary amides in supercritical CO2: Acetanilide, propanamide, and butanamide, J. Chem. Eng. Data. 58 (2013) 2110–2115. doi:10.1021/je400357t.

[167] Y.P. Chen, Y.M. Chen, M. Tang, Solubilities of cinnamic acid, phenoxyacetic acid and 4-

methoxyphenylacetic acid in supercritical carbon dioxide, Fluid Phase Equilib. 275 (2009) 33–38. doi:10.1016/j.fluid.2008.09.009.

[168] N. Lamba, R.C. Narayan, J. Modak, G. Madras, Solubilities of 10-undecenoic acid and geraniol in supercritical carbon dioxide, J. Supercrit. Fluids. 107 (2016) 384–391. doi:10.1016/j.supflu.2015.09.026.

[169] H. Higashi, Y. Iwai, K. Miyazaki, Y. Ogino, M. Oki, Y. Arai, Measurement and correlation of solubilities for trifluoromethylbenzoic acid isomers in supercritical carbon dioxide, J. Supercrit. Fluids. 33 (2005) 15–20. doi:10.1016/j.supflu.2004.03.006.

[170] H. Perrotin-Brunel, M.J.E. Van Roosmalen, M.C. Kroon, J. Van Spronsen, G.J. Witkamp, C.J. Peters, Solubility of cannabinol in supercritical carbon dioxide, J. Chem. Eng. Data. 55 (2010) 3704–3707. doi:10.1021/je100245n.

[171] J.P. Coelho, K. Bernotaityte, M.A. Miraldes, A.F. Mendonça, R.P. Stateva, Solubility of ethanamide and 2-propenamide in supercritical carbon dioxide. Measurements and correlation, J. Chem. Eng. Data. 54 (2009) 2546–2549. doi:10.1021/je900109b.

[172] J. Lim, H. Kim, H.K. Cho, M.S. Shin, Solubility of hinokitiol in supercritical fluids; measurement and correlation, Korean J. Chem. Eng. 28 (2011) 2319–2323. doi:10.1007/s11814-011-0112-7.

[173] M.S. Shin, H. Kim, Solubility of iodopropynyl butylcarbamate in supercritical carbon dioxide, Fluid Phase Equilib. 270 (2008) 45–49. doi:10.1016/j.fluid.2008.05.010.

[174] J. Fan, Y. Hou, W. Wu, J. Zhang, S. Ren, X. Chen, Levulinic acid solubility in supercritical carbon dioxide with and without ethanol as cosolvent at different temperatures, J. Chem. Eng. Data. 55 (2010) 2316–2321. doi:10.1021/je900727r.

[175] S.N. Reddy, G. Madras, Solubilities of Benzene Derivatives in Supercritical Carbon Dioxide, J. Chem. Eng. Data. 56 (2011) 1695–1699. doi:10.1021/je100863p.

[176] H.C. De Sousa, M.S. Costa, P. Coimbra, A.A. Matias, C.M.M. Duarte, Experimental determination and correlation of meloxicam sodium salt solubility in supercritical carbon dioxide, J. Supercrit. Fluids. 63 (2012) 40–45. doi:10.1016/j.supflu.2011.12.004.

[177] R. Ch, C. Garlapati, G. Madras, Solubility of n-(4-ethoxyphenyl)ethanamide in supercritical carbon dioxide, J. Chem. Eng. Data. 55 (2010) 1437–1440. doi:10.1021/je900614f.

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

[179] C. Si-Moussa, A. Belghait, L. Khaouane, S. Hanini, A. Halilali, Novel density-based model for the correlation of solid drugs solubility in supercritical carbon dioxide, Comptes Rendus Chim. 20 (2017) 559–572. doi:10.1016/j.crci.2016.09.009.

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

[181] A.Z. Hezave, M. Lashkarbolooki, A new simple correlation for calculating solubility of drugs in supercritical carbon dioxide, J. Theor. Comput. Chem. 12 (2013). doi:10.1142/S0219633613500624.

[182] M. Abdi-Khanghah, A. Bemani, Z. Naserzadeh, Z. Zhang, Prediction of solubility of N-alkanes in supercritical CO2 using RBF-ANN and MLP-ANN, J. CO2 Util. 25 (2018) 108–119. doi:10.1016/j.jcou.2018.03.008.

[183] A. Ray, T. Halder, S. Jena, A. Sahoo, B. Ghosh, S. Mohanty, N. Mahapatra, S. Nayak, Application of artificial neural network (ANN) model for prediction and optimization of coronarin D content in Hedychium coronarium, Ind. Crops Prod. 146 (2020) 112186. doi:10.1016/j.indcrop.2020.112186.

[184] W.H. Chen, S.H. Hsu, H.P. Shen, Application of SVM and ANN for intrusion detection, Comput. Oper. Res. 32 (2005) 2617–2634. doi:10.1016/j.cor.2004.03.019.

[185] D.K. Agrafiotis, W. Cedeno, V.S. Lobanov, On the use of neural network ensembles in QSAR and QSPR, J. Chem. Inf. Comput. Sci. 42 (2002) 903–911.

[186] T.N.G. Borhani, A. Afzali, M. Bagheri, QSPR estimation of the auto-ignition temperature for pure

hydrocarbons, Process Saf. Environ. Prot. 103 (2016) 115–125. doi:10.1016/j.psep.2016.07.004.

[187] G. Sodeifian, N. Saadati Ardestani, S.A. Sajadian, H.S. Panah, Measurement, correlation and thermodynamic modeling of the solubility of Ketotifen fumarate (KTF) in supercritical carbon dioxide: Evaluation of PCP-SAFT equation of state, Fluid Phase Equilib. 458 (2018) 102–114. doi:10.1016/j.fluid.2017.11.016.

[188] G. Sodeifian, S.A. Sajadian, Experimental measurement of solubilities of sertraline hydrochloride in supercritical carbon dioxide with/without menthol: Data correlation with a pressure of about 60, J. Supercrit. Fluids. 149 (2019) 79–87. doi:10.1016/j.supflu.2019.03.020.

[189] G. Sodeifian, S.A. Sajadian, N.S. Ardestani, Determination of solubility of Aprepitant (an antiemetic drug for chemotherapy) in supercritical carbon dioxide: Empirical and thermodynamic models, J. Supercrit. Fluids. 128 (2017) 102–111. doi:10.1016/j.supflu.2017.05.019.

[190] G. Sodeifian, N. Saadati Ardestani, S.A. Sajadian, H. Soltani Panah, Experimental measurements and thermodynamic modeling of Coumarin-7 solid solubility in supercritical carbon dioxide: Production of nanoparticles via RESS method, Fluid Phase Equilib. 483 (2019) 122–143. doi:10.1016/j.fluid.2018.11.006.

[191] G. Sodeifian, S.A. Sajadian, Solubility measurement and preparation of nanoparticles of an anticancer drug (Letrozole) using rapid expansion of supercritical solutions with solid cosolvent (RESS-SC), J. Supercrit. Fluids. 133 (2018) 239–252. doi:10.1016/j.supflu.2017.10.015.

[192] G. Sodeifian, F. Razmimanesh, S.A. Sajadian, Solubility measurement of a chemotherapeutic agent (Imatinib mesylate) in supercritical carbon dioxide: Assessment of new empirical model, J. Supercrit. Fluids. 146 (2019) 89–99. doi:10.1016/j.supflu.2019.01.006.

[193] G. Sodeifian, R. Detakhsheshpour, S.A. Sajadian, Experimental study and thermodynamic modeling of Esomeprazole (proton-pump inhibitor drug for stomach acid reduction) solubility in supercritical carbon dioxide, J. Supercrit. Fluids. 154 (2019) 104606. doi:10.1016/j.supflu.2019.104606.

[194] G. Sodeifian, F. Razmimanesh, S.A. Sajadian, H. Soltani Panah, Solubility measurement of an antihistamine drug (Loratadine) in supercritical carbon dioxide: Assessment of qCPA and PCP-SAFT equations of state, Fluid Phase Equilib. 472 (2018) 147–159. doi:10.1016/j.fluid.2018.05.018.

[195] G. Sodeifian, S.A. Sajadian, R. Derakhsheshpour, Experimental measurement and thermodynamic modeling of Lansoprazole solubility in supercritical carbon dioxide: Application of SAFT-VR EoS, Fluid Phase Equilib. 507 (2020) 112422. doi:10.1016/j.fluid.2019.112422.

[196] G. Sodeifian, F. Razmimanesh, S.A. Sajadian, Prediction of solubility of sunitinib malate (an anti-cancer drug) in supercritical carbon dioxide (SC–CO2): Experimental correlations and thermodynamic modeling, J. Mol. Liq. 297 (2020) 111740. doi:10.1016/j.molliq.2019.111740.

[197] G. Sodeifian, F. Razmimanesh, N. Saadati Ardestani, S.A. Sajadian, Experimental data and thermodynamic modeling of solubility of Azathioprine, as an immunosuppressive and anti-cancer drug, in supercritical carbon dioxide, J. Mol. Liq. 299 (2020). doi:10.1016/j.molliq.2019.112179.

[198] G. Sodeifian, F. Razmimanesh, S.A. Sajadian, S.M. Hazaveie, Experimental data and thermodynamic modeling of solubility of Sorafenib tosylate, as an anti-cancer drug, in supercritical carbon dioxide: Evaluation of Wong-Sandler mixing rule, J. Chem. Thermodyn. 142 (2020) 105998. doi:10.1016/j.jct.2019.105998.

[199] G. Sodeifian, S.M. Hazaveie, S.A. Sajadian, N. Saadati Ardestani, Determination of the Solubility of the Repaglinide Drug in Supercritical Carbon Dioxide: Experimental Data and Thermodynamic Modeling, J. Chem. Eng. Data. (2019). doi:10.1021/acs.jced.9b00550.

[200] G. Sodeifian, S.M. Hazaveie, S.A. Sajadian, F. Razmimanesh, Experimental investigation and modeling of the solubility of oxcarbazepine (an anticonvulsant agent)in supercritical carbon dioxide, Fluid Phase Equilib. 493 (2019) 160–173. doi:10.1016/j.fluid.2019.04.013.

[201] G. Sodeifian, N. Saadati Ardestani, S.A. Sajadian, Solubility measurement of a pigment (Phthalocyanine green) in supercritical carbon dioxide: Experimental correlations and thermodynamic modeling, Fluid Phase Equilib. 494 (2019) 61–73. doi:10.1016/j.fluid.2019.04.024.

[202] M. Hamadache, O. Benkortbi, S. Hanini, A. Amrane, L. Khaouane, C. Si Moussa, A Quantitative Structure Activity Relationship for acute oral toxicity of pesticides on rats: Validation, domain of application and prediction, J. Hazard. Mater. 303 (2016) 28–40. doi:10.1016/j.jhazmat.2015.09.021.

[203] V.H. Masand, V. Rastija, PyDescriptor: A new PyMOL plugin for calculating thousands of easily understandable molecular descriptors, Chemom. Intell. Lab. Syst. 169 (2017) 12–18.

[204] I. V Tetko, J. Gasteiger, R. Todeschini, A. Mauri, D. Livingstone, P. Ertl, V.A. Palyulin, E. V Radchenko, N.S. Zefirov, A.S. Makarenko, others, Virtual computational chemistry laboratory--design and description, J. Comput. Aided. Mol. Des. 19 (2005) 453–463.

[205] M. Hamadache, S. Hanini, O. Benkortbi, A. Amrane, L. Khaouane, C.S. Moussa, Artificial neural network-based equation to predict the toxicity of herbicides on rats, Chemom. Intell. Lab. Syst. 154 (2016) 7–15.

[206] A.H. Elsheikh, S.W. Sharshir, M. Abd Elaziz, A.E. Kabeel, W. Guilan, Z. Haiou, Modeling of solar energy systems using artificial neural network: A comprehensive review, Sol. Energy. 180 (2019) 622–639. doi:10.1016/j.solener.2019.01.037.

[207] A.M. Ghaedi, A. Vafaei, Applications of artificial neural networks for adsorption removal of dyes from aqueous solution: A review, Adv. Colloid Interface Sci. 245 (2017) 20–39. doi:10.1016/j.cis.2017.04.015.

[208] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, Stat. Comput. 14 (2004) 199–222. doi:10.1023/B:STCO.0000035301.49549.88.

[209] V. Vapnik, Statistical Learning Theory, New York:, John Wiley & Sons, Inc, 1998.

[210] G. Sodeifian, S.A. Sajadian, F. Razmimanesh, N.S. Ardestani, A comprehensive comparison among four different approaches for predicting the solubility of pharmaceutical solid compounds in supercritical carbon dioxide, Korean J. Chem. Eng. 35 (2018) 2097–2116. doi:10.1007/s11814-018-0125-6.

[211] A. Aminian, Estimating the solubility of different solutes in supercritical CO2 covering a wide range of operating conditions by using neural network models, J. Supercrit. Fluids. 125 (2017) 79–87. doi:10.1016/j.supflu.2017.02.007.

[212] B. Vaferi, M. Karimi, M. Azizi, H. Esmaeili, Comparison between the artificial neural network, SAFT and PRSV approach in obtaining the solubility of solid aromatic compounds in supercritical carbon dioxide, J. Supercrit. Fluids. 77 (2013) 44–51. doi:10.1016/j.supflu.2013.02.027.

[213] A. Eslamimanesh, F. Gharagheizi, A.H. Mohammadi, D. Richon, Artificial Neural Network modeling of solubility of supercritical carbon dioxide in 24 commonly used ionic liquids, Chem. Eng. Sci. 66 (2011) 3039–3044. doi:10.1016/j.ces.2011.03.016.

[214] J. Méndez-Santiago, A.S. Teja, The solubility of solids in supercritical fluids, Fluid Phase Equilib. 158–160 (1999) 501–510. doi:10.1016/s0378-3812(99)00154-5.

[215] S.A. Shojaee, H. Rajaei, A.Z. Hezave, M. Lashkarbolooki, F. Esmaeilzadeh, Experimental investigation and modeling of the solubility of carvedilol in supercritical carbon dioxide, J. Supercrit. Fluids. 81 (2013) 42–47. doi:10.1016/j.supflu.2013.04.013.

[216] S.A.L.I. Shojaee, H. Rajaei, A.L.I.Z. Hezave, Experimental solubility measurement of cephalexin in supercritical carbon dioxide, Chem. Ind. Chem. Eng. Q. 20 (2014) 387–396. doi:10.2298/CICEQ121128021S.

[217] M. Lashkarbolooki, A.Z. Hezave, Y. Rahnama, R. Ozlati, H. Rajaei, F. Esmaeilzadeh, Solubility of cyproheptadine in supercritical carbon dioxide; experimental and modeling approaches, J. Supercrit. Fluids. 84 (2013) 13–19. doi:10.1016/j.supflu.2013.09.004.

[218] A. Zeinolabedini Hezave, H. Rajaei, M. Lashkarbolooki, F. Esmaeilzadeh, Analyzing the solubility of fluoxetine hydrochloride in supercritical carbon dioxide, J. Supercrit. Fluids. 73 (2013) 57–62. doi:10.1016/j.supflu.2012.11.005.

[219] S.A. Shojaee, A.Z. Hezave, S. Aftab, M. Lashkarbolooki, F. Esmaeilzadeh, Solubility of gabapentin in supercritical carbon dioxide, J. Supercrit. Fluids. 78 (2013) 1–6. doi:10.1016/j.supflu.2013.02.003.

[220] H. Rajaei, A.Z. Hezave, M. Lashkarbolooki, F. Esmaeilzadeh, Representing experimental solubility of phenylephrine hydrochloride in supercritical carbon dioxide and modeling solute solubility using semi-empirical correlations, J. Supercrit. Fluids. 75 (2013) 181–186. doi:10.1016/j.supflu.2012.11.014.

[221] S.A. Shojaee, H. Rajaei, A.Z. Hezave, M. Lashkarbolooki, F. Esmaeilzadeh, Experimental measurement and correlation for solubility of piroxicam (a non-steroidal anti-inflammatory drugs (NSAIDs)) in

supercritical carbon dioxide, J. Supercrit. Fluids. 80 (2013) 38–43. doi:10.1016/j.supflu.2013.03.015.

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

Bibliographie de l'auteur

https://doi.org/10.15255/KUI.2020.004
KUI-40/2020
Original scientific paper
Received January 18, 2020
Accepted May 13, 2020

Artificial Neural Network and Support Vector Regression Applied in Quantitative Structure-property Relationship Modelling of Solubility of Solid Solutes in Supercritical CO_2

M. Moussaoui,[a,b,*] *M. Laidi*,[a] *S. Hanini*,[a] and *M. Hentabli* [a]

This work is licensed under a Creative Commons Attribution 4.0 International License

[a] Laboratory of Biomaterials and Transport Phenomena (LBMPT), University of Médéa, Médéa, Algeria
[b] University of Bouira, Bouira, Algeria

Abstract

In this study, the solubility of 145 solid solutes in supercritical CO_2 ($scCO_2$) was correlated using computational intelligence techniques based on Quantitative Structure-Property Relationship (QSPR) models. A database of 3637 solubility values has been collected from previously published papers. Dragon software was used to calculate molecular descriptors of 145 solid systems. The genetic algorithm (GA) was implemented to optimise the subset of the significantly contributed descriptors. The overall average absolute relative deviation MAARD of about 1.345 % between experimental and calculated values by support vector regress SVR-QSPR model was obtained to predict the solubility of 145 solid solutes in supercritical CO_2, which is better than that obtained using ANN-QSPR model of 2.772 %. The results show that the developed SVR-QSPR model is more accurate and can be used as an alternative powerful modelling tool for QSAR studies of the solubility of solid solutes in supercritical carbon dioxide ($scCO_2$). The accuracy of the proposed model was evaluated using statistical analysis by comparing the results with other models reported in the literature.

Keywords

Solubility, solid solutes, supercritical-fluids, computational intelligence techniques, quantitative structure-property relationship

1 Introduction

Supercritical carbon dioxide ($scCO_2$) is generally used in separation processes applied in the food, chemistry, pharmaceutical, and other industries. The design and optimisation of extraction, fractionation, and purification processes are mainly based on the knowledge of the solubility of solid solutes in $scCO_2$.[1] The experimental measurement of the solubility (thermo-physical property) of such compounds in $scCO_2$ is laborious and costly.[2] To avoid expensive and tedious experiments, and to fulfil the lack of solubility data and/or pure component property data required to estimate solubility, a need exists to develop flexible and robust predictive models to estimate the solubility of solid solutes in supercritical solvents using limited information. Literature reports three major modelling approaches have been used to model the solubility of a solid solute in $scCO_2$ using equations of state,[3] semi empirical equations,[2] and computational intelligence techniques such as artificial neural networks and support vector machine.[4] The solid solute solubility in $scCO_2$ is mostly modelled using density-based correlations, such as in [5]. Moreover, MLP-ANN also has been used successfully to model the solid solute solubility in $scCO_2$ based only on thermodynamic parameters.[6] The least square support vector machine LS-SVM may be considered an alternative for classical approaches, such as semi empirical correlations to model the solid solute solubility in $scCO_2$.[7] The most common type of ANN model that is being used nowadays is multilayer perceptron (MLP) feedforward neural network (FNN) trained by back-propagation (BP) algorithm.[8] This ANN type has also been successfully used for mapping complex highly nonlinear input/output relationships between one dependent variable and more independent variables of any systems. But there are numerous limitations, for instance, difficulty in deciding the criterion for the number of hidden neurons and layers, as well as over-fitting.[9] To avoid these problems, SVM is formulated as a quadratic optimisation problem and ensures a global optimal solution. Some studies have proven the superiority of SVM over ANN.[10] SVM method has gained a wide range of engineering applications in forecasting and regression analysis due to its attractive features and remarkable generalization performance.[11] Therefore the present study was carried out to evaluate the predictive performance of computational intelligence techniques including MPL-FNN as the most used neural network type and SVM to foretell the solubility of 145 different solid solutes in $scCO_2$ based on a mixture between thermodynamic properties and molecular descriptors. The novelty of the developed model in contrast to previous models is its ability to extrapolate or interpolate the solubility of several solid solutes in $scCO_2$ with only one combination of the fitted parameters obtained during the training stage. In addition, the independent parameters were a mixture between thermodynamic properties and structural descriptors. Moreover, a simple and convivial graphical user interface was designed to compute the solubility without learning MATLAB software. Finally, a sensitivity analysis was computed by calculating the relative importance of each input parameter on the output.

* Corresponding author: Assist. Professor Mohammed Moussaoui
Email: m.moussaoui@univ-bouira.dz

Chemistry in Industry
JOURNAL OF CHEMISTS & CHEMICAL ENGINEERS OF CROATIA
Publisher: Croatian Society of Chemical Engineers

Asst. Professor **MOHAMMED MOUSSAOUI**
Laboratory of Biomaterials and Transport Phenomenon (LBMPT)
Faculty of Technology, University Yahia Fares of Médéa
26 000 Algeria

December 8, 2020

DECLARATION OF ACCEPTANCE FOR PUBLICATION

Dear Prof Moussaoui,

it is our pleasure to inform you that your paper entitled:

Critical Properties and Acentric Factors of Pure Compounds Modelling Based on QSPR-SVM with Dragonfly Algorithm – KUI-63/2020 (*Mohammed Moussaoui, Maamar Laidi, Salah Hanini, Abdallah Abdellah El Hadj, Mohamed Hentabli*) after the conclusion of the reviewing/revision process has been accepted for publication in the journal *Kemija u industriji* under the category **ORIGINAL SCIENTIFIC PAPER** it will be published in one of the following issues of the journal.

ZAGREB HDKI HRVATSKO DRUŠTVO KEMIJSKIH INŽENJERA I TEHNOLOGA HRVATSKA

Yours sincerely,

Prof. Nenad Bolf
Editor-in-Chief