REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE YAHIA FARES DE MEDEA



Faculté de **T**echnologie

Département de **G**énie des **P**rocédés et d'**E**nvironnement

En vue de l'obtention du diplôme de MASTER en Génie des Procédés

Filière : Génie des Procédés
Spécialité : Génie Chimique

La prédiction de la rétention des pesticides lors de nanofiltration et d'osmose inverse par les modèles QSPR-RNA

Présenté par :

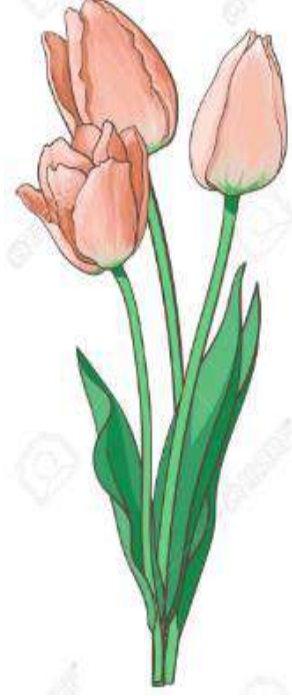
- M^{me} CHABANE Yasmine
- M^{elle} KOUL Asmaa

Devant le jury composé de :

Nom & Prénom	Grade & Etablissement	En Qualité de
Mr. LAIDI Maamar	Professeur, UYFM	Présendent
M^{me}. CHABA MOUNA Siham	Maitre de conférence classe B, UYFM	Examinatrice
M^{me}. MAOUZ Hadjira	Docteur à l'UYFMN	Examinatrice
Mr. HANINI Salah	Professeur, UYFM	Co-promoteur
M^{elle}. AMMI Yamina	Maitre de conférences classe A, UYFM	Promotrice

Promotion 2021-2022


بسم الله الرحمن الرحيم



Dédicaces

A mes parents qui mes très chers et dont le soutien m'a toujours aidé à réussir dans la vie j'espère qu'ils sont fiers de moi et mon travail, car je suis fierté d'avoir des tels parents

Ames très chères grandes parentes Miloud, el hadj, Yamna et Kelthoum

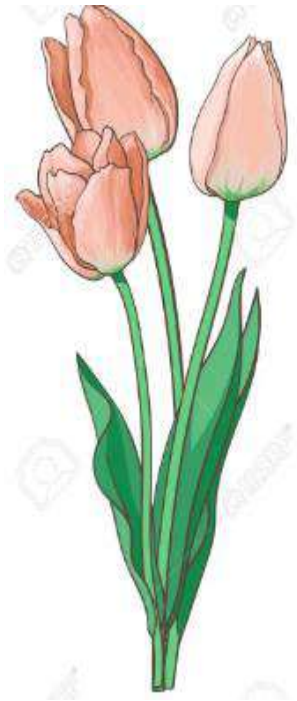
Mes chères sœurs Sabrina, safia et Soundous et mon cher frère Ayoub.

À toutes la famille koul et chellah

A tous mes amis et mes collègues, surtout mes amies Yasmin, Asmaa, Wafa, Imane, Nihad, Sarah et Khadîdja

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

Asmaa



Dédicaces

J'ai le grand honneur de dédier ce modeste travail : A ceux qui m'ont encouragé et soutenu moralement et matériellement pendant les moments les plus difficiles et durant toute ma vie. A mon très cher père et ma très chère mère

A mon très cher mari hamza qui m'a donné la possibilité d'arriver là où je suis, je remercie Allah de vous avoir amené dans ma vie.

A mon deuxième papa, mon oncle khaled ben moussa

A mon très cher frère zakaria et A mes beaux frères billel et khaled

A mes adorables sœurs malak et hiba et ma belle sœur maroua leur soutien est sans failles.

A mes oncles et mes tantes et tous mes cousines et cousins

*Toute la famille « **CHABANE**» et «**BEN MOUSSA**».*

*« **BEN MAMI**».*

*A Mes chers amies **Asmaa, Imane et Wafa** pour leur amitié sincère et pour tous les bons moments passés ensemble.*

Pour ceux que je n'ai pas cités bien sûr, ne croyez pas que je vous ai oublié, je vous porte toujours dans mon cœur.

A tous les étudiants « science de technologie 2017-2022 ».

YASMINE

Remerciements et gratitude

Nous tenons tout d'abord à remercier Allah le tout puissant et miséricordieux, qui nous a donné la force et la patience d'accomplir ce modeste travail.

En second lieu, Nous voudrions, exprimer nos sincères et ineffables remerciements au docteur ***Yamina AMMI*** *qui a assuré l'encadrement de ce mémoire et qui a bien voulu nous accueillir au sein de son équipe. Tout au long de ce mémoire, elle nous n'a cessé de suivre et de s'intéresser à son avancement. Qu'elle trouve ici l'expression de notre reconnaissance et de notre profonde gratitude.*

Le Professeur ***Salah HANINI*** *s'est occupé du co-encadrement de ce travail. Son aide précieuse dans la compréhension des éléments de base des réseaux de neurones nous a été bénéfique. Nous lui réitérons à cette occasion nos sincères remerciements pour sa disponibilité et sa compréhension, qu'il trouve ici l'expression de notre profonde gratitude.*

Nos vifs remerciements vont également aux membres du jury pour l'intérêt qu'ils ont porté à notre recherche en acceptant d'examiner notre travail et de l'enrichir par leurs propositions, Monsieur ***Maamar LAIDI*** *professeur à l'UYFM et Madame* ***Siham CHABA MOUNA*** *Maître de conférences classe B à l'UYFM.*

Nous souhaitons remercier les professeurs département de génie des procédés de l'environnement pendant les cinq années précédentes.

Enfin, nous tenons également à remercier toutes les personnes qui ont participé de près ou de loin à la réalisation de ce travail.

Résumé

La prédiction de la rétention des pesticides lors de nanofiltration et d'osmose inverse par les modèles QSPR-RNA

L'objectif assigné à ce travail est d'élaborer un modèle optimal de réseau de neurones basé sur la relation quantitative structure-propriété ($QSPR\text{-}RN_{optimal}$) pour modéliser la rétention des pesticides par les membranes d'osmose inverse et de nanofiltration.

Un modèle neuronal de type feedforward ($QSPR\text{-}RN_{optimal}$) caractérisé par une structure (treize neurones dans la couche d'entrée, dix-sept neurones une couche cachée et un neurone dans la couche de sortie) est construit dans le but de prédire la rétention des pesticides. Un ensemble de227 points de données pour $QSPR\text{-}RN_{optimal}$ a été utilisé pour tester le réseau de neurones. 80%, 20% de l'ensemble des données sont été utilisées respectivement pour l'apprentissage, le test de modèle. Les valeurs de la rétention calculées pour la phase de test ont été comparées aux valeurs de la rétention mesurées, de bonne corrélation a été trouvée (R =0.9028). L'analyse de sensibilité par la méthode de "poids" a révélé que les variables les plus pertinentes qui peuvent influer sur la rétention de pesticides par les membranes d'OI/NF exercent des contributions importantes ($IR > 5\%$).

Mots-clés : Prédiction, Rétention, Pesticides, Nanofiltration, Osmose inverse, QSPR-RNA

Abstract

Prediction of pesticides retention during nanofiltration and reverse osmosis by QSPR-RNA models

The objective assigned to this work is to develop an optimal neural network model based on the quantitative structure-property relationship (QSPR-RNoptimal) to model the retention of pesticides by reverse osmosis and nanofiltration membranes.

A feedforward neural model (QSPR-RNoptimal) characterized by a structure (thirteen neurons in the input layer, seventeen neurons in a hidden layer, and one neuron in the output layer) is constructed in order to predict pesticide retention. A set of 227 data points for QSPR-RNoptimal was used to test the neural network. 80% and 20% of the whole dataset were used for training, and model testing, respectively. The retention values calculated for the test phase were compared with the measured retention values, and a good correlation was found (R = 0.9028). The sensitivity analysis by the "weight" method revealed that the most relevant variables that can influence the retention of pesticides by RO/NF membranes exert significant contributions (IR>5%).

Keywords: Prediction, Retention, Pesticides, Nanofiltration, Reverse osmosis, QSPR-RNA

ملخص

التنبؤ باحتفاظ المبيدات أثناء الترشيح النانوي والتناضح العكسي بواسطة نماذج QSPR-RNA

الهدف المخصص لهذا العمل هو تطوير نموذج شبكي عصبي مثالي يعتمد على العلاقة الكمية بين التركيب والخصائص (QSPR-RN Optimal) لنمذجة الاحتفاظ بالمبيدات عن طريق التناضح العكسي وأغشية الترشيح النانوي.

تم إنشاء نموذج عصبي متقدم (QSPR-RN Optimal) يتميز ببنية (ثلاثة عشر عصبون في طبقة الإدخال، وسبعة عشر عصبون في طبقة مخفية وعصبون واحد في طبقة الإخراج) من أجل التنبؤ باحتفاظ المبيدات. تم استخدام مجموعة من 227 نقطة بيانات لـ QSPR-RN Optimal لاختبار الشبكة العصبية. تم استخدام 80٪، 20٪ من مجموعة البيانات بأكملها للتدريب واختبار النماذج على التوالي. تمت مقارنة قيم الاحتفاظ المحسوبة لمرحلة الاختبار مع قيم الاحتفاظ المقاسة، وتم العثور على ارتباط جيد (R = 0.9028). كشف تحليل الحساسية باستخدام طريقة "الوزن" أن أكثرالمنغيرات ذات الصلة والتي يمكن أن تؤثر على احتفاظ المبيدات بواسطة أغشية RO / NF تقدم جميع المدخلات مساهمات كبيرة (IR> 5٪).

كلمات المفتاح التنبؤ، الاحتفاظ، مبيدات الآفات، الترشيح الدقيق، التناضح العكسي، QSPR-RNA

Abréviations i
Nomenclature et symboles ii
Liste des tableaux iii
Liste des figures iv

Introduction générale 1

Chapitre I. Rétention des pesticides par les membranes

I.1. Introduction 04
I.2. Généralités 04
I.2.1. Définition de la membrane 04
I.2.2. Classification des membranes 05
I.2.2.1. Nature chimique 05
I.2.2.2. Porosité 06
I.2.2.3. Morphologie 06
I.2.2.4. Classification d'après la géométrie de préparation des membranes 07
I.2.2.5. Classification d'après le mécanisme de séparation 10
I.2.3. Osmose inverse et nanofiltration 11
I.2.4. Sélectivité d'une membrane 13
I.2.5. Influence des paramètres opératoires 13
I.2.5.1. Pression 13
I.2.5.2. Température 14
I.2.5.3. Taux de récupération "Taux de conversion" 14
I.3. Interactions et propriétés de séparation 14
I.3.1. Encombrement stériques (effet du tamisage) 15
I.3.1.1. État physique de la membrane 16
I.3.1.2. Taille moléculaire 18
I.3.2. Répulsion électrostatique (effet de charge) 20
I.3.2.1. Charge de la surface 20
I.3.2.2. Charge de soluté 21
I.3.2.3. Influence du pH 21
I.3.3. Interaction hydrophobe/adsorption 21
I.3.3.1. Hydrophobicité de la membrane 21
I.3.3.2. Hydrophobicité des solutés 22
I.4. Applications des procédés d'OI et de NF 23
I.5. Avantages et inconvénients des procédés d'OI et de NF 24
I.6. Revue des travaux de la rétention des pesticides lors d'OI/NF 24
I.7. Conclusion 26

Chapitre II. QSPR-Réseaux de neurones artificiels

II.1. Introduction 27
II.2. Méthodes d'apprentissages statistiques 28
II.3. Relations quantitatives structure-Activités (QSAR) 28
II.3.1. Outils de QSAR 29
II.3.2. Données biologiques 29
II.3.3. Descripteurs moléculaires 29
II.3.4. Formulation mathématique des QSAR 31
II.4. Réseaux de neurones artificiels (RNA) 32

II.4.1. Histoire des réseaux de neurones artificiels 32
II.4.2. Neurone biologique 33
II.4.3. Neurone formel 34
II.4.4. Architecture d'un réseau de neurones 36
II.4.5. Perceptron multicouches 38
II.4.6. Types d'apprentissage des réseaux de neurones 39
II.4.7. Réseaux de neurones les plus utilisés 40
II.4.8 Conception d'un modèle neuronal 41
II.4.9. Détermination des entrées/sorties du réseau de neurones 42
II.4.10. Construction de la base de données et choix des échantillons 42
II.4.11. Définition de la structure du réseau 43
II.4.12. Apprentissage 44
II.4.13. Evaluation des performances du réseau de neurones 46
II.5. Conclusion 47

Chapitre III. Matériels et méthodes

III.1. Introduction 48
III.2. Collecte des données, prétraitement et analyse 48
III.3. Développement des modèles QSAR 54
III.4. Analyse de sensibilité par la méthode de "poids" 56
III.5. Diagnostic des points aberrants et domaine d'applicabilité du modèle neuronal optimisé 58
III.6. Conclusion 59

Chapitre IV. Résultats et discussion

IV.1. Introduction 60
IV.2. Modèles QSAR-RN 60
IV.2.1. Performances des modèles QSAR-RN 60
IV.2.2. Expressions mathématiques des modèles neuronaux optimisés 66
IV.2.3. Analyse de sensibilité par la méthode de "poids" 69
IV.3. Diagnostic des points aberrants et domaine d'applicabilité du modèle neuronal Optimisé 70
IV.4. Conclusion 72

Conclusion générale et perspectives 73
Références bibliographiques 75
Annexe A 79
Annexe B 82
Annexe C 85

Symbole	Signification
ART	Adaptive Resonance Theory
F	Fonction d'activation
HB	Molécule est hydrophobe
HL	Molécule est hydrophile
I	Composante hydrophobique
Ii	Composante électronique
Iii	Composante stérique (ES, RM)
J	Nombre de descripteurs utilisés.
L'IUPAC	International Union of Pure and Applied Chemistry
MF	Microfiltration
MWCO	Poids moléculaire de coupure
NF	Nano filtration
OI	Osmose inverse
PMC	Perceptron Multicouches
QSAR	Relations quantitatives structure-Activités
QSTR	Relations quantitatives structure- Toxicités
QSPR	Relations quantitatives structure-proprietés
RFR	Réseaux à fonction de bases radiales
RNA	Réseaux de neurones artificiels
SOM	Cartes auto-organisatrices de Kohonen
TFC	Thin-film composite
Tr	Taux de rejet
UF	Ultrafiltration
π, ES, RM et σ	Paramètres fonctions des substituants
b_0	Constante
b_i	Un coefficient
D_I	Descripteur moléculaire
w_0	entrée constante appelée "biais"
w_i	Poids de connexions

Lecture	Désignation	Unité
π	Pression osmotique de la solution	(Pa)
C	Concentration de la solution	(mol.m^{-3})
C_{sa}	Concentration d'un soluté dans l'alimentation	(kg/m^3).
C_{sp}	Concentration d'un soluté dans le perméat	(kgm^3)
DCO	Demande Chimique en Oxygène	(mg.L^{-1})
Depth	Profondeur moléculaire	(nm)
EqWidth	Largeur équivalente moléculaire	(nm)
Length	Longueur moléculaire	(nm)
$logk_{ow}$	Logarithme du coefficient de partage octanol-eau	(-)
log 1/C	Logarithme inversede l'activité	(-)
MW	Masse molaire	(g.mol^{-1})
Q_A	Débit volumique d'Alimentation	(m3/h)
Q_P	Débit volumique de Perméat	(m3/h)
R	Constante des gaz parfait	(8.31 J $mol^{-1}k^{-1}$)
REQM	Racine de l'Erreur Quadratique Moyenne	(%)
T	Température	(k)
y	Taux de récupération	(-)

Lettres Grecques

Lecture	Désignation	Unité
α	Pente de la droite du meilleur ajustement linéaire	(-)
β	Ordonnée à l'origine de la droite du meilleur ajustement linéaire	(-)

Exposants

Lecture	Désignation
cal	Calculée
exp	Expérimentale
H	Cachée (Hidden)
I	Entrée (Input)
O	Sortie (Output)

Chapitre I. Rétention des pesticides par les membranes

Tableau I.1.	Classification IUPAC des pores suivants leur taille.	06
Tableau I.2.	Caractéristiques de procédés de séparations.	12
Tableau I.3.	Avantages et inconvénients de procédés d'OI et NF.	26

Chapitre II. QSPR-Réseaux de neurones artificiels

		35
Tableau II.1.	Fonctions de transfert usuelles, *f(x)*.	
Tableau II.2.	Résume les algorithmes d'apprentissage par rétro-propagation implanté dans la boite à outils des RN du programme MATLAB®.	45

Chapitre III. Matériels et méthodes

Tableau III.1.	Taille de la base de données et le nombre de composés organiques utilisés dans ce travail.	48
Tableau III.2.	Logiciels utilisés pour calculer LogD et le moment dipolaire et les descripteurs géométriques.	49
Tableau III.3.	Entrées et sortie utilisées pour le développement des modèles QSAR (QSAR1, QSAR2 etQSAR3).	50
Tableau III.4.	Analyse statistique des données d'entrée et de sortie (BD1).	51
Tableau III.5.	Analyse statistique des données d'entrée et de sortie (BD2).	52
Tableau III.6.	Analyse statistique des données d'entrée et de sortie (BD3).	52

Chapitre IV. Résultats et discussion

Tableau IV.1.	Structure des modèles RN optimisés.	61
TableauIV.2.	Vecteurs de régression linéaire issue de l'équation linéaire : [équation linéaire : $y^{cal} = \alpha y^{exp} + \beta$, $\boldsymbol{\alpha}$(pente), $\boldsymbol{\beta}$(ordonnée à l'origine), $\boldsymbol{R}$(coefficient de corrélation)] et la racine de l'erreur quadratique moyenne ($\boldsymbol{REQM}$).	64
Tableau IV.3.	Poids et biais du QSPR-RN $_{optimal}$.	68

Annexe A

Tableau 1	Liste de 23 pesticides utilisées pour développer les modèles QSAR (RN).	79

Annexe C

Tableau 1	Matrice des coefficients de corrélation pour la Base des données (BD3)	86

Chapitre I. Rétention des pesticides par les membranes

Figure I.1.	Principe de fonctionnement d'une membrane sélective	05
Figure I.2.	Membrane symétrique poreuse	06
Figure I.3.	Membrane symétrique dense	07
Figure I.4.	Représentation schématique d'une membrane composite	07
Figure I.5.	Représentation schématique d'un module tubulaire	08
Figure I.6.	Représentation schématique d'un module à fibre creuses	09
Figure I.7.	Représentation schématique d'un module spiralé	09
Figure I.8.	Gamme de rétention de différentes particules par les procédés membranaires	10
Figure I.9.	Principe des phénomènes d'osmose inverse	11
Figure I.10.	La molécule de polyamide et ses équilibres acido-basiques	15
Figure I.11.	Illustration de principe de détermination du seuil de coupure d'une membrane, la rétention en fonction de la masse molaire (défini ici pour un taux de rejet de 0.90)	17
Figure I.12.	Calcul des dimensions de l'octanol (Length, Width et Depth) de la structure 3D	20
Figure I.13.	Mesure de l'angle de contact par la méthode de la goutte déposée	22
Figure I.14.	Nombre de publications traitant la rétention des molécules organiques par (a) : nanofiltration (b) osmose inverse	25

Chapitre II. QSPR-Réseaux de neurones artificiels

Figure II.1.	Représentation d'un neurone biologique	33
Figure II.2.	Neurone formel	34
Figure II.3.	Mise en correspondance neurone biologique / neurone artificiel par Claude Touzet	36
Figure II.4.	Réseaux de neurones les plus utilisés	37
Figure II.5.	Structure d'un réseau de neurones non bouclé.	38

Chapitre III. Matériels et méthodes

Figure III.1.	Rétentions expérimentales en fonction de poids moléculaires : (a) BD1, (b) BD2, (c) BD3.	54
Figure III.2.	Organigramme de développement et l'optimisation de l'architecture d'un QSPR-RNA.	55
Figure III.3.	Organigramme de la méthode de "poids".	57

Chapitre IV. Résultats et discussion

Figure IV.1.	Coefficient de corrélation pour chaque model neuronal "QSPR-RN1, QSPR-RN2 et QSPR-RN3" : (a) division 01, (b) division 02 et (c) division 03.	61
Figure IV.2.	Comparaison entre les valeurs expérimentales et les valeurs calculées pour QSPR-RNoptimal : (a) la phase d'apprentissage, (b) la phase de test, (c) la phase totale.	64
Figure IV.3.	Performances des modèles neuronaux (QSAR-RN*1*, QSAR-RN*2* et QSAR-RN*3*) pour : (a) la phase d'apprentissage, (c) la phase de test, (d) la phase totale.	65

Figure IV.4. Représentation schématique du modèle neuronal "QSPR-RNoptimal" optimisé. 66
Figure IV.5. Architecture finale du réseau de neurone optimal de type feedforwardpour la prédiction de la rétention des pesticides par les membranes d'OI/NF. 69
Figure IV.6. Présentation des pourcentages de l'importance relative (*IR*) de chaque variable d'entrée sur la rétention des pesticides lors d'OI/NF 70
Figure IV.7. Diagnostic des données potentiellement suspectes et domaine d'applicabilité du modèle neuronal optimal : (a) pour la phase test et (b) pour la phase totale. 71

Annexe B

Figure 1. Capture d'écran du site intitulé Chemspider. 82
Figure 2. Capture d'écran des logiciels : (a) ChemBioDraw2D et (b) ChemBio3D. 83
Figure 3. Capture d'écran de logiciel Hyperchem 83
Figure 4. Capture d'écran de logiciel STATISTICA. 84
Figure 5. Capture d'écran de logiciel MATLAB 2010. 84

Un pesticide est défini comme "une substance ou une association de substances destinée à repousser, détruire ou combattre les ravageurs, y compris les vecteurs de maladies humaines ou animales, ou toutes autres espèces végétales ou animales nocives ou gênantes au cours de la production, la transformation ou l'entreposage des produits agricoles". Considérés comme "tueurs de fléaux" par certains étymologistes, ce sont des molécules dont les propriétés toxiques permettent de lutter contre des organismes nuisibles à l'homme et/ou à son environnement (Hennia 2019). L'utilisation généralisée des pesticides en agriculture a été créditée de l'augmentation spectaculaire des rendements des cultures en réduisant les menaces naturelles et est devenue une partie intégrante des pratiques agricoles. L'utilisation sanitaire des pesticides a été, elle aussi, couronnée de succès à la suite de l'éradication de certaines maladies. Par ailleurs, les pesticides ayant des propriétés antiseptiques et désinfectantes ont eu leur part d'utilisation en médecine. La gamme des pesticides disponibles s'est alors rapidement diversifiée et leur usage a connu un très fort développement au cours des décennies suivantes. Petit à petit, du fait de leur facilité d'emploi et de leur efficacité, ils se sont imposés dans les usages domestiques (anti-moustiques par exemple), dans le traitement de la voierie publique, des plantes ornementales des voies ferrés, des routes, des espaces verts et des jardins. Durant la seconde moitié du vingtième siècle, l'utilisation des pesticides a doublé tous les dix ans. De plus, la résistance acquise par certaines espèces vis-à-vis des pesticides a contraint les utilisateurs de ces produits à augmenter la quantité à utiliser afin que l'efficacité d'action soit optimale d'une part et à la diversification des molécules actives (recherche de nouveaux pesticides plus actifs) d'autre part (Hamdache 2014).

Le problème de la contamination des eaux par les pesticides et autres micropolluants organiques inquiète de plus en plus les producteurs d'eau potable. La diversité de ces produits et le niveau des concentrations atteint dans les eaux brutes ces dernières années exigent des méthodes d'élimination efficaces qui garantissent en bout de chaîne de traitement des concentrations résiduelles inférieures aux limites maximales fixées par les instances européennes : 0.1 ug/L pour chaque substance prise individuellement et 0.5 ug/L pour les totaux pesticides plus métabolites. Pour éliminer les pesticides de l'eau, plusieurs techniques sont utilisées, seules ou en association, avec plus ou moins de succès. Le développement de nouvelles membranes a ouvert des perspectives d'application dans le domaine du traitement de l'eau potable par osmose inverse basse pression et par nanofiltration. En plus de ses capacités d'adoucissement de l'eau, cette dernière peut aussi bien éliminer la matière organique dissoute, que la couleur et certains pesticides. Mais à cause d'une consommation élevée d'énergie, cette technique a vu son essor freiné (Boussahel et al. 2022).

Les méthodes de séparation membranaire sont un procédé de séparation de fluides utilisant comme agent séparant une membrane synthétique qui est une couche mince de matière. L'épaisseur d'une membrane peut varier de 100 nm à un peu plus de 1 cm. Elle permet l'arrêt ou le passage sélectif de certaines substances dissoutes ou non dans un mélange, entre les deux milieux qu'elle sépare. La partie du mélange retenue par la membrane est appelée rétentat (ou concentrat) alors que celle qui traverse cette dernière est appelée perméat. La séparation se fait sous l'action d'une force motrice de transfert selon un mécanisme de séparation défini. Les caractéristiques des membranes sont déterminées par deux paramètres : la perméabilité et la sélectivité. On distingue quatre techniques répertoriées sous les noms de microfiltration (MF), ultrafiltration (UF), nanofiltration (NF) et osmose inverse (OI). La micro-, l'ultra- et la nanofiltration mettent en jeu des membranes possédant une structure poreuse tandis que les membranes d'osmose inverse sont des matériaux denses. La microfiltration est utilisée pour séparer des espèces dont la taille est de l'ordre du micromètre (cellule animale, levures, bactéries…), l'ultrafiltration pour séparer des espèces de taille comprise entre ~ 2 et 100 nm (peptides, protéines, virus…) et la nanofiltration pour traiter les espèces de taille inférieure à 2 nm (sucres, colorants, sels…). Tout comme la nanofiltration, l'osmose inverse, permet de filtrer des solutions contenant des espèces de très faibles masses molaires. Elle est généralement utilisée lorsqu'une rétention totale des solutés est souhaitée (Bouranene 2008).

La littérature met en lumière plusieurs points. D'une part, la nanofiltration et l'osmose inverse sont des procédés de plus en plus répandus et de nombreuses études indiquent qu'il permet de traiter différents types d'effluents en retenant les pesticides. D'autre part, le rôle complexe et encore mal cerné des nombreuses interactions qui peuvent s'établir entre les solutés, l'eau et la membrane. De nombreux modèles de transferts existent mais ils ne décrivent ou n'expliquent pas toujours correctement le procédé et les résultats obtenus dans le cas du transfert de molécules organiques nécessitent des modifications importantes.

La modélisation est une autre approche permettant de diminuer le nombre d'expériences à effectuer pour déterminer les paramètres optimaux pour la rétention des pesticides lors de nanofiltration et d'osmose inverse. Si elle représente bien le procédé, les influences des paramètres optimaux pourront être évaluées sans aucune expérience et n'auront qu'à être validés sur le site. 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 des nouvelles méthodes de calcul et de modélisation telles que les réseaux de neurones artificiels "RNA", les algorithmes génétiques et la logique floue. L'efficacité, la fiabilité et la

robustesse de ces méthodes par rapport aux difficultés de l'utilisation des approches classiques, ont été démontrées par leurs nombreuses applications dans le domaine du génie des procédés. Le principal avantage des RNA, comparé aux modèles de connaissance classique, est leur capacité à modéliser ces procédés sans avoir une connaissance approfondie de la phénoménologie mise en jeu. Un autre avantage important des RNA. Bien que les réseaux de neurones artificiels (RNA) soient très appliqués dans le domaine du génie des procédés en raison de la précision des résultats qu'ils produisent et de la simplicité relative de la mise en œuvre. Peu de travaux de recherche utilisent les RNA comme moyen de calcul des phénomènes de transports par d'osmose inverse et de nanofiltration lors de la rétention des pesticides (Ammi 2016).

L'objectif de ce travail est l'application des QSPR-RN de type feedforwardet pour modéliser la rétention des pesticides par les membranes d'OI/NF. Ainsi, nous avons organisé ce mémoire en quatre chapitres :

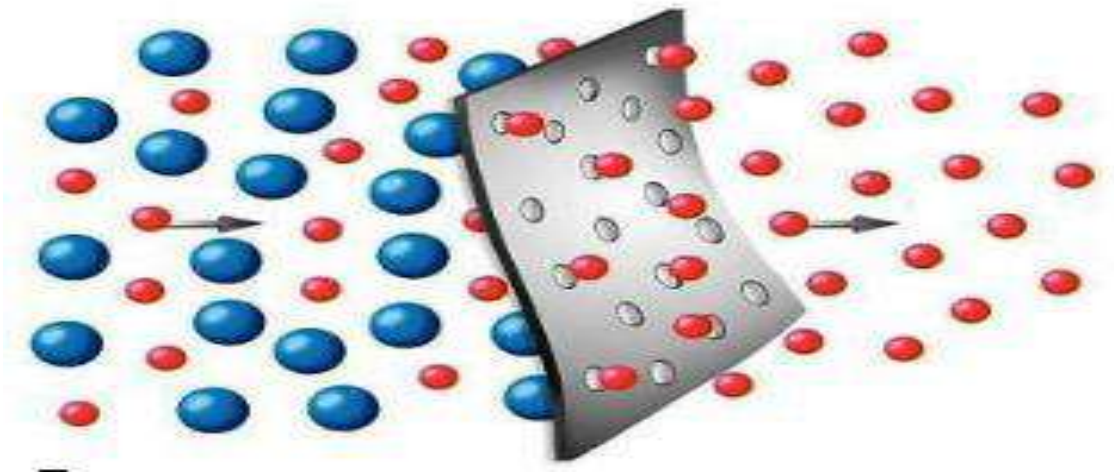
- ✓ Le premier chapitre est consacré à la description de la technologie membranaire, on donnera dans un premier temps un aperçu sur les généralités sur les membranes, ensuite les interactions et les propriétés de séparation, puis les applications et les avantages et une revue des travaux de la rétention des pesticides lors d'OI/NF.
- ✓ Le deuxième chapitre est dédié à la présentation des modèles de réseaux de neurones artificiels basés sur la relation quantitative structure-activité "QSAR-RN".
- ✓ Nous présentons dans le troisième chapitre les matériels et les différentes procédures de modélisations utilisées. Il comptera trois axes : dans le premier axe, on examinera les données collectées, prétraitement et analyse. Le second axe sera consacré aux techniques les plus courantes pour établir des modèles QSPR-RN. Le troisième axe traitera l'analyse de sensibilité basée la méthode de "poids". Le quatrième et dernier axe présentera le diagnostic des points aberrants et domaine d'applicabilité du modèle neuronal optimisé.
- ✓ Dans le quatrième chapitre, les différents résultats prometteurs obtenus sont discutés, il est organisé en quatre parties : dans une première partie, on présentera les performances des modèles QSPR-RN, ensuite l'extraction de formule mathématique à partir de ce modèle, par la suite, on présentera les résultats de l'analyse de sensibilité par la méthode de "poids", Puis on y discutera sur un diagnostic des points aberrants et domaine d'applicabilité du modèle neuronal optimisé.

Les conclusions tirées de ce travail ainsi que les études qui peuvent être envisagées à l'avenir sont présentées à la fin de ce mémoire.



Chapitre I

Rétention des pesticides par les membranes



I.1. Introduction

La filtration membranaire est de plus en plus utilisée comme procédé de séparation dans de nombreux domaines notamment dans le cycle de l'eau (potabilisation de l'eau, traitement des effluents, réutilisation de l'eau, adoucissement, dessalement,…). La filtration membranaire est basée sur l'application d'une différence de pression qui permet le transfert du solvant à travers une membrane dont la taille des pores assure la rétention de solutés. Ces opérations, classées selon une taille des pores décroissantes, sont : la microfiltration (MF) l'ultrafiltration (UF), la nano filtration (NF) et l'osmose inverse (OI).

Les procédés de séparation membranaire ont pris leur essor dans les années 70 en particulier grâce au développement des matériaux et des techniques de fabrication des membranes qui offrent une meilleure rentabilité dans tous les termes de consommation énergétique par apport à des techniques de séparation conventionnelles telle que la distillation, des conditions de fonctionnement tolérables, un traitement souhaitable, et qu'elles favorisent le développement dans toutes les branches de l'industrie. Parmi eux, l'osmose inverse et nano filtration sont utilisées en traitement de séparation des différentes molécules organiques et inorganiques qui permettent une séparation conforme aux normes actuellement en vigueur (Ammi 2016).

Ce chapitre s'attache à faire les points sur les généralités des membranes d'osmose inverse et de nano filtration, les interactions et les propriétés de séparation des molécules organiques, les applications des procédés d'OI/NF, les avantages et les inconvénients des procédés d'OI/NF et une revue des travaux de la rétention des pesticides lors d'OI/NF**.**

I.2. Généralités

I.2.1. Définition de la membrane

Une membrane est une couche mince de la matière condensée ayant des propriétés physiques et mécaniques propres. La membrane permsélective est une interphase séparant deux milieux, et contrôlant le transport des constituants d'une manière spécifique. La partie du mélange retenue par la membrane est appelée retentât et ce qui traverse cette dernière est appelée perméat. La membrane peut être homogène ou composite, dense ou poreuse, isotrope ou anisotrope, solide ou liquide, ionique ou neutre. Il a fallu attendre les années 60 pour voir apparaitre les premières membranes organiques industrielles (Figure I.1). Il s'agissait de membrane hautement perméable dont la structure asymétrique permet d'obtenir à la fois une bonne sélectivité de séparation et une haute perméabilité aux solutés ioniques ou macro moléculaires (Hamani 2016).

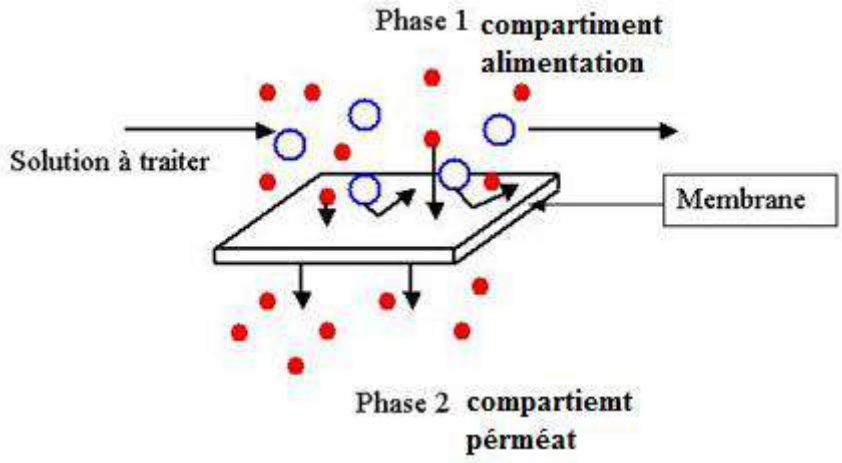


Figure I.1. Principe de fonctionnement d'une membrane sélective (https://culturesciences.chimie.ens.fr/thematiques/chimie-des-materiaux/chimie-des-surfaces/les-techniques-membranaires-differences-et).

I.2.2. Classification des membranes

La comparaison des membranes, selon un type de séparation donné, repose essentiellement sur la mesure de leurs performances (perméabilité au solvant, sélectivité ou rétention, et durée de vie) dans les conditions réelles. On peut cependant estimer qu'une membrane est plus apte qu'une autre à la séparation, grâce à certains critères tels que la sélectivité, la perméabilité et la durée de vie. La classification des membranes peut se faire selon plusieurs critères tels-que : le mécanisme de séparation, la morphologie, la géométrie, la nature chimique, la nature physique (Hamani 2016).

I.2.2.1. Nature chimique

- **Membrane organique :** sont fabriquées à partir de polymères organiques (acétate de cellulose, polysulfones, polyamides, etc..). Les qualités de ces matériaux leur confèrent une grande adaptabilité aux différentes applications. Environ 90% des membranes d'ultrafiltration et les microfiltrations sont constitués de membranes organiques.
- **Membrane minérale :** sont à la base, de silice, de carbone, d'oxyde de graphite et de métaux frittés. Ces matériaux possèdent une grande résistance physicochimique, mécanique et thermique. Leur durée de vie excède de trois fois ces telle des membranes organiques mais leur cout est sensiblement plus élevé (Ammi 2016).

I.2.2.2. Porosité

- **Membrane poreuse :** ce type de membrane se compose de pores dont la dimension est supérieure aux dimensions moléculaires. L'IUPAC (International Union of Pure and Applied Chemistry) distingue trois catégories de pores selon leurs tailles (Tableau I.1).

Tableau I.1. Classification IUPAC des pores suivants leur taille (Hamani 2016).

Dénomination	**Micropore**	**Mésopore**	**Macropore**
Taille de pore	<2 nm	2-5 nm	>50nm

- **Membrane dense :** ce terme signifie une absence de porosité de dimension microscopique, ces membranes ne présentent aucune porosité. Elles consistent en un film dense à travers lequel le perméat est transporté par diffusions sous l'effet d'une grandeur de potentiel électrique, de concentration ou d'une force de pression. Si la solubilité diffère, les membranes denses peuvent séparer des composés de taille voisine (Hamani 2016).

1.2.2.3. Morphologie

- **Membrane symétrique (isotrope) :** la membrane symétrique est composée d'un seul matériau ayant une composition sensiblement uniforme dans toute leur épaisseur. Elle peut être dense (absence de pores) ou poreuse (Figure I.2 et I.3). De telles membranes forment la couche active sur toute leur épaisseur (Ammi 2016).

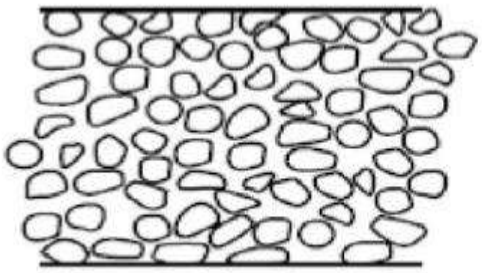


Figure I.2. Membrane symétrique poreuse(Hamani 2016).



Figure I.3. Membrane symétrique dense (Hamani 2016).

- **Membrane asymétrique ou anisotrope :** la membrane asymétrique ou **anisotrope a une porosité** variable dans son épaisseur. Elle présente deux parties distinctes :
  - La peau ou couche active, de très faible épaisseur et porosité, assure la sélectivité de la membrane.
  - La sous-couche poreuse ou support, garantit une résistance mécanique : elle représente en général la plus grande partie de l'épaisseur de la membrane et sa porosité interne est très élevée entre 60 et 80 %.
- **Membrane composite :** Elle est constituée d'un assemblage de deux ou plusieurs couches de matériaux différents et présentant différentes porosités. La membrane composite comporte deux types de couches : les couches actives (couche filtrant) et le support assurant de bonnes propriétés mécaniques (Figure I.4) (Ammi 2016).

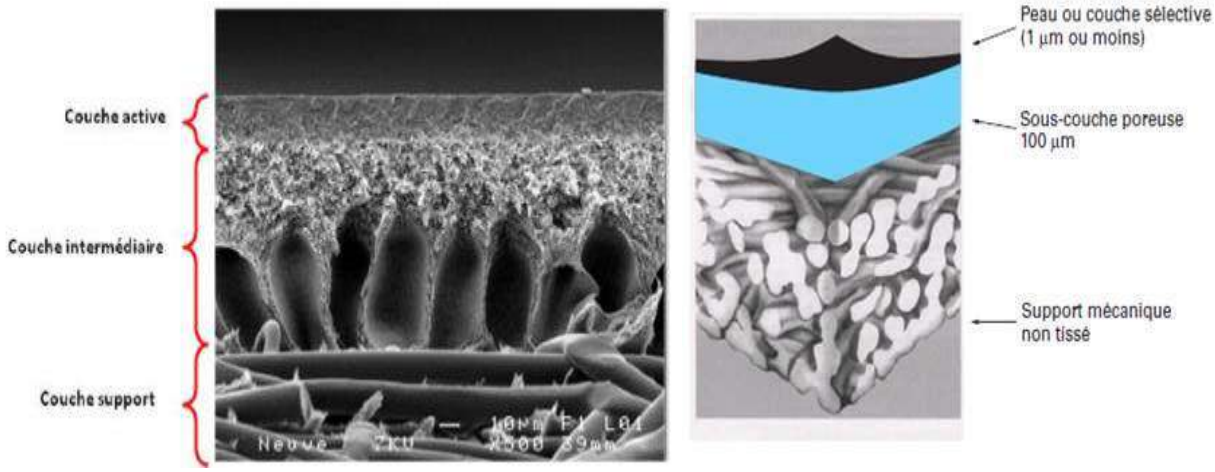


Figure I.4. Représentation schématique d'une membrane composite (Chouachi et Kaddouri 2021).

1.2.2.4. Classification d'après la géométrie de préparation des membranes

Les membranes peuvent être classées en fonction de la forme sous laquelle elles sont fabriquées :

- ❖ Membranes planes ;
- ❖ Membranes spirales ;
- ❖ Membranes cylindriques ;
- ❖ Membranes tubulaires (ayant le diamètre interne plus grand que 3 mm) ;
- ❖ Membranes creuses (ayant le diamètre plus petit que 3 mm).

La forme des membranes conditionne la manière dont elles sont assemblées pour fabriquer des modules (Ammi 2016).

- **Modules membranaires**

La séparation se déroule dans un module de filtration pouvant prendre plusieurs formes :

- **Module plan :** le système utilise des membranes sous forme de feuilles. Il s'agit du type de configuration la plus simple. La surface filtrante reste toutefois faible par rapport à l'encombrement engendré.
- **Module tubulaire :** Il est constitué par des faisceaux de tubes de 4 à 25mm de diamètre intérieur. Ils sont basés sur une technologie simple, une consommation élevée en énergie est nécessaire pour une très faible surface d'échange par unité de volume (compacité réduite) (Figure I.5).

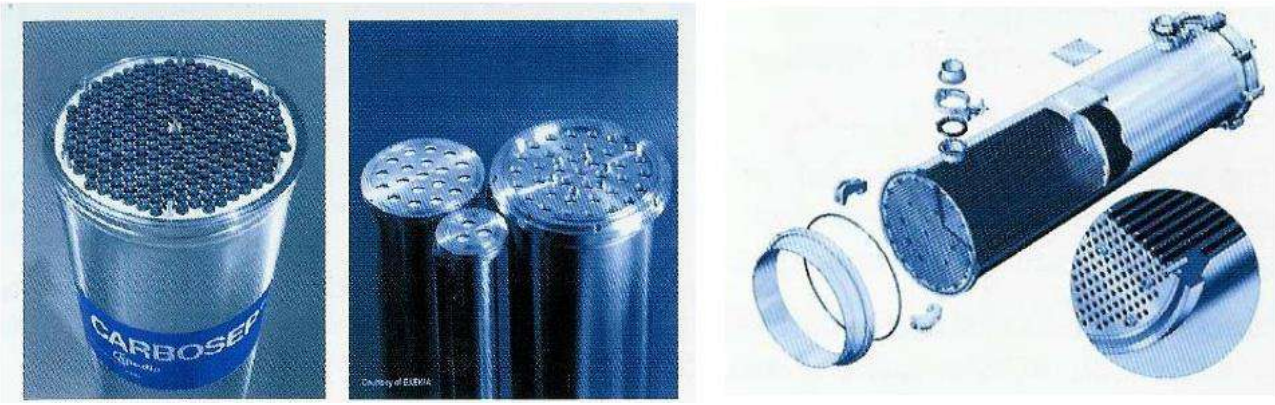


Figure I.5. Représentation schématique d'un module tubulaire (Ammi 2016).

- **Fibre creuse :** Elle rassemble un ensemble de fibres creuses de diamètre externe inférieur à 1 millimètre en un faisceau. Cette configuration leur procure la plus forte densité d'écoulement par module (Figure I.6).

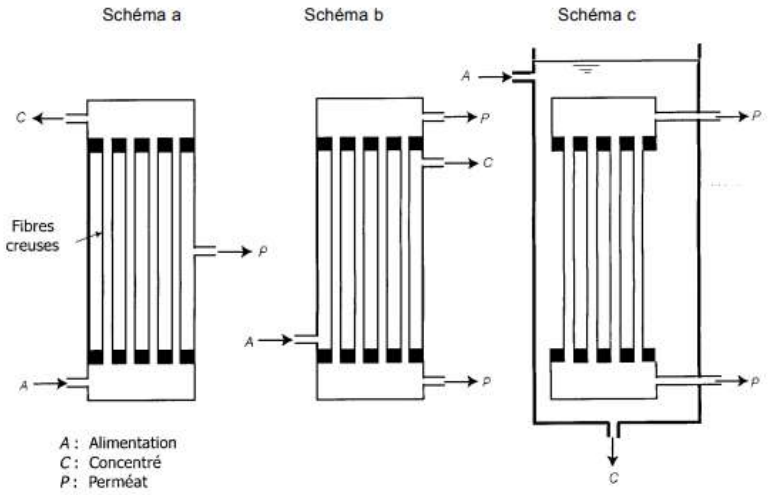
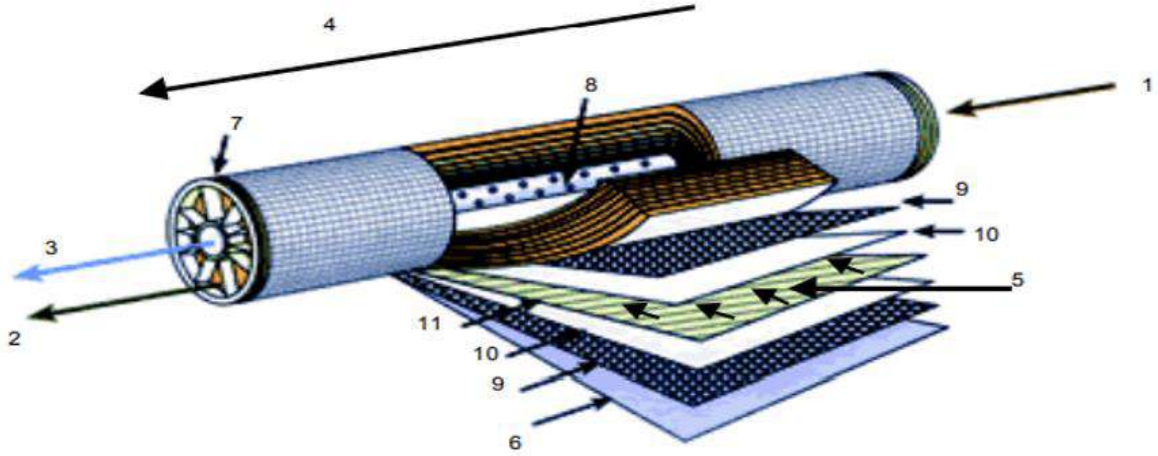


Figure I.6. Représentation schématique d'un module à fibres creuses (Berland et Juery 2002).

- **Module spiralé :** Au sein des modules spirales, une membrane plane est enroulée sur elle-même autour d'un tube poreux qui recueille le filtrat. On obtient ainsi un cylindre multi-couches où le perméat s'écoule selon un chemin spiralé vers le tube poreux tandis que l'alimentation circule axialement dans les canaux.



1 Entrée d'eau
2 Sortie de concentrat
3 Sortie de perméat
4 Sens d'écoulement de l'eau brute
5 Sens d'écoulement du perméat
6 Matériau de protection
7 Joint d'étanchéité entre module et enveloppe
8 Perforations collectant le perméat
9 Espaceur
10 Membrane
11 Collecteur de perméat

Figure I.7. Représentation schématique d'un module spiralé (Berland et Juery 2002).

1.2.2.5. Classification d'après le mécanisme de séparation

La classification la plus répandue des membranes d'après le mécanisme de séparation est la suivante :

a. Membranes poreuses (MF, UF, NF), où les effets de tamisage et de friction jouent des rôles importants, comme en NF, les forces de surface peuvent jouer un rôle important. Les pores peuvent être subdivisés en plusieurs catégories :

- Macropores-plus grands que 50 nm (MF) ;
- Mésopores-2 à 50 nm (UF) ;
- Micropores-plus petits que 2 nm (NF).

b. Membranes denses (perméation de gaz, pervaporation, osmose inverse (OI) ; ces membranes peuvent être considérées comme des milieux denses où la diffusion des espèces a lieu dans les volumes libres situés entre les chaînes moléculaires du matériel de la membrane ;

c. Membranes échangeuses des ions : Ces membranes sont considérées comme un type spécial de membranes non poreuses, consistant dans des gels denses, ayant une charge positive (échangeuses des anions) ou négative (échangeuses des cations).

Il a noté que la communauté scientifique ne s'entende pas sur cette classification considèrent que les membranes d'OI sont des membranes poreuses. La figure I.8 montre schématiquement la correspondance entre les opérations de filtration par membrane, la taille des pores et les composés retenus (Ammi 2016).

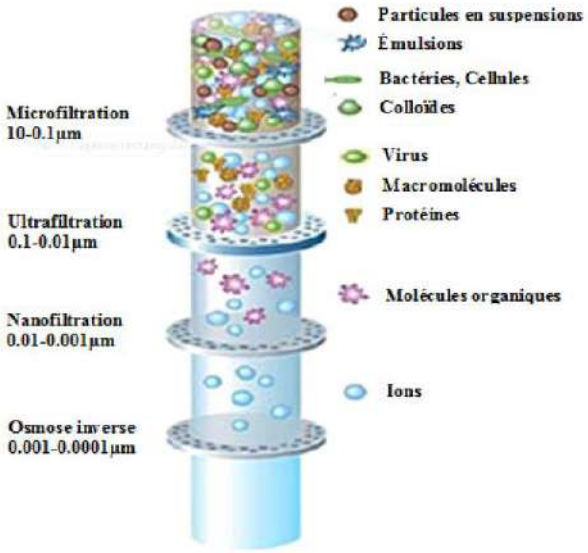


Figure I.8. Gamme de rétention de différentes particules par les procédés membranaires (Ammi 2016).

1.2.3. Osmose inverse et nano filtration

L'**osmose inverse** est un procédé de filtration tangentielle qui permet l'extraction d'un solvant, le plus souvent l'eau par perméation sélective à travers une membrane dense sous l'action d'un gradient de pression (Figure I.9).

- Elle s'oppose au phénomène naturel d'osmose qui tend à transférer le solvant d'une solution diluée vers une solution concentrée mise en contact par une membrane sélective sous l'action du gradient de concentration.
- Lorsqu'une pression est appliquée sur le compartiment le plus concentré, le flux de solvant diminue et tend vers une valeur nulle pour une pression égale à la pression osmotique de la solution.
- Lorsque la pression appliquée est supérieure à cette pression osmotique, le flux s'inverse : C'est le phénomène d'osmose inverse (Ammi 2016).

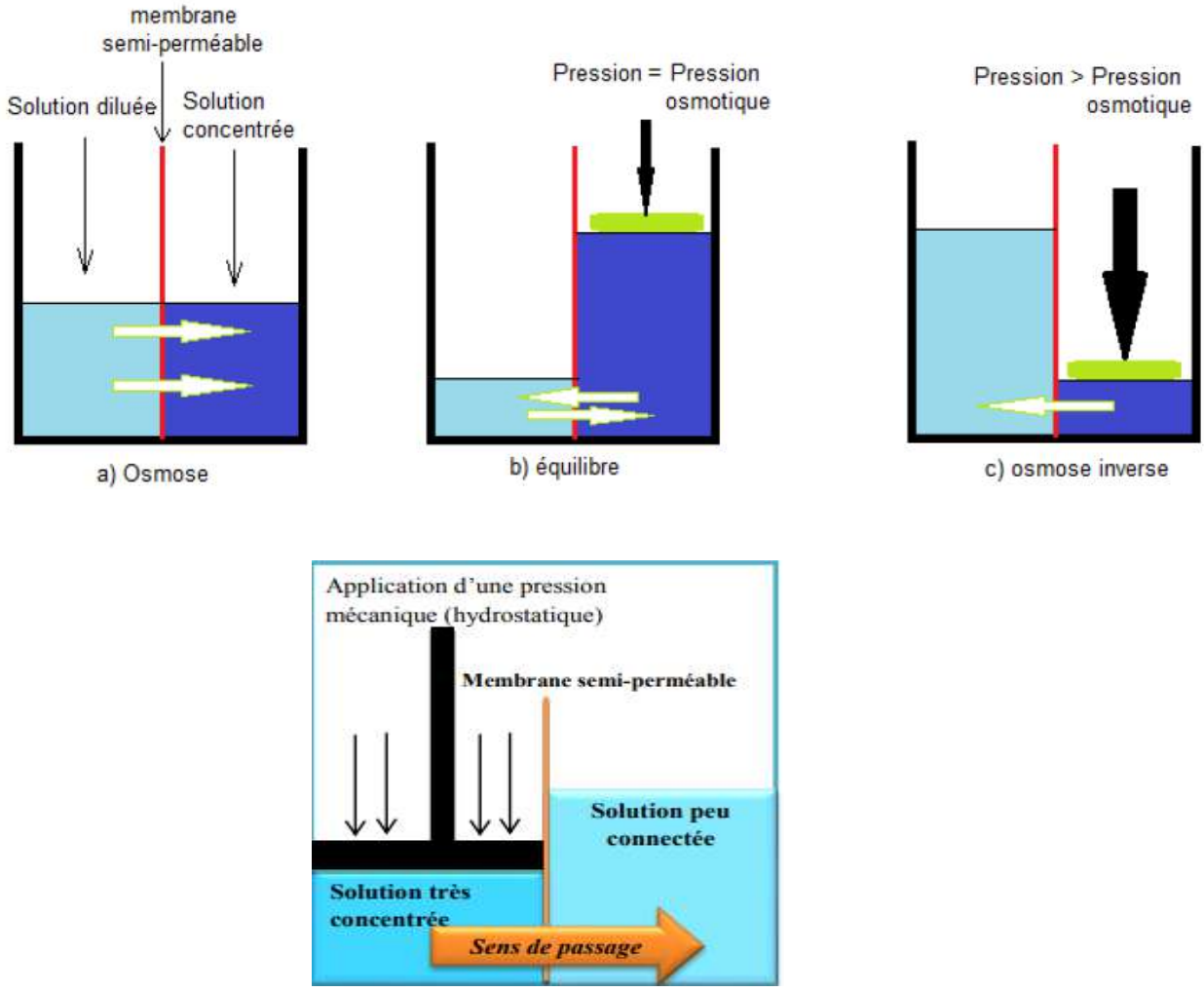


Figure I.9. Principe des phénomènes d'osmose inverse (Ammi 2016 et Chouachi et 2021).

Pour les solutions diluées, considérées comme thermodynamiquement idéales, la pression osmotique est calculée à partir de l'équation de Van't Hoff :

$$\pi = CRT \qquad (I.1)$$

π : pression osmotique de la solution (Pa)

C : concentration de la solution (mol.m^{-3})

R : constante des gaz parfait (8.31 J $mol^{-1}k^{-1}$)

T : température (K) (Ammi 2016).

La **nano filtration (NF)** est une technique intermédiaire entre l'osmose inverse et l'ultrafiltration, elle permet de retenir des molécules de faible poids moléculaire (MW ≥300 g.mol^{-1}) et des ions divalents, sans modifier son affinité vis à vis les ions monovalents. Les caractéristiques d'utilisation sont les suivantes :

- Pression de travail 10 à 40 bars,
- Diamètre des pores de la membrane 1 nm,
- Débit volumique 50 à 100 l.h−1mm^{-2}.

Les membranes de nanofiltration sont en fait des membranes classiques greffées par des groupements hydrophiles chargés négativement. Elles ont été testées avec succès en traitement des eaux pour éliminer à la fois les ions divalents (Ca^{+2},Mg^{+2}) et certains pesticides organochlorés. Les sels ionisés multivalents (calcium, magnésium, aluminium, sulfates…) et les composés organiques non ionisés de masse molaire supérieure à environ 250 g/mol sont, par contre, fortement retenus. Ces membranes ont des seuils de coupure entre 300 et 1000 g/mol (Ammi 2016).

Le tableau I.2 donne les caractéristiques des différents procédés membranaires à gradient de pression.

Tableau I.2. Caractéristiques de procédés de séparations (Nechad 2009).

Procédés	**Origine de la sélectivité**	**Force motrice (bar)**	**Diamètre des pores (μm)**	**Seuil de coupure (kDa)**	**Consommation énergétique (kWh/m^3)**	**Procédés concurrents**
Microfiltration MF	Différence de taille entre particules ou molécules à séparer	Pression 0.1 à 3	0.1 à 10	-	1-10	Filtration Centrifugation

Ultrafiltration UF	Déférence de taille et de charge entre particules a séparé	Pression 3 à 10	0.01 à 0.1	1-300	1-10	Evaporation Distillation Echange d'ion
Nanofiltration NF	Déférence de taille et de charge entre particules ou molécules à séparer	Pression 10 à 50	~10^{-3}	0.2-1	5-50	Evaporation OI Echange d'ion
Osmose inverse OI	Différence de solubilité et de diffusion à travers la membranaire des molécules à séparer	Pression 30 à 100	Membrane dense	<0.2	10-200	Evaporation Distillation Echange d'ion

1.2.4. Sélectivité d'une membrane

La sélectivité d'une membrane est, en général, définie par le taux de rejet (taux de rétention) de l'espèce (sel, macromolécule, particule) que la membrane est censée retenir

$$Tr = \frac{Csa - Csp}{Csa} = 1 - \frac{Csp}{Csa} \qquad (I.2)$$

Où :

Csp: Concentration d'un soluté dans le perméat (kgm^3) ;

Csa: Concentration d'un soluté dans l'alimentation (kg/m^3).

I.2.5. Influence des paramètres opératoires

I.2.5.1. Pression

La pression de fonctionnement (Ptm) doit être supérieure à la pression osmotique. L'étude menée par Sridhar et al. 2002 a montré que pour une pression appliquée inférieure à 8.9 bar, aucun perméat n'est produit lors du traitement d'un effluent issu d'industrie d'huile végétale. Ainsi, la pression efficace (Ptm-Π) correspond à la pression qui conduit réellement à la production du perméat. Plus cette pression est élevée, plus le flux du perméat est important. De même, le taux de rétention augmente lorsque la pression augmente. D'après Lin et al. 2004, la DCO (Demande Chimique en Oxygène) du perméat passe de 6.7 à 3.1 mg.L $^{-1}$ pour des pression de 2.1 et 6.3 bar avec un taux de rétention augmente de 96.9% à 98.6%. Sridhar et al. 2002 ont montré que cette

augmentation du taux de rétention est due à l'augmentation de la pression transmembranaire ainsi que des flux du solvant et du flux du soluté. La concentration du soluté dans le perméat diminue. Par conséquent, le taux de rétention augmente (Ammi 2016).

I.2.5.2. Température

L'augmentation de la température engendre une diminution de la viscosité des solutions, et par conséquent une augmentation de la densité de flux du perméat, d'environ 3% par degré Celsius. Le taux de rétention diminue, la diffusivité de l'eau et des solutés augmente avec la température (Girard et Fukumoto 2000). Pour une pression donnée, le compactage de la membrane est d'autant plus important que la température est élevée. Il est donc important de vérifier que la membrane et le module aient une tenue suffisante à la température de fonctionnement choisie en général, la température maximale recommandée est de l'ordre de 50°C.

I.2.5.3. Taux de récupération "Taux de conversion"

Le taux de récupération "y" d'un système d'OI/NF est le rapport entre le débit de permeat "Q_P" et le débit d'alimentation "Q_A" (Chabane 2015) :

$$y = \frac{Q_P}{Q_A} \qquad (I.3)$$

L'augmentation du débit d'alimentation engendre une augmentation de la vitesse de circulation le long de la membrane. Plus elle est importante, plus la concentration aux abords de la membrane sont moins importantes aussi que la densité de flux du perméat au cours du temps (Cornelissen et al. 2001). Plus le taux de récupération est élevé, plus le flux diminue, aussi la concentration côté retentât augmente (Sagne 2008).

I.3. Interactions et propriétés de séparation

Dans le cas des molécules organiques, les propriétés de séparation ne sont pas facilement prédictibles. En effet, il a été montré qu'elles résultent de trois interactions qui affectent leur rétention lors d'OI/NF et peuvent être signalées comme suit : encombrements stériques (effet du tamisage) (Altalyan et al. 2016), répulsion électrostatique (effet de charge) et des interactions hydrophobes /adsorption (Sagne 2008). Rappelons tout d'abord la structure du polyamide, polymère majoritairement utilisé comme couche active des membranes d'osmose inverse et de nanofiltration (Figure I.10).

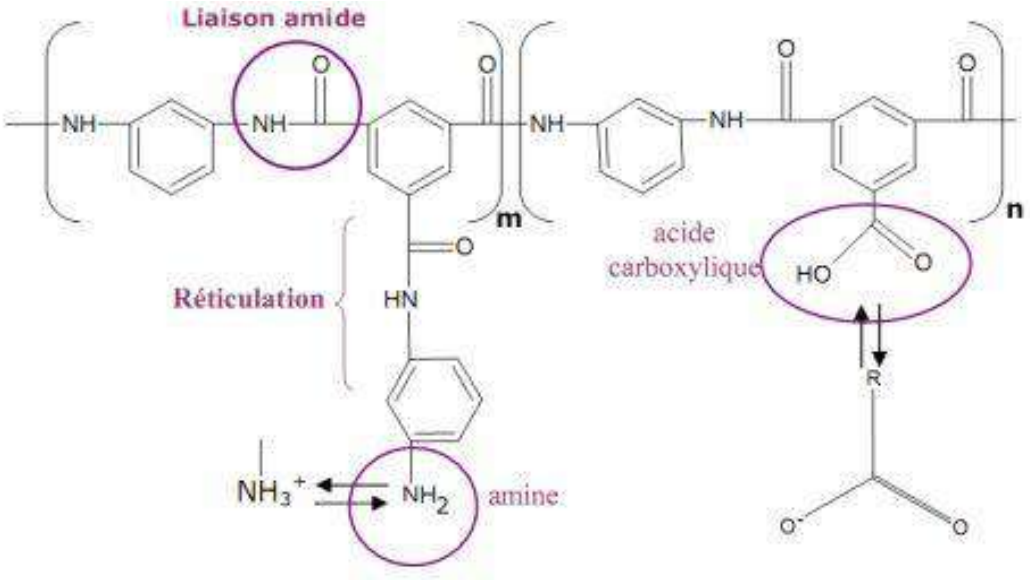


Figure I.10. La molécule de polyamide et ses équilibres acido-basiques (Sagne 2008).

De par cette structure, plusieurs types d'interactions peuvent se produire entre les molécules en solution et la surface de la membrane (Sagne 2008).

- Les fonctions acides carboxyliques et amines libres (non engagées dans une liaison de réticulation) sont susceptibles de s'ioniser, arborant ainsi respectivement une charge négative et positive et pouvant alors s'engager dans des liaisons électrostatiques fortes. Elles peuvent également participer à des liaisons hydrogènes.
- Les noyaux aromatiques peuvent développer des interactions π-π avec le nuage¬ électronique d'une autre molécule aromatique, pour former un nuage d'électrons global plus stable.
- Les liaisons amides peuvent développer des interactions de Lifshitz-van der Waals plus¬ faibles, par les doublets d'électrons libres des atomes d'azote et d'oxygène.

I.3.1. Encombrement stériques (effet du tamisage)

La taille moléculaire des molécules organiques et le seuil de coupure (Sc ou MWCO) de membrane ont confirmé l'importance des paramètres dimensionnels pour la rétention des molécules sur les membranes d'OI/NF (Yangali-Quintanilla et al. 2010a ; Dolar et al. 2011). Plusieurs auteurs ont sélectionné les descripteurs de taille moléculaire sachant que la rétention est liée aux mécanismes d'exclusion de taille entre membranes-solutés (Kiso et al. 2001a ; Ozaki et al. 2002 ; Schafer et al. 2003 ; Kimura et al. 2004 ; Nghiem et al. 2004 ; Yangali-Quintanilla et al. 2009b-2010b).

I.3.1.1. État physique de la membrane

Les propriétés physiques de la membrane comme le nombre de pores, la forme ainsi que la rugosité de la surface jouent un grand rôle dans la rétention des solutés.

La porosité est exprimée par la taille, la distribution de la taille et le nombre de pores effectifs dans la couche active. Le terme "pore" pour les membranes d'OI/NF correspond à un espace vide dans le polymère par lequel le transport du fluide peut avoir lieu sous l'action d'une force motrice. Dans une membrane dense, les pores sont des vides de moins de 0.5 nm de diamètre qui sont formés naturellement dans les polymères solides. Ces pores, ou défauts de densité, sont causés par des irrégularités de l'entremêlement des chaînes de polymères. Ils peuvent aussi être créés par les contractions locales des polymères cristallisables lors de leur cristallisation. Ces pores peuvent être circulaires ou non, fermés ou ouverts et forment un réseau continu d'interconnections. Ils sont susceptibles d'accueillir de petites molécules qui au cours de leur transport à travers ce pore interagissent fortement avec ses parois (Kosutic et al. 2000).

Les indicateurs de porosité de membrane peuvent être le rejet du sel ou le seuil de coupure, attribués par le fabricant en cours de l'expérience.

A. Seuil de coupure ou Seuil de rétention "Sc ou MWCO"

Les membranes d'OI/NF sont caractérisées par leur seuil de coupure. Il s'agit de la masse molaire de la plus petite molécule arrêtée à 90% (ou 95% selon le fournisseur) par la membrane dans des conditions opératoires données. Il est exprimé généralement en Dalton (unité équivalente à des grammes par mole). Le seuil de coupure renseigne l'utilisateur sur la plage d'utilisation de la membrane, mais sa valeur exacte ne doit être considérée qu'à titre indicatif. En effet, les molécules ont des géométries variées et il est possible que des espèces de masse molaire supérieure au seuil de coupure soient retenues par la membrane, de même que des espèces de masse moléculaire légèrement inférieure la traversent. Il est donc important de retenir que le seuil de coupure est défini par rapport à un soluté donné. On peut ainsi caractériser une membrane d'OI/NF, pour une séparation donnée, dans des conditions fixées, par une courbe de variation du taux de rejet d'un soluté en fonction de sa masse moléculaire (Figure I.11) (Lanteri et al. 2009).

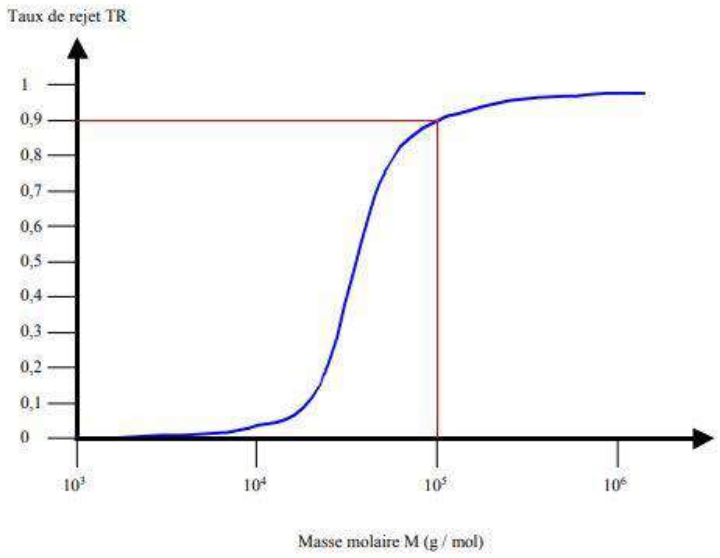


Figure I.11. Illustration du principe de détermination du seuil de coupure d'une membrane, la rétention en fonction de la masse molaire (défini ici pour un taux de rejet de 0.90) (Jean-Marc et al. 2002).

Plus le Sc d'une membrane est faible, plus elle peut retenir de petites molécules ou des colloïdes. Le Sc est relié principalement à la taille des pores de la membrane, il est influencé par (Anselme et Jacobs 1996) :

- ✓ la forme de la molécule à filtrer ;
- ✓ sa charge ;
- ✓ le pH ;
- ✓ la pression d'opération ;
- ✓ le flux de perméation ;
- ✓ la charge de la membrane.

B. Rejets des sels

• Rejet du sel de chlorure de sodium SR (NaCl)

Le premier paramètre donnant une indication sur la "porosité" de la membrane est son taux de rétention en chlorure de sodium (NaCl), mesuré par le fabricant dans des conditions standards, qui lui sont spécifiques. Il a été bien corrélé au taux de rétention de certains pesticides, de polysaccharides et d'alcools (Bellona et al. 2004). Étude d'Agenson et al. 2003 sur 36 composés organiques semi-volatils et volatils d'intérêt environnemental, de masse molaire (MW) comprise

entre 70 et 300 g/mol a montré qu'en général les taux de rétention les plus forts sont obtenus pour les membranes ayant un taux de rétention en NaCl le plus élevé.

Dans le cas de l'osmose inverse, le soluté de référence est souvent le chlorure de sodium (NaCl), la déminéralisation des eaux étant l'application la plus importante. Certaines membranes développées pour le dessalement de l'eau de mer ont un taux de rejet au chlorure de sodium de 99 % environ. D'autres, développées pour le dessalement des eaux saumâtres, présentent un taux de rejet au NaCl de 96 %. Dans le cas de l'ultrafiltration, l'efficacité de la membrane est, en général, caractérisée par le seuil de coupure (Cut-off en anglais). Il s'agit de la masse molaire (g/mol) correspondant à une rétention pratiquement totale (90 % le plus souvent) d'une macromolécule déterminée (Jean-Marc et all. 2002).

- **Rejet du sel de sulfate de magnésium SR($MgSO_4$)**

Le rejet du sel de sulfate de magnésium incorpore les encombrements stériques et les effets de répulsion électrostatique qui sont liés à la structure de la membrane et les conditions d'exploitation. L'utilisation du seuil de coupure était acceptable à des fins de modélisation, toute fois les membranes d'OI/NF avec une large gamme de MWCO est difficile à estimer la rétention de contaminants. Le rejet du sel de sulfate de magnésium peut être plus approprié (YangaliQuintanilla et al. 2009b et 2010b).

Kosutic et al. 2000 ont étudié la porosité des membranes commerciales en polyamides et ont rapporté que les différences de performances sont principalement dues par des différences de porosité. Par ailleurs, ils ont indiqué que la variation de la pression du travail a une influence sur cette porosité et donc sur les performances des membranes. Pour certaines membranes, l'augmentation de la pression conduit à un effet de compaction qui fait diminuer la densité de flux du perméat et la taille des pores tout en augmentant leur nombre.

Le pH peut avoir une influence sur la porosité : à pH élevé, la répulsion entre les groupements acides au sein de la membrane peut entrainer une augmentation de la taille des pores (Berg et al. 1997).

I.3.1.2. Taille moléculaire

La taille moléculaire est un paramètre principal qui détermine la rétention de molécules organiques avec les membranes d'OI/NF. Plusieurs paramètres peuvent être utilisés pour représenter la taille moléculaire : le poids, la longueur, la largeur, la profondeur et la largeur

équivalente qui sont calculés à partir de la structure moléculaire. Van der Bruggen et al. 1998 ont montré qu'il existe une bonne corrélation entre la rétention et les différents paramètres de taille (Figure I.9).

A. Poids moléculaire "MW"

Le poids moléculaire est un paramètre facilement disponible, il fournit seulement une estimation approximative de la taille moléculaire. Il est utilisé pour avoir une première idée sur la taille relative des molécules étudiées. Il ne peut pas être utilisé pour la modélisation, car il ne représente pas la géométrie de la molécule. Pour les molécules non chargées, peu polaires, la masse molaire peut fournir une première information sur la rétention avec de bonnes corrélations (Sadmani 2014b). Cependant, il ne renseigne pas sur l'encombrement stérique de la molécule. L'évaluation de la taille et de la géométrie de la molécule couplée à la taille de pores peut être un meilleur descripteur que la masse molaire. En effet, des pesticides ayant une section et non une masse molaire plus faible sont moins bien retenus. Le poids moléculaire est considéré comme une variable de descripteur moins significatif pour la modélisation de la rétention (Yangali-Quintanilla et al. 2008).

B. Longueur moléculaire "Length"

La longueur moléculaire (Length) a été définie comme la distance entre les deux atomes les plus éloignés de la molécule (en tenant compte des forces des Van der Waals). La ligne droite reliant les deux atomes les plus éloignés a été définie comme la L-axe(Santos et al. 2006).

C. Largueur moléculaire "Width"

La molécule est projetée sur le plan perpendiculaire à l'axe L. Le rectangle de surface minimale entourant la projection de la molécule est sélectionné. L'aire du rectangle (S) est calculée et la largeur moléculaire (Width) est définie comme étant la moitié de la racine carrée de S. La racine carrée de S correspond à la longueur du côté du carré dont la superficie est égale à la surface du rectangle (Kiso et al. 2010).

$$Width = \frac{1}{2}\sqrt{S_{min}} \qquad (I.4)$$

D. Profondeur moléculaire "Depth"

La largeur moléculaire et la profondeur moléculaire (la largeur supérieure à la profondeur) sont mesurées par la projection de la molécule sur le plan perpendiculaire à l'axe de la longueur (Kiso et al. 1992 ; Santos et al. 2006).

$$Depdth = \frac{1}{2}\sqrt{S_{max}} \qquad (I.5)$$

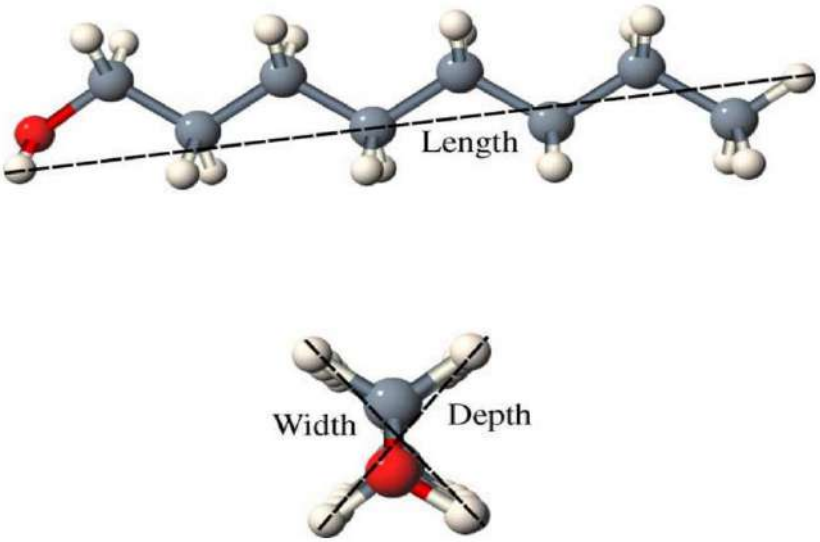


Figure I.12. Calcul des dimensions de l'octanol (Length, Width et Depth) de la structure 3D (Santos et al. 2006).

E. Largeur équivalente moléculaire "*EqWidth*"

La largeur moléculaire équivalente est définie comme étant la moyenne géométrique de la largeur et de la profondeur (Santos et al.2006 ; Linares et al. 2011).

On peut calculer la largeur équivalente par la formule suivante :

$$EqWidth = \sqrt{Width * Depth} \qquad (I.6)$$

I.3.2. Répulsions électrostatiques (effet de charge)

Les interactions électrostatiques ont souvent été rapportées comme participantes de manière importante au mécanisme de rétention. Elles s'établissent entre deux entités chargées (molécules organiques et membranes) (Van der Bruggen et al. 1998).

I.3.2.1. Charge de la surface

En général les membranes d'OI/NF thin-film composite (TFC) portent une charge négative (Norberg et al. 2007) qui permet de minimiser l'adsorption des composés colmatant chargés négativement et d'augmenter la rétention des sels ionisés dans l'eau. La charge de la membrane est généralement quantifiée par mesure du potentiel zêta (ζ) (Bellona et Drewes 2005).

Le pH de l'alimentation a une influence sur la charge de la membrane, il joue sur la protonation des groupes fonctionnels. A pH élevé, la membrane est chargée négativement du fait de la dissociation du groupement carboxylique et à pH faible elle est chargée positivement du fait de l'ionisation du groupement amines. Le pH pour lequel le potentiel zêta est nul est appelé point isoélectrique ou pH d'ionisation de la membrane (Desmukh et Childress 2001).

I.3.2.2. Charge du soluté

Comme nous venons de voir, les membranes d'osmose inverse et de nanofiltration présentent une charge généralement négative. Il est donc clair qu'un soluté chargé négativement sera mieux retenu, par répulsion électrostatique, qu'un soluté neutre. Le respect de l'électro neutralité implique que les composés chargés positivement le seront également. Les molécules ionisables sont donc en général mieux retenues sous leur forme ionisée. Cet effet est d'autant plus marqué que la molécule est petite, car sa densité de charge est alors plus élevée. Pour des molécules de masses molaires proches, celles de même signe que la membrane sont mieux retenues que les neutres, elles-mêmes mieux retenues que celles de charges opposées à celles de la membrane. Pour ces dernières, l'attraction électrostatique favorise leur passage. Pour les plus grosses molécules les effets stériques restent prédominants (Ammi 2016).

I.3.2.3. Influence du pH

Comme nous l'avons vu précédemment, la charge des membranes dépend du pH de la solution avec laquelle elles sont en contact. Suite à cela, il peut se créer des interactions électrostatiques entre la membrane et les solutés chargés, ionisés ou non, présents en solution. Le positionnement de ceux-ci près de la surface modifie alors la charge apparente de la membrane. Par ailleurs, si le pH n'a pas d'influence sur le taux de rétention des molécules non chargées, alors il joue un grand rôle sur celui des molécules ionisables puisqu'il gouverne leur ionisation. En effet, la charge des espèces acido-basiques faibles varie avec le pH (Ammi 2016).

I.3.3. Interaction hydrophobes/adsorption

L'adsorption est fortement liée à l'hydrophobicité des molécules, à leur capacité à former des liaisons hydrogènes et à leur polarité.

I.3.3.1. Hydrophobicité de la membrane

La mesure de l'angle de contact α que forme l'eau sur une membrane, permet d'avoir une idée de l'hydrophobicité de cette membrane. Cette dernière est supposée gouverner la mouillabilité de la membrane en particulier par l'eau et donc avoir une influence sur ses performances. Il existe trois méthodes de mesure de l'angle de contact : la bulle captive, la technique de la lame de Wilhelmy et la goutte déposée. Par l'utilisation de la méthode de la goutte déposée sur une surface solides (Figure I.13), deux cas peuvent être présentés (Sagne 2008) :

- α est supérieur à 90°, la membrane est dite hydrophobe
- α est inférieur à 90°, la membrane est dite hydrophile.

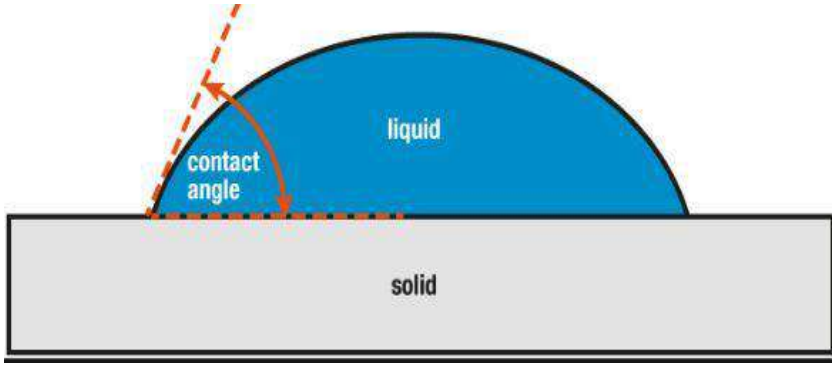


Figure I.13. Mesure de l'angle de contact par la méthode de la goutte déposée (https://www.linseis.com/fr/grandeurs-mesurees/angle-de-contact/).

La plupart des membranes d'OI/NF sont hydrophiles possédant des valeurs de l'angle de contact généralement inférieur à 90°.

I.3.3.2. Hydrophobicité des solutés

Le caractère hydrophobe ou hydrophile des solutés va donc jouer un rôle sur leur rétention. Il est évalué par le logarithme du coefficient de partage octanol-eau ($logk_{ow}$), plus souvent exprimé sous forme logarithmique ($logk_{ow}$, logP). S'il est supérieur à 2, la molécule est hydrophobe (HB) ; s'il est inférieur, la molécule est hydrophile (HL).

Le logarithme du coefficient de partage n-octanol/eau, LogP, a été utilisé comme un indice hydrophobe, et est défini comme suit (Équation I.7)

$$Logk_{ow} = \log\left(\frac{C_0}{C_W}\right) \quad (I.7)$$

Où C_0et $C_W$sont les concentrations des molécules dans les couches n-octanol et l'eau (Taheran et al. 2016).

Pour les solutés ionisables, le caractère hydrophobe est fonction du pourcentage des chargé contre les espèces non chargées, il dépend donc du pH. Par conséquent, un coefficient de partage apparent à un pH donné est calculé pour les solutés ionisables. Le coefficient de distribution octanol-eau dépend du pH, D, est le rapport entre les concentrations d'équilibres de toutes les espèces (ionisé et non) d'une molécule dans l'octanol à la même espèce dans la phase aqueuse. Le logarithme du coefficient de distribution peut être calculé à partir du pKa et $logk_{ow}$ du soluté (Équations I.8 et I.9) (Verliefde et al. 2008).

Pour les acides :

$$\mathrm{LogD_{(pH)}} = \mathrm{logk_{ow}} - \log\left(1 + 10^{(\mathrm{pH-pKa})}\right) \quad (I.8)$$

Pour les basses :

$$\mathrm{LogD_{(pH)}} = \mathrm{logk_{ow}} - \log(1 + 10^{(\mathrm{pKa-pH})}) \quad (I.9)$$

On parle également des molécules polaires/apolaires. Les molécules hydrophobes sont en général apolaires. Le dipôle est dirigé vers la membrane de telle sorte que le côté du dipôle de charge opposée est plus proche de la membrane. Cette orientation n'est pas statique, afin elle doit être considérée comme une tendance statistique des molécules en mouvement rapide d'avoir cette orientation préférentielle. Le dipôle dirigé est donc attiré dans les pores et entre ainsi plus facilement dans la structure de la membrane. Les molécules ayant un moment dipolaire élevé ont des rétentions inférieures par rapport à des molécules non polaires. Cet effet, est indépendant du signe de la charge de la membrane (positif ou négatif) (Sagne 2008).

I.4. Applications des procédés d'OI et NF

Les différentes applications envisagées pour les procédés OI et NF sont résumées comme suites (Ammi 2016) :

- **Industries diverses :**
  - Récupération des métaux ;
  - Production de l'eau ;
  - Traitement de surface ;
  - Effluents industriels.
- **Industrie agro-alimentaire :**
  - Filière œnologique ;
  - Lactosérum ;
  - Recyclage des protéines ;
  - Effluents agro-alimentaires ;
  - Industrie pharmaceutique et biotechnologique: Sérums, principes actifs, antibiotique.
- **Eau potable, dessalement, saumure :**
  - Élimination nano-polluants organiques ;
  - Élimination des particules ;
  - Adoucissement ;
  - Déminéralisation.

I.5. Avantages et inconvénients des procédés d'OI et NF

Les avantages et les inconvénients des procédés d'OI et NF sont regroupés dans le tableau I.3.

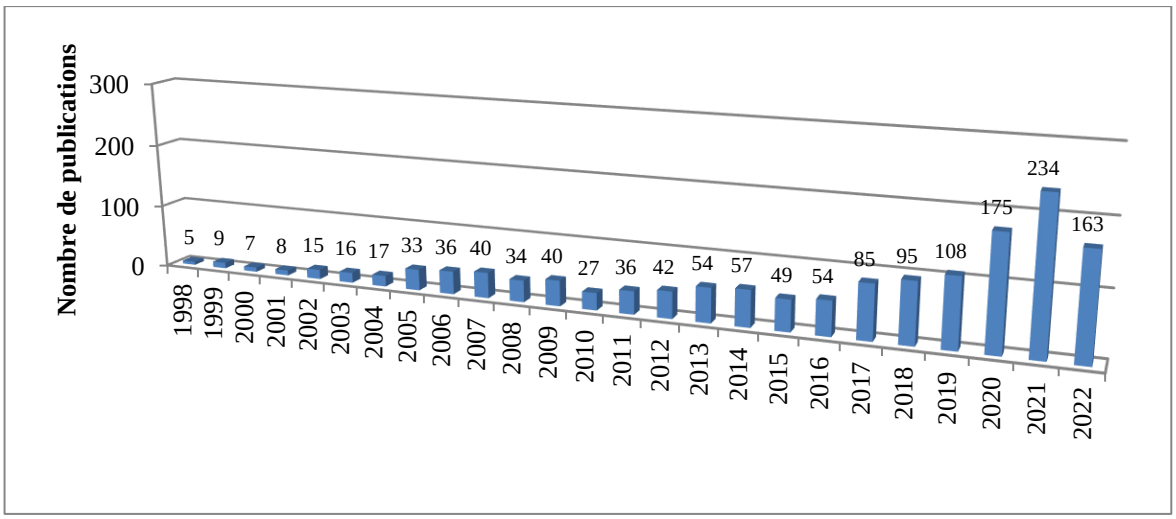
Tableau I.3. Avantages et inconvénients de procédés d'OI et NF (Ammi 2016).

Avantages	Inconvénients
♣ Un taux de conversion élevé ♣ (supérieur à 55%). Une consommation spécifique ♣d'énergie basse (3 à 5 Kwh/m^3). Un coût d'investissement ♣relativement bas. Une durée de mise en œuvre réduite. ♣ Opération à température ambiante ce ♣qui évite la dégradation de molécules fragiles dans le domaine agroalimentaire. Pas	♣ Sensibilité à la variation de la qualité ♣d'eau brute et surtout aux pollutions. Nécessité d'un personnel qualifié. ♣ Exploitation relativement difficile ♣surtout au niveau du prétraitement. Un coût de maintenance élevé dû aux ♣frais de renouvellement des membranes d'une durée de vie ne dépassant pas 7 ans à présent. Baisse de la perméabilité et ♣modification de la sélectivité en cas de

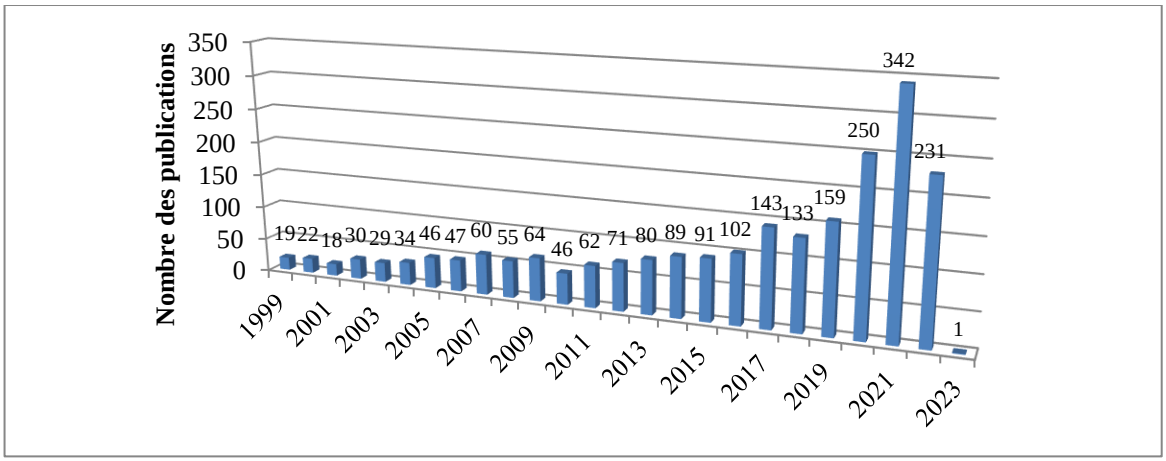
d'intervention de réactifs ♣chimiques comme des agents d'extraction qui sont des sources de pollution.	colmatage des membranes.

I.6. Revue des travaux de la rétention des pesticides lors d'OI/NF

De nombreux travaux réalisés au niveau international mettent en évidence la rétention des pesticides par d'OI/NF. La base de données consultée est Science Direct. La figure I.13, reporte le nombre de publications de 1998 à 2022 pour les membranes de la nanofiltration et d'osmose inverse.



(b)



(b)

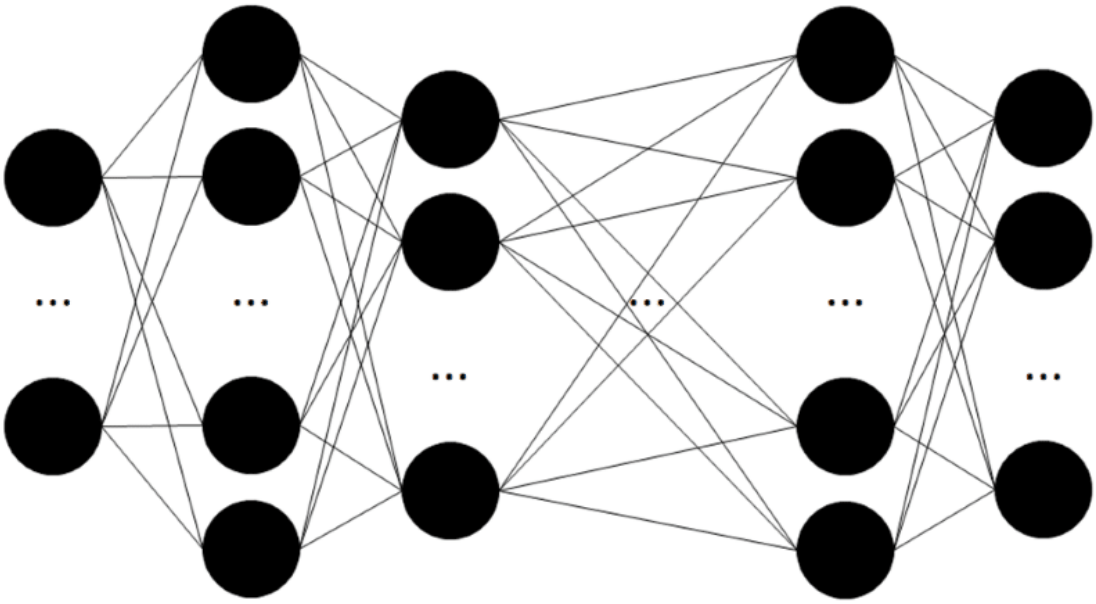
Figure I.13. Nombre de publications traitant la rétention des molécules organiques par (a) : nanofiltration (b) osmose inverse, (Science Direct le 22/05/2022).

I.7. Conclusion

La rétention des pesticides par les membranes de la nanofiltration et d'osmose inverse dépend d'un grand nombre de facteurs liés à la fois à la membrane (structure physique, charge, hydrophobicité), à la molécule (taille, structure, charge, polarité ou hydrophobicité) et aux conditions expérimentales (pH, pression, taux de récupération et température). Les interactions mises en place sont à la fois d'ordre stériques, électrostatiques et dues au caractère hydrophobe.

Chapitre II

OSPR-Réseau de Neurones Artificiels



II.1. Introduction

La Chemoinformatique est un nouveau domaine apparu il y a quelques années. Il concerne le développement, la création, l'organisation, le stockage, la diffusion, l'analyse, la visualisation et l'utilisation de l'information chimique, elle représente un domaine en évolution rapide. Actuellement, les méthodes spécifiques de la chemoinformatique sont devenues indispensables pour le développement de nouveaux composés, matériaux et processus.

La chemoinformatique est une discipline scientifique qui a beaucoup évoluée durant ces 40 dernières années. Cette science est à l'interface entre la chimie et l'informatique. Il a été rendu compte que dans beaucoup de domaines de la chimie moderne, le nombre important de données et d'informations fourni par la recherche en chimie peut seulement être analysé par des méthodes informatiques. De plus, les nouveaux problèmes rencontrés en chimie sont si complexes que l'utilisation de l'informatique a été rendue indispensable. C'est pourquoi des méthodes ont été développées pour prédire les propriétés chimiques, physiques et biologiques des matériaux et des composés chimiques. Ces méthodes ont aussi été développées pour la représentation des médicaments, l'élucidation des structures cristallines, dans la prédiction des réactions chimiques et dans la représentation des molécules organiques. La recherche et le développement en chemoinformatique sont essentiels pour :

- L'amélioration de la compréhension des phénomènes chimiques ;
- Permettre à l'industrie chimique de rester compétitive dans le marché mondial (http://www.chemspider.com/Chemical-Structure.6886.html?rid=0ae6e03a-77d9-40b1a7cd-44816ea08f94&page_num=0).

Ces dernières années, il y a eu plusieurs tentatives pour faire progresser l'utilisation des relations quantitatives structure-propriétés (QSPR) comme une approche viable pour développer des modèles pilotés par les données pour décrire la performance des procédés membranaires (Dornier et al. 1995 ; Sahoo et al. 2006 ; Libotean et al. 2009). Peu d'étude sont également appliqué les modèles QSPR à la modélisation de la rétention des composés organiques par les membranes d'osmose inverse (OI) et de nanofiltration (NF) (Libotean et al. 2008, Yangali-Quintanilla et al. 2008, 2009b et 2010b, Shahmansouri et al. 2013, Sadmani et al. 2014a, Sadmani et al. 2014b, Ammi et al. 2015, 2017, 2018, 2020, 2021, 2021).

Ce chapitre a été consacré à des modèles de réseaux de neurones artificiels basés sur la relation quantitative structure-activité "QSAR-RN". On parlera sur la méthode d'apprentissage statistique (relation quantitative structure-activité "QSAR" et les réseaux de neurones artificiels "RNA"). Une

conclusion clôturera ce chapitre et sera une synthèse par laquelle on justifiera le choix du thème de ce mémoire.

II.2. Méthodes d'apprentissages statistiques

La théorie de l'apprentissage statistique est un système d'apprentissage automatique à partir des domaines de la statistique et de l'analyse fonctionnelle. La théorie de l'apprentissage statistique traite du problème de la recherche d'une fonction prédictive basée sur des données. La théorie de l'apprentissage statistique a conduit à des applications dans des domaines tels que la vision par ordinateur, la reconnaissance de la parole, la bioinformatique (https://fr.wikipedia.org/wiki/Bio-informatique).

II.3. Relations quantitatives structure-Activités (QSAR)

Initiées sur la base de relations empiriques durant la seconde partie du 19ème siècle, les QSAR 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 suites aux travaux de Hansch et Fujita (Devillers, 2013) 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 activités d'intérêt biologique.

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

- Prédire l'activité biologique et les propriétés physico-chimiques.
- Comprendre et rationaliser les mécanismes d'action dans une série de produits chimiques.
- Économiser 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.

La capacité à prédire une activité biologique est utile dans un certain nombre de situations. Il existe un grand nombre d'applications de ces modèles QSAR au sein de l'industrie, du milieu universitaire et des organismes gouvernementaux chargés de la réglementation. Nous résumons ci-après quelques cas d'utilisations de ces modèles :

- l'optimisation de l'activité pharmacologique, biocide ou pesticide ;

- la conception rationnelle de nombreux autres produits tels que des agents tensio-actifs, des parfums, des colorants ou des produits chimiques ;
- l'identification des composés dangereux à des stades précoces de développement ;
- la conception de la toxicité et des effets secondaires pour les nouveaux composés ;
- la prédiction de la toxicité pour les humains pour les multiples expositions (délibérée, occasionnelle ou professionnelle) et de l'écotoxicité ;
- la prédiction d'une variété de propriétés physico- chimiques des molécules (qu'il s'agisse de produits pharmaceutiques, de pesticides, de produits d'hygiène personnelle, la chimie fine, etc.) (Hamadache 2014).

II.3.1. Outils de QSAR

Les modèles QSAR permettent de quantifier la relation entre la structure et l'activité biologique et d'apprécier l'impact de diverses propriétés physico-chimiques sur cette activité. L'élaboration de tels modèles QSAR repose sur l'utilisation de trois outils : une ou plusieurs données biologiques 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 (Hamadache 2014).

II.3.2. Données biologiques

Les données biologiques sont habituellement exprimées sur une échelle logarithmique en raison de la relation linéaire entre la réponse et le logarithme de dose dans la région centrale de la courbe de log dose-réponse. Les logarithmes inverses de l'activité (log 1/C) sont également utilisés pour obtenir des valeurs mathématiques plus élevées lorsque les structures sont biologiquement très efficaces. Parmi les données biologiques utilisées dans l'analyse de QSAR, on trouve la constante de vitesse, la constante d'inhibition, … etc (Hamadache 2014).

II.3.3. 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 (Poezevara, 2011) : les descripteurs expérimentaux (par exemple : la lipophilie [logP], la polarisabilité, la densité électronique, le potentiel d'ionisation) et les descripteurs moléculaires théoriques. Ces derniers sont à nouveaux classés en trois catégories en fonctionde leur dimension (Hong 2008 et Hamadache 2014).

A. **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.

B. **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.

C. Descripteurs 3D

Ces descripteurs 3D reflètent les propriétés physiques dans la plupart des cas, mais sont insuffisants pour expliquer de façon satisfaisante certaines propriétés ou activités, telles les activités biologiques. 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 l'activité biologique de molécules. 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. Par exemple, le programme appelé DRAGON permet de calculer différents groupes de descripteurs : auto correlations 2D, descripteurs géométriques, descripteurs RDF, descripteurs 3D-MoRSE, descripteurs WHIM, descripteurs GETAWAY, …etc (Hamadache 2014).

II.3.4. Formulation mathématique des QSAR

Les équations QSAR prennent la forme générale suivante :

$$QSAR \begin{cases} Activité\ biologique\ =\ fonction\ (Descripteurs) \\ Activité\ thérapeutique\ =\ fonction\ (Descripteurs) \end{cases}$$

$$QSPR : Propriétés = f(Descripteurs)$$

$$QSTR : Toxicités = f(Descripteurs)$$

Dans l'équation, l'activité biologique est normalement exprimée comme log [1/ (symbole de concentration)], où C'est généralement la concentration minimum requise pour causer une réponse biologique définie. La relation mathématique entre un paramètre ou plusieurs paramètres physico-chimiques et l'activité biologique d'un composé peut être exprimée à travers l'équation de Hansch :

$$\log\left(\frac{1}{c}\right) = b_0 + \sum_{i=1}^{j}(b_i D_i) \qquad \text{(II. 1)}$$

Dans cette équation, C'est la concentration de substance nécessaire pour obtenir une réponse biologique, $b_0$est une constante, b_i est un coefficient, D_I est un descripteur moléculaire et j est le nombre de descripteurs utilisés.

Les relations QSARs les plus connues ont été introduites par Hansh et Fujita (Poezevara, 2011). Une des équations les plus utilisées présume que les propriétés physiques responsable de l'activité biologique peuvent être séparées en trois catégories : (i) une composante hydrophobique (π), (ii) une composante électronique (σ) et (iii) une composante stérique (ES, RM) :

$$log\left(\frac{1}{c}\right) = a + b\pi + c\,\pi^2 + dE_s + \ eRM + f\sigma \qquad \text{(II. 2)}$$

a, b, c, d, e et f sont des coefficients déterminés durant l'étude QSAR, tandis que π, ES, RM et σ sont des paramètres fonctions des substituants. L'équation de Hansch est utilisée pour prévoir l'activité biologique de composés dont la structure est semblable à ceux utilisés pour former cette équation. Cependant, dans le cas où la valeur prévue est différente de celle déterminée par l'expérimentation, cela peut signifier que l'activité biologique réelle des composés n'est pas prise en compte par les paramètres mis en jeu dans le processus de construction de l'équation (Hamadache 2014).

II.4. Réseaux de neurones artificiels (RNA)

II.4.1. Histoire des réseaux de neurones artificiels

Un réseau de neurones artificiels est un système informatique matériel et / ou logiciel dont le fonctionnement est calqué sur celui des neurones du cerveau humain. Il s'agit là d'une variété de technologie Deep Learning (apprentissage profond), qui fait elle-même partie de la sous-catégorie d'intelligence artificielle du Machine Learning (apprentissage automatique).

Le concept des réseaux de neurones artificiels futinventé en 1943 par deux chercheurs de l'Université de Chicago : le neurophysicien Warren McCullough, et le mathématicien Walter Pitts. Dans un article publié dans le journal Brain Theory, les deux chercheurs présentent leur théorie selon laquelle l'activation de neurones est l'unité de base de l'activité cérébrale.

En 1957, le Perceptron fut inventé. Il s'agit du plus ancien algorithme de Machine Learning, conçu pour effectuer des tâches de reconnaissance de patterns complexes. C'est cet algorithme qui permettra plus tard aux machines d'apprendre à reconnaître des objets sur des images.

Malheureusement, à l'époque, les réseaux de neurones étaient limités par les ressources techniques. Par exemple, les ordinateurs n'étaient pas assez puissants pour traiter les données nécessaires au

fonctionnement des réseaux de neurones. C'est la raison pour laquelle la recherche dans le domaine des Neural Networks est restée en sommeil durant de longues années.

Il aura fallu attendre le début des années 2010, avec l'essor du Big Data et du traitement massivement parallèle, pour que les Data Scientists disposent des données et de la puissance de calcul nécessaires pour exécuter des réseaux de neurones complexes. En 2012, lors d'une compétition organisée par Image Net, un Neural Network est parvenu pour la première fois à surpasser un humain dans la reconnaissance d'image.

C'est la raison pour laquell**e** cette technologie est de nouveau au cœur des préoccupations des scientifiques. A présent, les réseaux de neurones artificiels ne cessent de s'améliorer et d'évoluer de jour en jour (https://www.lebigdata.fr/reseau-de-neurones-artificiels definition#:~:text=(apprentissage%20automatique).,L'histoire%20des%20r%C3%A9seaux%20de%20neurones%20artificiels,et%20le%20math%C3%A9maticien%20Walter%20Pitts).

II.4.2. Neurone biologique

L'idée de réseaux de neurones artificiels a été inspirée de la manière par laquelle les neurones biologiques traitent les informations. Le cerveau humain est composé de cellules distinctes appelées neurones formant une ensemble dense d'environ 10 à 100 milliards d'unités interconnectées. Tel qu'il est illustré sur la figure II.1, le neurone est une cellule composée d'un corps cellulaire et d'un noyau. Le corps cellulaire sera méfié pour former les dendrites. C'est par les dendrites que l'information est acheminée de l'extérieur vers le soma (corps du neurone) (Ammi 2016).

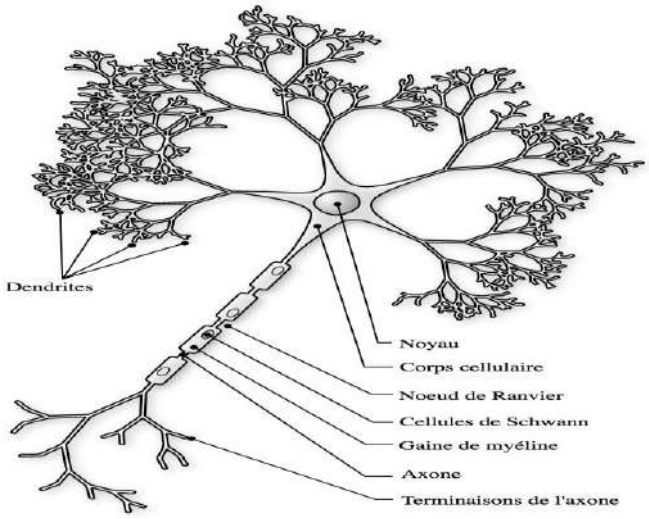


Figure II.1. Représentation d'un neurone biologique par Nicolas Rougier(https://scilogs.fr/intelligence-mecanique/reprenons-bases-neurone-artificiel-neurone-biologique/).

Un neurone biologique reçoit des entrées ou signaux transmis par d'autres neurones (interaction dendrites- synapse). Au niveau du corps (soma), le neurone analyse et traite ces signaux en les sommant. Si le résultat obtenu est supérieur au seuil d'activation (ou d'excitabilité), il envoie une décharge alors nommé potentiel d'action le long de son axone vers d'autres neurones biologiques.

II.4.3. Neurone formel

Le neurone formel est l'élément essentiel d'un réseau de neurones. C'est un opérateur mathématique très simple, dont on peut facilement calculer la valeur numérique (Hamadache 2014). Son fonctionnement est schématisé sur la figure II.2.

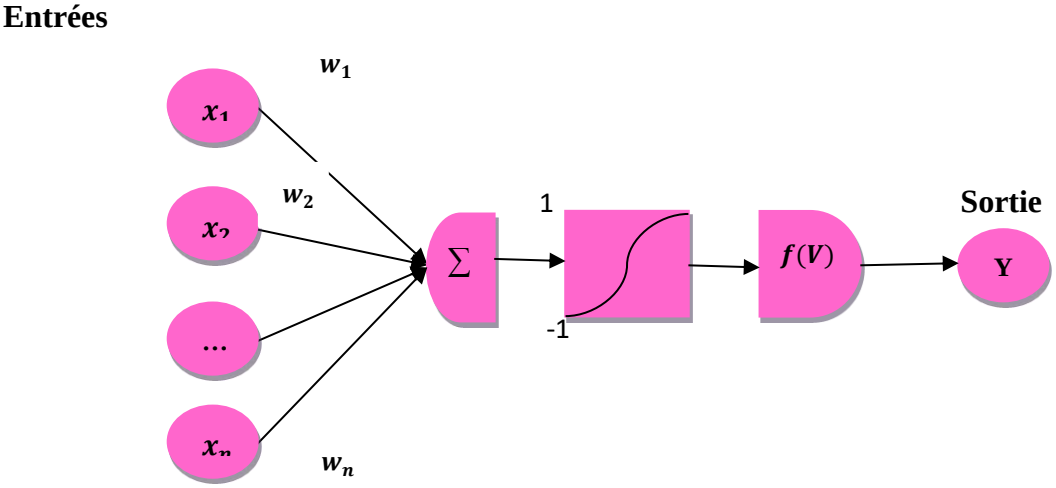


Figure II.2. Neurone formel (Ammi 2016).

Un neurone formel est une fonction algébrique paramétrée, non linéaire en ses paramètres. Ses entrées peuvent être les sorties d'autres neurones ou des entrées de signaux extérieurs. Sa sortie est une fonction non linéaire f d'une combinaison linéaire v des entrées (x_i). Le potentiel v le plus fréquemment utilisé est la somme pondérée des entréesx_i pondérées par les coefficients (w_i) également appelés poids de connexions :

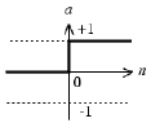
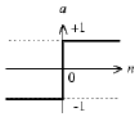
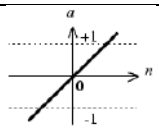
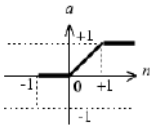
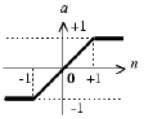
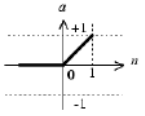
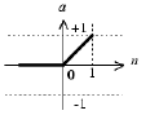
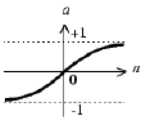
$$v = w_0 + \sum_{i=1}^{n} (w_i x_i) \qquad \text{(II. 3)}$$

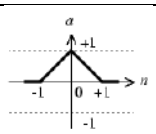
Précisons que le poids $w_0$est affecté à une entrée constante appelée "biais". Une fonction f, appelée fonction d'activation, est appliquée à ce potentiel v. Le choix d'une fonction d'activation se révèle dans certains cas être un élément constitutif important des réseaux de neurones. Les fonctions mathématiques les plus souvent utilisées dans notre travail sont regroupées dans le tableau II.1. À la

différence des neurones biologiques dont l'état est binaire, la plupart des fonctions de transfert sont continues, offrant une infinité de valeurs possibles comprises dans l'intervalle [0, +1] ou [-1, +1]. Un neurone formel, réalise donc l'opération suivante (Ammi2016) :

$$y = f(v) = f\left(w_0 + \sum_{i=1}^{n}(w_i x_i)\right) \qquad \text{(II. 4)}$$

Tableau II.1. Fonctions de transfert usuelles, a=f(n) (Khaouane 2013).

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 ≤ n ≤1 a=1 si n > 1		Satlin
Linéaire saturée Symétrique	a= −1 si n < −1 a= n si −1≤ n ≤1 a=1 si n >1		Satlins
Linéaire positive	a= n si $n \geq 0$ a= 0 si n < 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	$a = 1 - \|n\|, -1 \leq n \leq 1$ a *=0 autrement*		Tribas

Un neurone formel est une représentation artificielle et schématique d'un neurone biologique (Figure II.3) :

- Les synapses sont modélisées par des poids,
- Le soma ou corps cellulaire est modélisé par la fonction de transfert, appelé aussi fonction d'activation
- L'axone par l'élément de sortie (https://scilogs.fr/intelligence-mecanique/reprenons-bases-neurone-artificiel-neurone-biologique/).

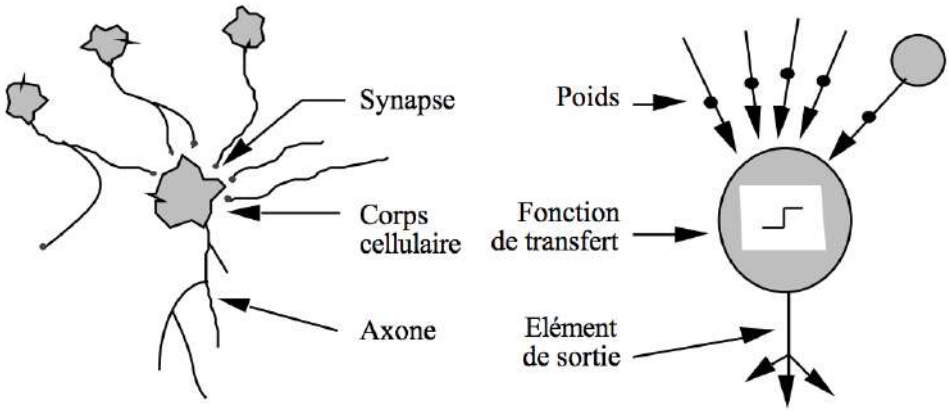


Figure II.3. Mise en correspondance neurone biologique / neurone artificiel par Claude Touzet(https://scilogs.fr/intelligence-mecanique/reprenons-bases-neurone-artificiel-neurone-biologique/).

II.4.4. Architecture d'un réseau de neurones

Un RNA est formé d'un ensemble de neurones formels interconnectés. Deux grands types d'architectures peuvent être distingués (Figure II.4) :

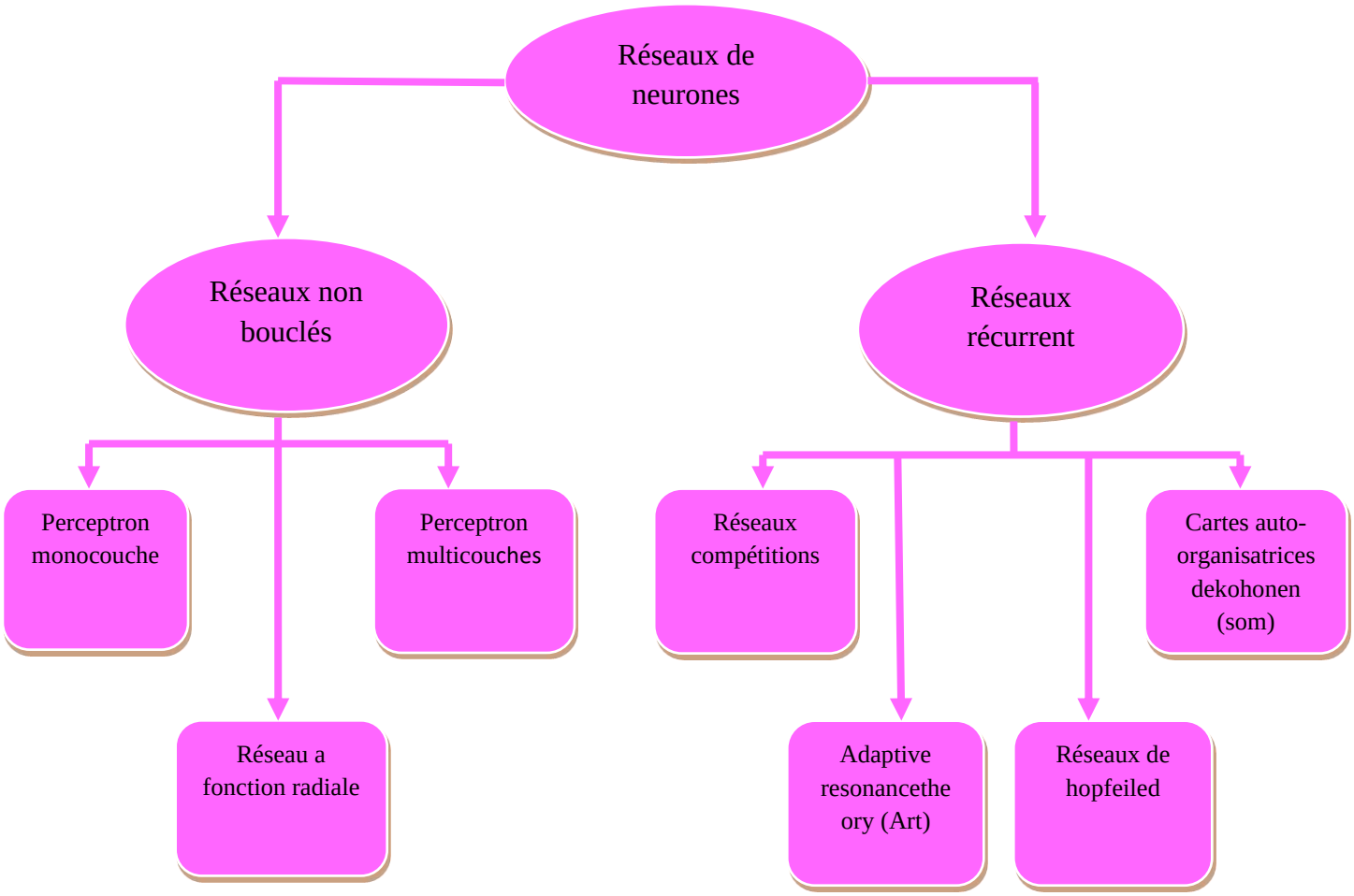


Figure II.4. Réseaux de neurones les plus utilisés.

- **Les réseaux de neurones non bouclés :** Ils réalisent une (ou plusieurs) fonction(s) algébrique(s) des entrées, par composition des fonctions réalisées par chaque neurone. Les réseaux de neurones non bouclés sont des outils statiques utilisés principalement pour effectuer des tâches d'approximation de fonctions non linéaires, de modélisation de processus statiques non linéaires (Figure II.5).

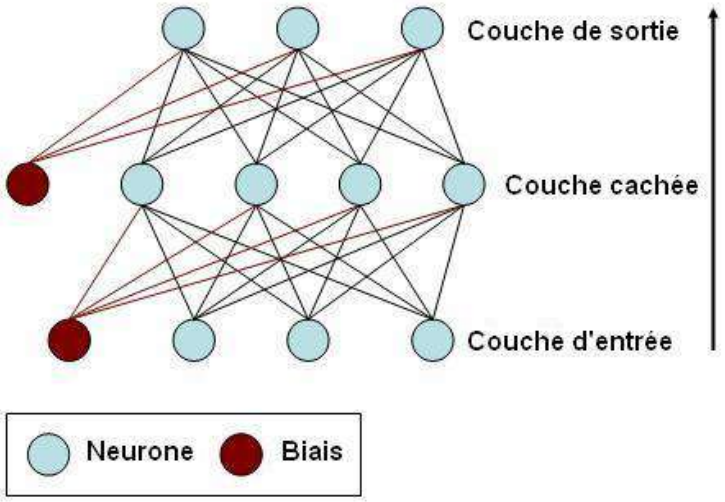


Figure II.5. Structure d'un Réseaux de neurones non bouclés (http://www.statsoft.fr/concepts-statistiques/reseaux-de-neurones-automatises/reseaux-de-neurones-automatises.htm#.YjJNDtXMLIU).

- **Les réseaux de neurones bouclés :** Ils constituent quant à eux un système dynamique "à temps discret", régi par une (ou plusieurs) équation(s) aux différences non linéaires, résultant de la composition des fonctions réalisées par chacun des neurones et des retards associés à chacune des connexions. Ils sont utilisés pour effectuer des tâches de modélisation et d'adaptation de systèmes dynamiques, de commande de processus ou de filtrage. Dans notre étude, nous avons adopté le modèle le plus couramment utilisé, le Perceptron Multicouches (PMC), illustré par la figure II.6 et dont les principes sont exposés dans ce qui suit (Ammi 2016).

II.4.5. Perceptrons multicouches

Le Perceptron multicouches "PMC" est un réseau de neurones non bouclé caractérisé par une structure bien particulière. La figure II.5 illustre la structure de ce type de réseau. Ces neurones sont organisés en couches successives où les informations circulent dans un seul sens, de la couche d'entrée vers la couche de sortie. Les neurones d'une même couche ne sont pas interconnectés. Un neurone ne peut envoyer son résultat qu'à un neurone situé dans une couche postérieure à la sienne. Le réseau est formé de plusieurs couches, qui s'appellent, par convention :

- couche d'entrée : cette couche représente toujours une couche associée aux entrées du système, composée des cellules d'entrée qui correspondent aux k variables d'entrée.

- couche(s) cachée(s) : chaque couche est composée de p neurones. Ces neurones n'ont aucun lien avec l'extérieur et sont appelés neurones cachés.
- couche de sortie : elle est constituée par la valeur de sortie de la combinaison linéaire des fonctions de transferts de la couche cachée. Les neurones sont reliés entre eux par des connexions pondérées. Ce sont les poids de ces connexions qui gouvernent le fonctionnement du réseau et programment une application de l'espace des entrées vers l'espace des sorties, à l'aide de la transformation non linéaire (Hamadache 2014).

II.4.6. Types d'apprentissage des réseaux de neurones

L'apprentissage est vraisemblablement la propriété la plus intéressante des réseaux de neurones. L'apprentissage est une phase du développement du réseau de neurones durant laquelle on calcule les poids des neurones de telle manière que les sorties du réseau soient aussi proches que possible des sorties désirées. Les procédures d'apprentissage sont divisées, elles aussi, en deux grandes catégories : apprentissage supervisé et apprentissage non supervisé (Khaouane 2013).

A. Apprentissage supervisé

L'apprentissage supervisé implique l'existence d'un professeur qui a pour rôle d'évaluer le succès (ou l'échec) du réseau quand il lui est présenté un stimulus connu (on dit que ce stimulus est un exemple appartenant à la base d'apprentissage). Cette supervision consiste à renvoyer au réseau une information lui permettant de faire évoluer ses connexions (parfois aussi sa propre architecture) afin de faire diminuer son taux d'échec. L'information peut être explicite, sous la forme d'une mesure de l'erreur commise, par exemple, ou globalement sur l'ensemble des exemples de la base. C'est ce qui se passe dans le cas du perceptron. Mais elle peut être plus implicite (apprentissage par renforcement), sous forme d'une simple appréciation (bon ou mauvais, punition ou récompense), sans mesure d'erreur et même être globale, sur l'ensemble des tâches que le réseau doit exécuter. Dans ce dernier cas, il est facile de comprendre que l'apprentissage est le plus difficile, la difficulté majeure consistant pour le réseau à identifier les étapes du processus qui sont responsables de l'échec ou du succès (Khaouane 2013).

B. Apprentissage non supervisé

L'apprentissage non supervisé implique la fourniture à un réseau autonome d'une quantité suffisante d'exemples contenant des répétitions (autrement dit de la redondance), telles que celui-ci en dégage les régularités automatiquement. Ces réseaux sont souvent appelés auto-organisateurs, ou encore à apprentissage compétitif. Dans l'apprentissage non supervisé, les données ne contiennent

pas d'informations sur une sortie désirée, il n'y a pas de superviseur ou expert humain. Il s'agit de déterminer les paramètres du réseau de neurones suivant un critère à définir. Dans ce cas, les exemples présentés à l'entrée provoquent une auto adaptation du réseau afin de produire des valeurs de sortie qui soient proches en réponse pour des valeurs d'entrées similaires. Ce type d'apprentissage possède souvent moins de complexité dans le calcul par rapport à l'apprentissage supervisé. Bien entendu, l'architecture du réseau, préalablement définie par son utilisateur, est une forme de supervision (Khaouane 2013).

II.4.7. Réseaux de neurones les plus utilisés

A. Perceptron monocouche (perceptron simple)

Le perceptron est le premier modèle des réseaux de neurones inventé en 1957 par Frank Rosenblatt (Rosenblatt 1958). C'est un réseau simple, puisqu'il ne se compose que d'une couche d'entrée et d'une couche de sortie. Il peut aussi être utilisé pour faire de la classification et pour résoudre des opérations logiques simples, il suit généralement un apprentissage supervisé (Talbi 2011).

B. Le perceptron multicouche (en anglais Multi Layer Perceptron)

C'est une extension du précédent, avec une ou plusieurs couches cachées entre l'entrée et la sortie (Keedwell et Narayanan 2005). Chaque neurone dans une couche est connecté à tous les neurones de la couche précédente et de la couche suivante. Il peut résoudre des problèmes non linéairement séparables et des problèmes logiques plus compliqués (Talbi 2011). Il suit un apprentissage supervisé, ce réseau a été choisi dans la suite de nos travaux.

C. Réseaux à fonction radiale

Les réseaux à fonction de bases radiales (RFR) ou plus simplement réseaux à bases radiales ont été proposés par Moody et Darken (Moody et Darken 1989). Les RFR seront donc employés dans les mêmes types de problèmes que les perceptrons multicouches à savoir, en classification et en approximation de fonctions, particulièrement. L'apprentissage le plus utilisé pour les RFR est le mode hybride (Talbi 2011).

D. Cartes auto-organisatrices de Kohonen (SOM)

Les cartes topologiques ou cartes auto organisatrices ont été introduites pour la première fois par Kohonen en 1982 (Kohonen 1982). Ce sont des réseaux à apprentissage non-supervisé qui

établissent une carte discrète, ordonnée topo-logiquement, en fonction de patterns d'entrée. Le réseau forme ainsi une sorte de treillis dont chaque nœud est un neurone associé à un vecteur de poids. La correspondance entre chaque vecteur de poids est calculée pour chaque entrée. Par la suite, le vecteur de poids ayant la meilleure corrélation, ainsi que certains de ses voisins, vont être modifiés afin d'augmenter encore cette corrélation (Talbi 2011).

E. Réseaux de Hopfield

Les réseaux de Hopfield (Hopfield 1982) sont des réseaux récurrents et entièrement connectés. Dans ce type de réseau, chaque neurone est connecté à chaque autre neurone et il n'y a aucune différence entre les neurones d'entrée et de sortie. Ils fonctionnent comme une mémoire associative non-linéaire et sont capables de trouver un objet stocké en fonction de représentations partielles ou bruitées. L'application principale des réseaux de Hopfield est l'entrepôt de connaissances, mais aussi la résolution de problèmes d'optimisation. Le mode d'apprentissage utilisé ici est le mode non-supervisé (Talbi 2011).

F. Réseaux ART

Les réseaux ART (en anglais Adaptive Resonance Theory) sont des réseaux à apprentissage par compétition (Carpenter et Grossberg1987 ; Carpenter et Grossberg 1988). Dans ce genre de réseau, les vecteurs de poids ne seront adaptés que si l'entrée fournie est suffisamment proche d'un prototype déjà connu par le réseau. Il est à noter qu'il existe deux principaux types de réseaux ART : les ART-1 pour des entrées binaires et les ART-2 pour des entrées continues. Le mode d'apprentissage des ART peut être supervisé ou non (Talbi 2011).

II.4.8. Conception d'un modèle neuronal

Les réseaux de neurones réalisent des fonctions non linéaires paramétrées. Leurs mises en œuvre nécessitent :

- ✓ La détermination des entrées et des sorties pertinentes, c'est-à-dire les grandeurs qui ont une influence significative sur le phénomène que l'on cherche à modéliser ;
- ✓ La construction de la base de données et le choix des échantillons ;
- ✓ La détermination du nombre de neurones cachés nécessaires pour obtenir une approximation satisfaisante ;
- ✓ La réalisation de l'apprentissage ;
- ✓ L'évaluation des performances du réseau de neurones à l'issue de l'apprentissage.

II.4.9. Détermination des entrées/sorties du réseau de neurones

Pour toute conception de modèle, la sélection des entrées doit prendre en compte deux points essentiels (Ammar 2007) :

- Premièrement, la dimension intrinsèque du vecteur des entrées doit être aussi petite que possible, en d'autres termes, la représentation des entrées doit être la plus compacte possible, tout en conservant pour l'essentiel la même quantité d'information et en gardant à l'esprit que les différentes entrées doivent être indépendantes.
- En second lieu, toutes les informations présentées dans les entrées doivent être pertinentes pour la grandeur que l'on cherche à modéliser : elles doivent donc avoir une influence réelle sur la valeur de la sortie.

II.4.10. Construction de la base de données et choix des échantillons

Le processus d'élaboration d'un réseau de neurones commence par la construction d'une base de données et le choix des échantillons. Il faut donc déterminer les grandeurs qui ont une influence significative sur le phénomène que l'on cherche à modéliser.

Cette étape est cruciale et aide à déterminer le type de réseau le plus approprié pour résoudre un problème donné. En particulier, la façon dont se présente l'échantillon conditionne le type de réseau, le nombre de cellules d'entrée et de sortie et la façon dont il faudra mener l'apprentissage, le test et la validation (Bishop 1995).

Afin de développer une application à base de réseaux de neurones, il est nécessaire de disposer de deux bases de données, une pour effectuer l'apprentissage et l'autre pour tester le réseau obtenu et déterminer ses performances.

Pour mieux contrôler la phase d'apprentissage, il est souvent préférable de posséder une troisième base de données appelée "base de validation croisée" (Stone 1977). Les avantages liés à l'utilisation de cette troisième base de données seront exposés dans les paragraphes suivants. Notons qu'il n'y pas de règle pour déterminer ce partage d'une manière quantitative, néanmoins chaque base doit satisfaire aux contraintes de représentativité de chaque classe de données et doit généralement refléter la distribution réelle, c'est-à-dire la probabilité d'occurrence des diverses classes (Nascimento et al. 2000).

Généralement, les bases de données subissent un prétraitement qui consiste à effectuer une normalisation appropriée tenant compte de l'amplitude des valeurs acceptées par le réseau. Les valeurs d'entrées et de sortie sont normalisées dans un intervalle spécifique afin de donner à chaque paramètre la même influence statistique. De plus, un problème additionnel peut survenir à cause d'une possible saturation de la sortie, si la somme pondérée des entrées du perceptron multicouches ($\sum(w_i x_i)$) est dans la partie asymptotique de la fonction d'activation (Dames 2005).

II.4.11. Définition de la structure du réseau

La détermination de l'architecture optimale est restée pendant longtemps un problème ouvert. Il existe actuellement diverses méthodes empiriques mettant notamment en jeu des tests statistiques pour une vaste classe de réseaux (Urbani et al. 1994). Comme nous l'avons mentionné, nous avons retenu le perceptron multicouche comme base du modèle. Nous structurons ce réseau en précisant le nombre de couches et de neurones cachés pour que le réseau soit en mesure de reproduire ce qui est déterministe dans les données.

A. Nombre de couches cachées

Mis à part les couches d'entrée et de sortie, il faut décider du nombre de couches intermédiaires ou cachées. Sans couche cachée, le réseau n'offre que de faibles possibilités d'adaptation. Néanmoins, il a été démontré qu'un perceptron multicouche avec une seule couche cachée pourvue d'un nombre suffisant de neurones, peut approximer n'importe quelle fonction avec la précision souhaitée (Hornik 1991).

B. Nombre de neurones cachés

Chaque neurone peut prendre en compte des profils spécifiques de neurones d'entrée. Un nombre plus important permet donc de mieux "coller" aux données présentées, mais diminue la capacité de généralisation du réseau. Il faut alors trouver le nombre adéquat de neurones cachés nécessaire pour obtenir une approximation satisfaisante. Le nombre de neurones cachés dépend du type de données dont on dispose. En général, ce nombre est déterminé par des règles empiriques. Suivant les auteurs, la taille de la couche cachée doit être (Dames 2005) :

- Soit égale à celle de la couche d'entrée ;
- Soit égale à 75 % de celle-ci ;
- Soit égale à la racine carrée du produit du nombre de neurones dans la couche d'entrée et de sortie.

Pour des cas plus sophistiqués, la recherche consisterait soit à :

- Estimer un réseau comportant de nombreux neurones puis à le simplifier par l'analyse des multi-colinéarités ou par une règle d'apprentissage éliminant les neurones inutiles (Dames 2005) ;
- Définir une architecture tenant compte de la structure des variables identifiées au préalable par une analyse en composantes principales (Bermejo et Cabestany 1999; Bermejo et Cabestany 2001)
- Augmenter progressivement la complexité du modèle à chaque itération, dite "Cascading" et réévaluer l'efficacité de généralisation à chaque addition d'un nouveau neurone (Dames 2005).

II.4.12. Apprentissage

Une fois l'architecture choisie, elle doit subir une phase d'apprentissage. Cette dernière consiste à calculer les coefficients synaptiques de telle manière que les sorties du réseau de neurones soient, pour les exemples utilisés lors de l'apprentissage, aussi proches que possibles des sorties "désirées". Celles-ci correspondent à la valeur de la fonction que l'on veut approcher ou à la sortie du processus que l'on veut modéliser (Dreyfus 2008).

L'apprentissage correspond donc à la phase du développement du réseau durant laquelle les coefficients synaptiques sont modifiés jusqu'à l'obtention du comportement désiré.

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 (Dreyfus 2008).

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 (en anglais 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 rétro-propagation du gradient de l'erreur (en anglais Back-propagation) (Dames 2005). On distingue deux étapes dans la phase d'apprentissage (Dames 2005) :

- Une phase de propagation, qui consiste à présenter une configuration d'entrée au réseau, puis à la propager à celle de sortie en passant par les couches cachées.
- Une phase de rétro-propagation, qui consiste, après le processus de propagation, à minimiser l'erreur commise sur l'ensemble des exemples présentés, erreur considérée comme une fonction des poids synaptiques. Cette erreur représente la somme des différences au carré entre les réponses calculées et celles désirées pour tous les exemples contenus dans l'ensemble de l'apprentissage.

Les algorithmes de rétro-propagation implantés actuellement dans les logiciels de RNA appartiennent à une des cinq classes suivantes (Si-Moussa 2009):

- Descente du gradient
- Descente des gradients conjugués
- Quasi-Newton
- Levenberg-Marquardt
- Régularisation

Tableau II.2 résume les algorithmes d'apprentissage par rétro-propagation implanté dans la boite à outils des RN du programme MATLAB® (Demuth et Beale 2000).

Tableau II.2. Résumé des caractéristiques des algorithmes d'apprentissage de rétro-propagation implantés dans MATLAB® (Khaouane 2013).

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, ne peut être utilisée que pour apprentissage en mode séquentiel

Descente du gradient avec taux d'apprentissage 9variable (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	
Descente des gradients conjugués avec la méthode de mise à jour de Fletcher-Reeves (traincgf)	Il possède la plus petite mémoire de stockage parmi les algorithmes de gradient conjugués.
Descente des gradients conjugués avec la méthode de mise à jour de Polak-Ribiére (traincgp)	Nécessite une mémoire de stockage légèrement plus grande que traincgf. Convergence plus rapide sur quelques problèmes.
Descente des gradients conjugués avec la méthode de reprises Powell-Beale (traincgb)	Nécessite une mémoire de stockage légèrement plus grande que traincgp. Convergence Généralement plus rapide.
Scaledconjugate 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 Hessiènne et nécessite plus de calcul à chaque itération que les algorithmes de gradients conjugués, mais possède habituellement une convergence plus rapide.

II.4.13. Evaluation des performances du réseau de neurones

Enfin, la dernière étape doit permettre d'estimer la qualité du réseau obtenu en lui présentant des exemples qui ne font pas partie de l'ensemble d'apprentissage. Elle permet à la fois d'apprécier les performances du système neuronal et de détecter le type de données qui posent problème. Si les

performances ne sont pas satisfaisantes, il faudra, soit modifié l'architecture du réseau, soit modifié la base d'apprentissage (Dames 2005).

II.5. Conclusion

La littérature met en évidence que la modélisation peut permettre de valider les hypothèses simplificatrices émises lors de la formulation mathématique des phénomènes de transfert. Peu de travaux ont appliqué les relations quantitatives structure-activité (QSAR) dans la modélisation des processus membranaires, ils peuvent être considérés comme des outils importants pour la prédiction de la rétention des pesticides lors des membranes d'osmose inverse et de nanofiltration.

En guise de conclusion, cette revue bibliographique nous a servi à fixer l'objectif de ce travail qui consiste en la prédiction de la rétention des pesticides lors d'osmose inverse et de nanofiltration en utilisant des réseaux de neurones artificiels (RNA) et les relations quantitatives structure-propriétés (QSPR). Toutefois, nous avons fait en sorte de combler toutes les lacunes relevées dans les publications antérieures et ce en :

- ✓ choisissant une base la plus large possible ;
- ✓ utilisant une large base de descripteurs moléculaires, caractéristiques de membrane et les conditions opératoires et en déterminer les plus pertinents ;
- ✓ procédant à une optimisation pour le choix de nombre de neurones optimaux dans la couche cachée et pour le choix des fonctions de transfert.

Chapitre III
Matériels et méthodes



III.1. Introduction

Les techniques de modélisation décrites dans le chapitre II ont été appliquées aux données expérimentales de la rétention des pesticides lors de nano filtration es osmose inverse.

Ce chapitre comportera trois axes. Dans le premier axe, on examinera les données collectées, prétraitement et analyse. Le second axe sera consacré aux techniques les plus récentes pour développer des modèles de réseaux de neurones artificiels (RNA) et relation quantitative structure-activité (QSPR). Le troisième axe traitera l'analyse de la sensibilité basée sur la méthode de "poids". Le quatrième axe présentera le diagnostic des points aberrants et domaine d'applicabilité du modèle neuronal optimisé.

III.2. Collecte des données, prétraitement et analyse

Dans cette étude, nous avons utilisé les données disponibles dans la littérature (Yoob et al. 2005 ; Konstantions et al. 2008 et 2011 ; Anaa et al. 2008 et 2009 ; Verliefde et al. 2007, 2008 et 2009 ; Chuyang et al. 2009 ; Comertona et al. 2009 ; Yangali-Quintanilla et al. 2008, 2009 et 2010 ; Huang et al. 2011 ; Riungu et al. 2012 ; Sadmani et al. 2014 ; Nikhbakht et al. 2019 ; Hai 2020 ; Jeong et al. 2021).

La liste des 23 pesticides est répertoriée dans annexe A. La taille de la base de données et le nombre de pesticides sont montrés dans le tableau III.1.

Tableau III.1. Taille de la base de données et le nombre de pesticides utilisés dans ce travail.

Base de données "BD"	QSPR-RN	Taille de base de données	Nombre de pesticides
BD1 avec SR	QSPR-RN1	253	21
BD2 avec SR(MgSO4)	QSPR-RN2	241	21
BD3 avec SR(NaCl)	QSPR-RN3	227	15

Le choix des variables d'entrées et de sortie a été basé sur les interactions entre les propriétés des pesticides, les caractéristiques de la membrane et les conditions opératoires de filtration pour la rétention des pesticides par les membranes d'OI/NF. Les entrées considérées dans cette étude sont :

les descripteurs moléculaires (Mw, dc, Log D, le moment dipolaire, la longueur moléculaire "length", la surface minimale "S_{min}", la surface maximale "S_{max}", la largeur moléculaire "Width", la profondeur moléculaire "Depth" et la largeur équivalente moléculaire "Eqwidth"), les caractéristiques de la membrane (le poids moléculaire de coupure "MWCO" , le rejet de sel de sulfate de magnésium le "SR($MgSO_4$)", "SR(NaCl)" charge de surface de la membrane zêta "potentiel " et l'hydrophobicité de la membrane "angle de contact") et les conditions opératoires (pH, pression, température et le taux de récupération).

Les valeurs de LogD, moment dipolaire, surface minimale "S_{min}", surface maximale "S_{max}" ont été calculées par les logiciels qui sont mentionnés dans le tableau III.2. Sur l'annexe B, nous donnons la capture d'écran de deux sites et deux logiciels utilisés.

Tableau III.2. Logiciels utilisés pour calculer LogD, le moment dipolaire, Smin, Smax et Length.

	Logiciels
LogD	Chemspider http://www.chemspider.com/
Moment dipolaire	Hyperchem
S_{min}	
S_{max}	
Length	ChemBio3D Ultra (Version 12 ; 2012)

Les valeurs de la largeur équivalente moléculaire (EqWidth) ont été calculées par l'équation I.6 (voir le premier chapitre). Le nombre total de variables (les propriétés des produits pharmaceutique, les caractéristiques membranaires et les conditions d'opératoires) a été réduit en utilisant des matrices de corrélation (voir annexe C).

Chaque matrice regroupe l'intensité de la liaison qui existe entre ces paramètres. Cette intensité de liaison a été obtenue par le calcul des coefficients de corrélation de Pearson (R) (Équation III.1) (Ammi 2016)

$$R(Z,Y) = \frac{\sum_i (Z_i - \bar{z})(y_i - \bar{y})}{\sqrt{\sum_i (z_i - \bar{z})^2}\sqrt{\sum_i (y_i - \bar{y})^2}} \qquad (III.1)$$

$Z_i$est la $i^{ème}$valeur du vecteur Z (d_c, LogD, Length, Dipole moment, S_{min} ,S_{max}, Depth, width, EqWidth, MWCO, SR (NaCl), angle de contact, potentiel zêta, pH, pression, température, taux de récupération et rétention) ; yi est la $i^{ème}$valeurdu vecteur Y (d_c, LogD, Length, Dipole moment, S_{min},

S_{max}, Depth, width, EqWidth, MWCO, SR (MgSO4), angle de contact, potentiel zêta, pH, pression, température, taux de récupération et rétention) ; $\bar{z}$ et $\bar{y}$sont respectivement les moyennesarithmétiques des vecteurs *Z* et *Y*. Le degré de coefficient de corrélation de Pearson doit être modéré (R<0.75) (Ammi 2016), On élimine l'entrée qui possède une relation entre eux et les l'entrée qui ont la même valeur pour plus de 0.75 des échantillons.

Pour chaque base de données (BD1, BD2 et BD3), un modèle QSPR a été élaboré et un ensemble de variables d'entrées et de sortie a été identifié (Tableau III.3).

Tableau III.3. Entrées et sortie utilisées pour le développement des modèles QSPR-RN (QSPR-RN1, QSPR-RN2 et QSPR-RN3)

	QSPR-RN1	**QSPR-RN2**	**QSPR-RN3**
Entrées	- dc - Log D - Moment dipolaire - Length - EqWidth - MWCO - SR (MgSO4 ou NaCl) - Zeta potentiel - Angle de contact - pH - Pression - Taux de récupération - Température	- dc - Log D - Moment dipolaire - Length - EqWidth - MWCO - SR (MgSO4) - Zeta potentiel - Angle de contact - pH - Pression - Taux de récupération - Température	- dc - Log D - Moment dipolaire - Length - EqWidth - MWCO - SR (NaCl) - Zeta potentiel - Angle de contact - pH - Pression - Taux de récupération - Température
Sortie	Rétention	Rétention	Rétention

L'analyse statistique des données d'entrées et de sortie, en termes du minimum "min", de la moyenne arithmétique (Équation III.2) "$\bar{x}$", du maximum "max" et de la déviation standard (écart type) "DS" (Équation III.3) (Khaouane 2013), sont présentées dans les tableaux III.4, III.5 et III.6.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \text{(III. 2)}$$

$$DS = \left(\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2\right)^{1/2} \qquad \text{(III. 3)}$$

Tableau III.4. Analyse statistique des données d'entrée et de sortie (BD1).

	Min	**Max**	**Moyenne**	**DS**
dc (g/mol)	0.288416	0.771483	0.684461	0.055827
LogD	-2.2300	178.5000	2.7028	11.4967
Moment dipolaire	0.040000	5.570000	2.267392	1.229808
Length (nm)	0.120000	3.510000	0.230167	0.362841
EqWidth	0.590000	1.080000	1.001000	0.065863
MWCO (Dalton)	96.0000	400.0000	206.6542	85.4807
SR(MgSO4 ou NaCl)	0.100000	0.992000	0.944408	0.167322
Zeta potentiel (mv)	-87.0000	-1.0000	-25.8939	21.9814
Angle de contact (°)	17.50000	94.70000	49.29300	15.95053
pH	3.39000	10.00000	7.25254	0.82768
Pression (KPa)	240.000	2500.000	898.301	634.904
Taux de récupération (%)	0.50000	90.13330	37.45385	27.80016
Température (C°)	13.00000	25.00000	22.82708	2.71797
Rétention %	2.0000	100	79.6376	20.6933

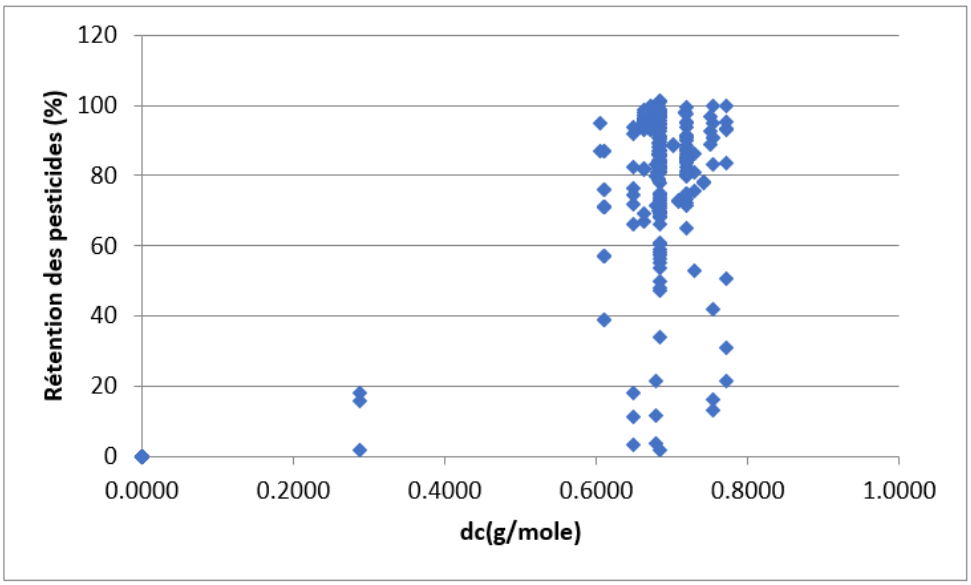
Tableau III.5. Analyse statistique des données d'entrée et de sortie (BD2).

	Min	Max	Moyenne	DS
dc (g/mol)	0.288416	0.771483	0.684329	0.054733
LogD	-2.2300	178.5000	2.6757	11.2234
Moment dipolaire	0.040000	5.570000	2.289857	1.227860
Length (nm)	0.120000	3.510000	0.226865	0.354382
EqWidth	0.590000	1.080000	1.001944	0.065033
MWCO (Dalton)	96.0000	400.0000	202.7659	85.7370
SR (MgSO4) (%)	0.100000	0.996000	0.864409	0.169754
Zeta potentiel (mv)	-87.00000	-1.0000	-25.3883	21.7658
Angle de contact (°)	7.00000	94.70000	49.77389	16.17660
pH	3.39000	10.00000	7.26036	0.81557
Pression (KPa)	240.000	2500.000	901.755	630.560
Taux de récupération(%)	0.50000	90.13330	37.49573	27.36695
Température (C°)	13.00000	25.00000	22.86706	2.674432
Rétention %	2.0000	100.0000	80.0860	20.4452

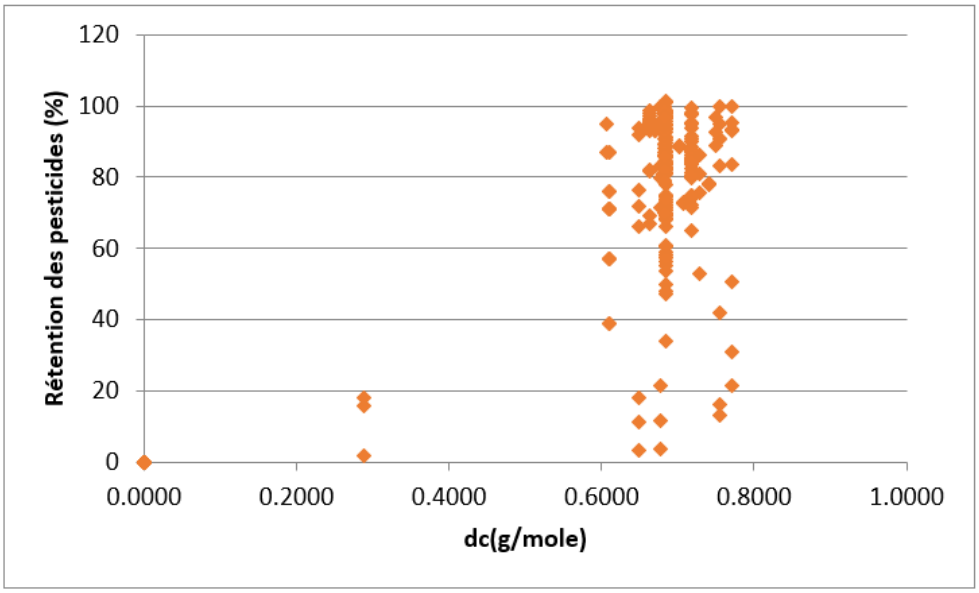
Tableau III.6. Analyse statistique des données d'entrée et de sortie (BD3).

	Min	Max	Moyenne	DS
MW (g/mol)	0.605768	0.7711483	0.686402	0.030812
LogD	-2.2300	178.5000	2.7426	11.88473
Moment dipolaire	0.040000	5.570000	2.132239	1.095541
Length (nm)	0.150000	3.510000	0.205354	0.309002
EqWidth	0.820000	1.080000	1.005841	0.047342
MWCO (Dalton)	96.0000	400.0000	203.5265	0.1647
SR (NaCl) (%)	0.500000	0.996000	0.855562	0.167642
Zeta potentiel (mv)	-87.0000	-1.0000	-26.6523	22.6300
Angle de contact (°)	7.00000	94.70000	49.60717	17.00488
pH	3.39000	10.00000	7.25270	0.84804
Pression (KPa)	240.000	250.000	848.273	581.506
Taux de récupération (%)	0.50000	90.13330	38.23196	26.93437
Température (C°)	13.00000	25.00000	22.97566	2.67965
Rétention %	3.3000	100.0000	80.7951	16.4439

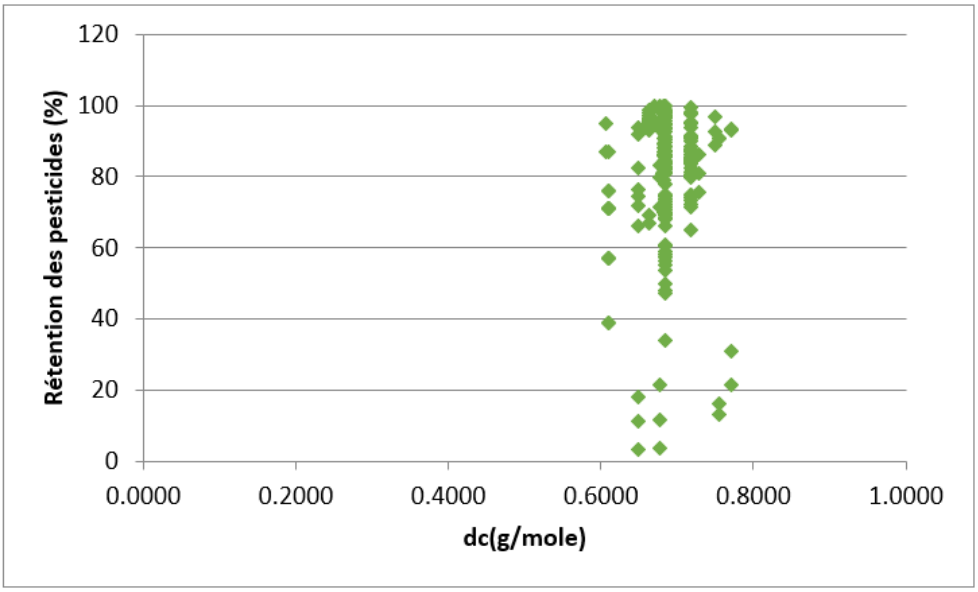
La figure III.1 représente les rétentions expérimentales en fonction le diamètre effectif dans l'eau "dc" pour les trois bases de données (les rétentions en fonction le diamètre effectif du compose organique dans l'eau "dc" sont indiquées par des cercles bleu pour BD1"253 points", des cercles orange pour BD2 "241 « des cercles verts pour BD3 "227"). Quand le diamètre effectif du compose organique dans l'eau "dc" varié de 0.7000g/mol à 0.8000g/mol, les rétentions expérimentales qui ont des pourcentages supérieurs à 60.5 % s'agglomèrent les unes aux autres.



(a)



(b)



(c)

Figure III.1. Rétentions expérimentales en fonction de poids moléculaires : (a) BD1, (b) BD2, (c) BD3.

III.3. Développement des modèles QSPR-RNA

Cette étape concerne la création des modèles de réseaux de neurones artificiels basées sur la relation quantitative structure propriété "QSPR-RNA".

Les réseaux de neurones artificiels (RNA) sont particulièrement bien adaptés pour les modèles QSAR/QSPR en raison de leur capacité à extraire des informations non-linéaires présentes dans la matrice de données (Yangjeh et Jenagharad 2009). Pour trouver un modèle non linéaire entre les propriétés des produits pharmaceutiques, les caractéristiques de la membrane et les conditions opératoires de filtration pour la rétention des produits pharmaceutiques par les membranes d'OI/NF, l'utilisation de la méthode RNA est nécessaire et est devenue une technique de modélisation importante. Elle est largement utilisée pour les études de QSPR non linéaires et est un outil puissant pour construire des modèles prédictifs.

Dans ce travail, une procédure basée sur le développement et l'optimisation de l'architecture d'un QSPR-RNA est décrite. Elle est basée, comme décrite sur l'organigramme ci-dessous.

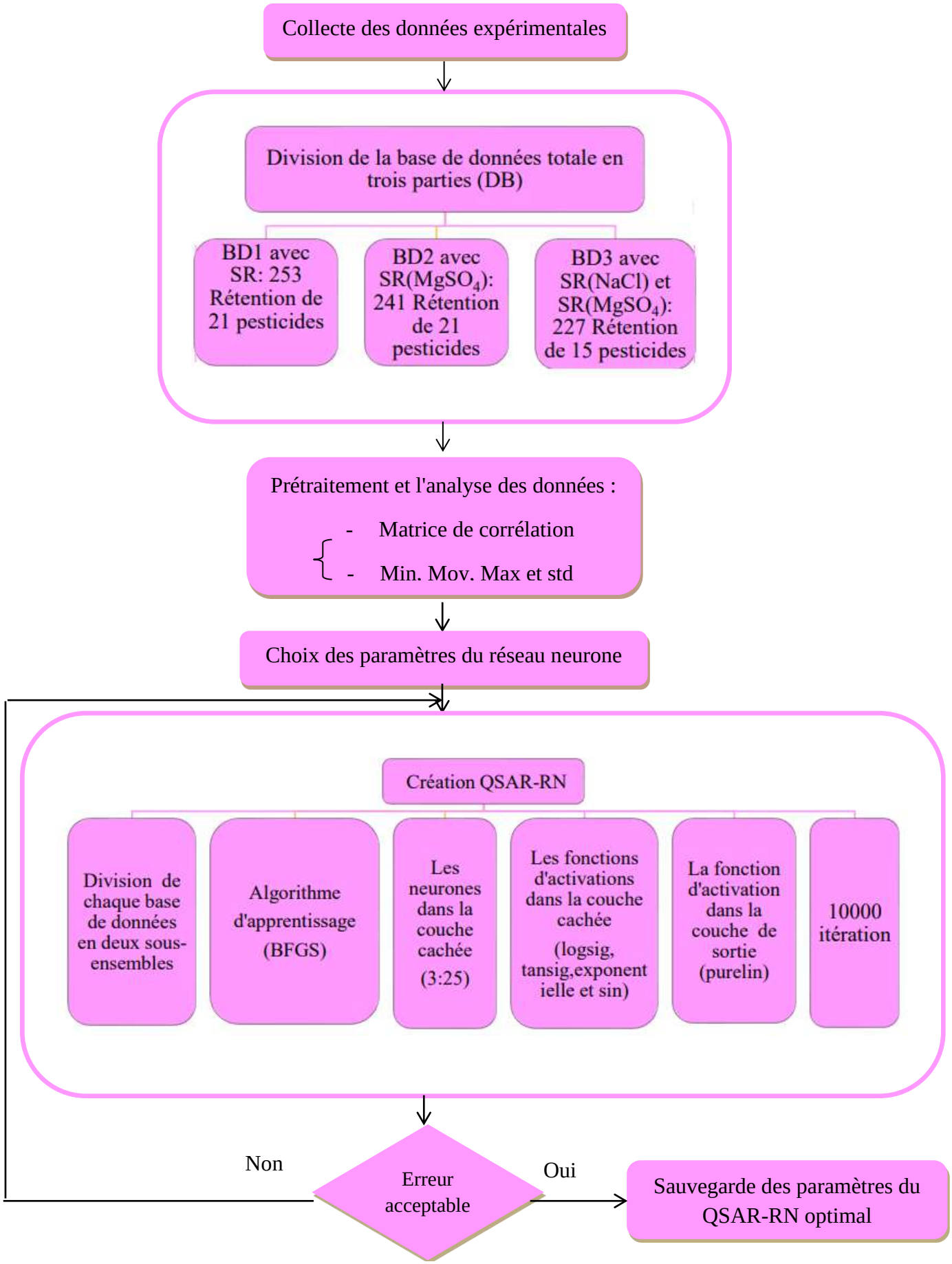


Figure III.3. Organigramme de développement et l'optimisation de l'architecture d'un QSPR-RNA.

L'algorithme d'apprentissage utilisé dans ce travail est le BFGS quasi-Newton (trainbfg). Chaque QSPR-RN contient trois couches de neurones ou des nœuds : une couche d'entrée avec 13 neurones dans la couche d'entrée, une couche cachée avec un nombre de neurones actifs optimisés lors de l'apprentissage et une couche de sortie avec une unité générant la valeur estimée de la rétention. Le nombre de neurones dans la couche cachée a été varié de 3 à 25 neurones. Les fonctions d'activations « La tangente hyperbolique sigmoïde (tansig), la sigmoïde logarithmique (logsig), l'exponentielle et sin » ont été utilisées dans la couche cachée. La fonction de transfert linéaire (purelin) a été utilisée dans la couche de sortie.

La procédure d'élaboration du QSAR-RN décrite ci-dessus a été réalisée en utilisant STATISTICA (Annexe B). Ce logiciel est utilisé pour optimiser l'architecture des quatre modèles neuronaux QSAR-RN employés pour la rétention des produits pharmaceutiques par les membranes d'OI/NF.

III.4. Analyse de sensibilité par la méthode de "poids"

La contribution des variables d'entrées (dc, *LogD*, Moment dipolaire**,** *Lenght*, *EqWidth*, MWCO, SR(NaCl), angle de contact, potentiels zêta, pH, pression, taux de récupération et température) sur la sortie a été déterminée par une analyse de sensibilité en utilisant la méthode de "poids" pour chaque réseau de neurone QSAR-RN

Cette méthode, proposée d'abord par Garson (Garson 1991) et reprise par Goh (Goh 1995), donne une quantification de l'importance relative (*IR*) des différentes entrées sur la sortie du RN (Gevrey et al. 2003). Elle se base sur le partitionnement des poids de connexions en :

- Poids de connexions entrée-cachée (w_{ji}^{I}) ;
- Poids de connexions cachée-sortie (w_{1j}^{H}).

Le processus de calcul, compose de quatre étapes, est schématisé sur l'organigramme ci-dessous (Khaouane 2013).

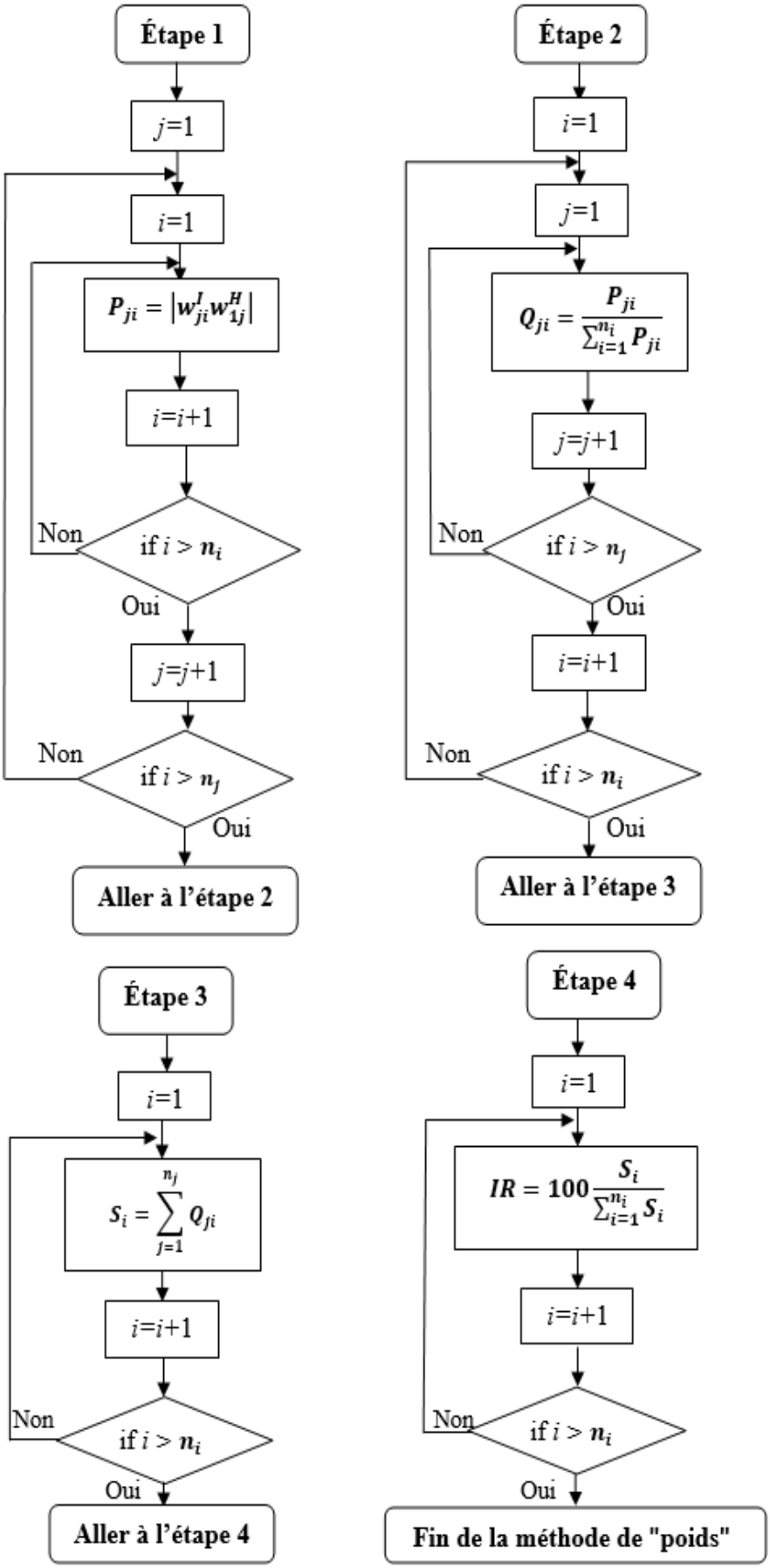


Figure III.3. Organigramme de la méthode de "poids" (Khaouane 2013).

III.5. Diagnostic des points aberrants et domaine d'applicabilité du modèle neuronal optimisé

Les diagrammes de dispersions des modèles optimisés ont montré un certain nombre des points éloignés de la première bissectrice que nous avons attribué à l'existence probable de points aberrants. Ceci nous a motivé à réaliser un diagnostic des données potentiellement suspectes selon la méthode décrite dans les deux références suivantes (Hosseinzadeh and Hemmati-Sarapardeh 2014) ; (Baghban, Mohammadi et al. 2017).

Les valeurs aberrantes sont bien connues en tant que points de données distincts de la tendance générale des principaux points de données. La détection des valeurs aberrantes, qui joue un rôle important dans la proposition de modèles, est utilisée pour reconnaître des données individuelles (ou des groupes de données) pouvant différer de la masse des données constituant un ensemble de données. À cette fin, il est essentiel de trouver des méthodes rigoureuses pour détecter les valeurs potentiellement suspectes afin d'éliminer des données expérimentales inexactes et d'améliorer la précision des modèles (Hosseinzadeh and Hemmati-Sarapardeh 2014) ; (Baghban, Mohammadi et al. 2017). Dans cette étude, la technique mathématique "Leverage" a été utilisée pour rechercher les valeurs aberrantes. Dans cette technique, les valeurs résiduelles sont d'abord calculées. Ensuite une matrice H de points de données d'entrée est créée sur la base de la formule suivante :

$$H = X(X^tX)^{-1}X^t \qquad \text{(III. 4)}$$

Où X fait référence à la matrice m × n (m et n représentent respectivement le nombre d'échantillons et les paramètres (variables d'entrée) du modèle). Les valeurs de Hat sont obtenues à partir de la diagonale principale de la matrice de H.

$$\text{Hat} = \text{diagonale}(H) \qquad \text{(III. 5)}$$

Le diagramme de William qui représente les résidus normalisés en fonction du "Leverage" levier (Hat) peut détecter graphiquement les valeurs aberrantes. Dans ce graphique, le levier critique (H^*) (seuil) est généralement fixé à la valeur donnée par l'équation suivante :

$$H^* = \frac{3(n+1)}{m} \qquad \text{(III. 6)}$$

Ou m et n représentent respectivement le nombre d'échantillons et les paramètres (variables d'entrée) du modèle.

Les résidus normalisés sont calculés à partir des données expérimentales et celle calculée par le modèle.

$$(R_Norm)_i = \frac{(y_i^{exp} - y_i^{cal})}{\sqrt{var(y^{exp} - y^{cal})}} \quad i = 1, \dots m \qquad (III.7)$$

Le résidu normalisé égal à trois est normalement considéré comme une valeur "seuil" pour accepter les points de données dans les écarts types compris sur une plage de ± 3 par rapport à la moyenne (pour couvrir 99% des données normalement distribuées). Si la majorité des points de données se situent dans les plages de $0 \leq$ Hat $\leq H^*$ et de $-3 \leq$ R_Norm ≤ 3, cela indique que le développement du modèle et ses prédictions sont effectués dans le domaine d'applicabilité, ce qui conduit à un modèle statistiquement valide. Ainsi on peut affirmer qu'il existe des points ‘’Good High Leverage’’ dans le domaine de $0 \leq$ Hat $\leq H^*$ et $-3 \leq$ R_Norm ≤ 3. Les points situés dans la plage de R_Norm <-3 ou R_Norm>3 (qu’ils soient plus grands ou plus petits que la valeur H^*) sont désignés comme points aberrants du modèle ou comme points "Bad High Leverage".

III.6. Conclusion

Nous avons décrit ici les outils nécessaires à la modélisation de la rétention des pesticides lors d’osmose inverse et de nano filtration par les modèles QSAR-RNA. Dans cette étape, nous avons présenté les données collectées, prétraitement et analyse. Après nous avons présenté les techniques les plus récentes pour établir des modèles de réseaux de neurones artificiels (RNA) et relation quantitative structure-activité (QSPR). Ensuite, nous avons exposé une analyse de sensibilité basée sur la méthode de "poids". Nous avons terminé ce chapitre par diagnostic des points aberrants et domaine d’applicabilité du modèle neuronal optimisé.

Le chapitre suivant sera consacré aux résultats et discussion pour mieux comprendre la rétention des pesticides lors d’osmose inverse et de nanofiltration par les différentes méthodes.

Chapitre IV

Résultats et discussion



IV.1. Introduction

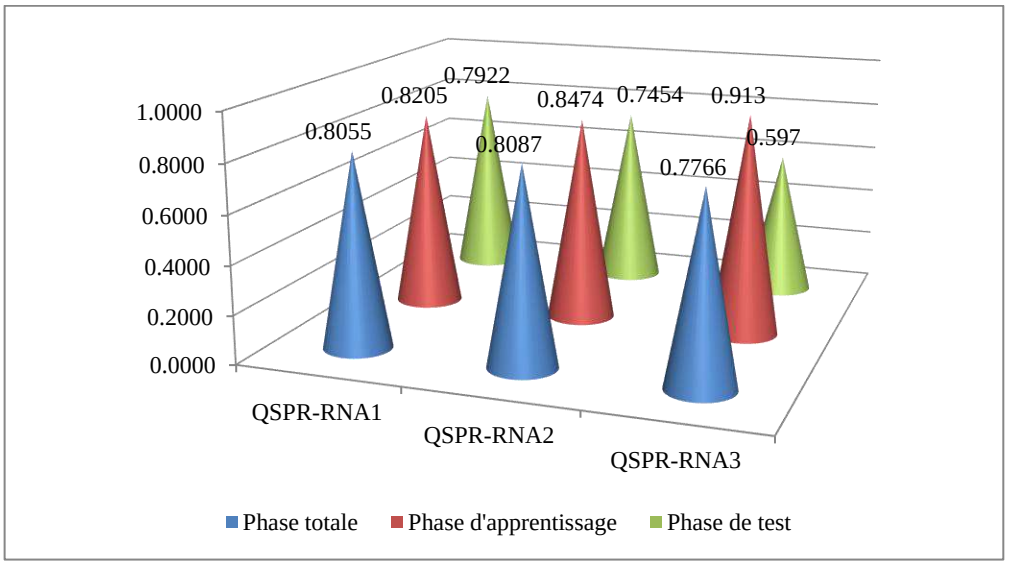
Dans ce chapitre, les résultats obtenus sont discutés, il est organisé en trois parties :

- Dans la première partie, on présentera les performances des modèles QSAR-RN. Ensuite l'extraction de formules mathématiques à partir de ces modèles.
- Par la suite, les résultats de l'analyse de sensibilité par la méthode de "poids dans la deuxième partie est présentée.
- La quatrième partie est consacrée au diagnostic des points aberrants et domaine d'applicabilité du modèle neuronal optimisé

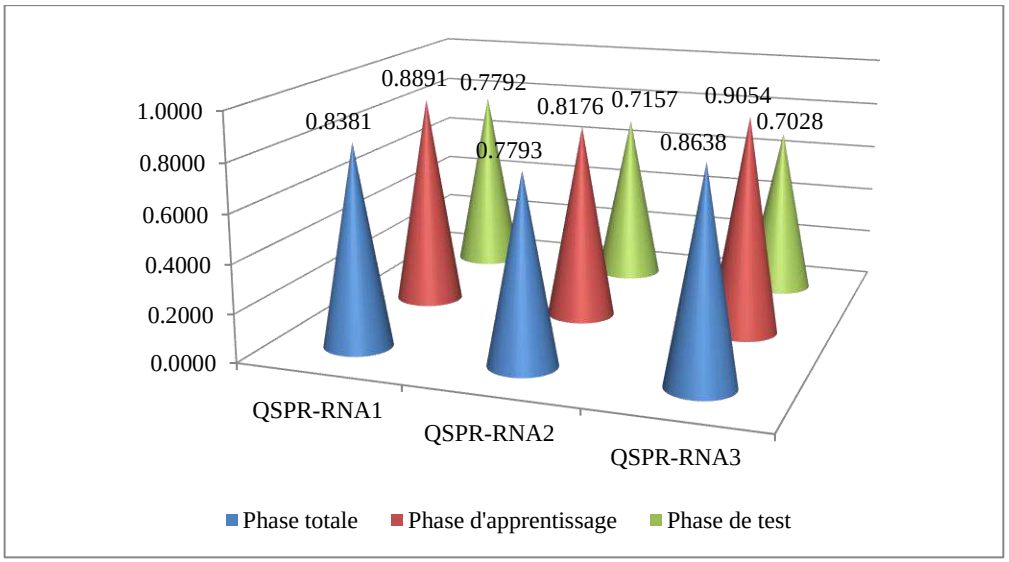
IV.2. Modèles QSAR-RN

2.1. Performances des modèles QSAR-RN

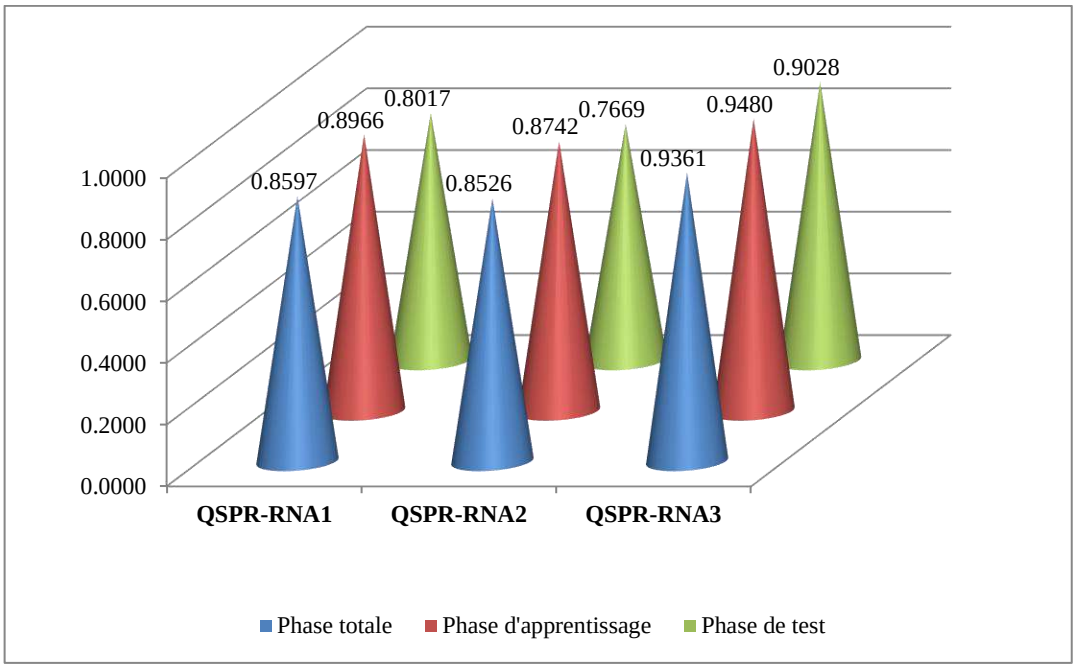
La figure IV.1 montre les coefficients de corrélation des modèles neuronaux "QSPR-RN1, QSPR-RN2 et QSPR-RN3" pour chaque division de base données (division 01, division 02 et division 03). D'après cette figure, on constate que la troisième division est une division optimale qui donne des meilleurs résultats par rapport les autres divisions (R = 0.9361pour la phase totale, 0.9480 pour la phase d'apprentissage, 0.902 pour la phase test).



(a)



(b)



(c)

Figure IV.1. Coefficient de corrélation pour chaque model neuronal "QSPR-RN1, QSPR-RN2 et QSPR-RN3" : (a) division 01, (b) division 02 et (c) division 03.

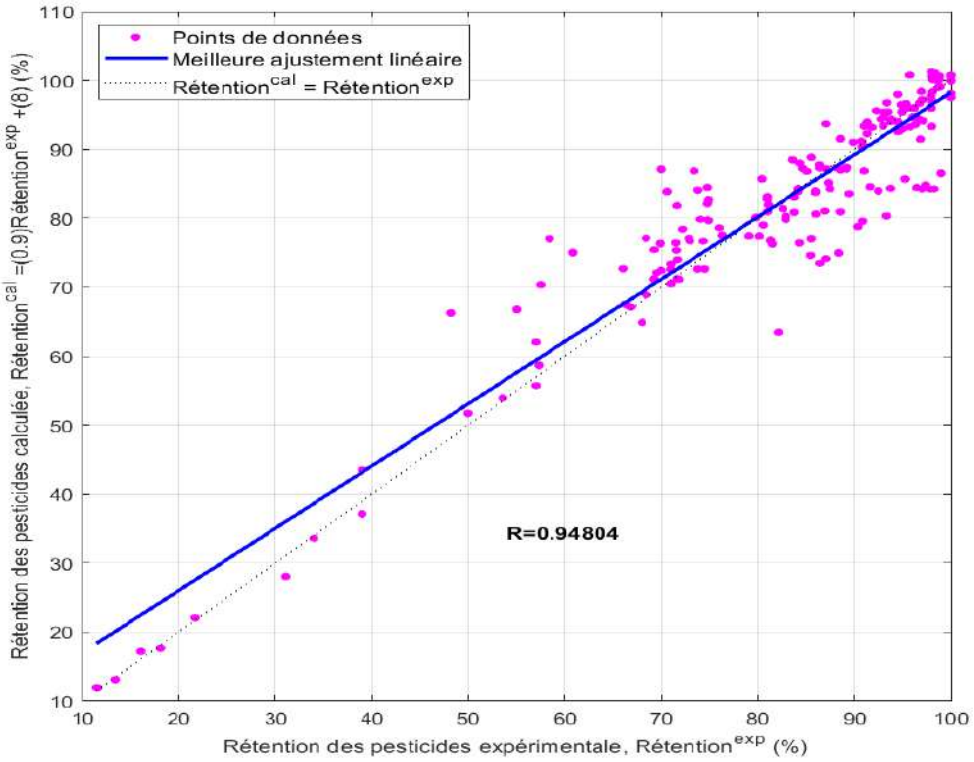
Le tableau IV.1 montre la structure des modèles QSPR-$RN_{optimal}$ optimisé.

Tableau IV.1. Structure des modèles RN optimisés.

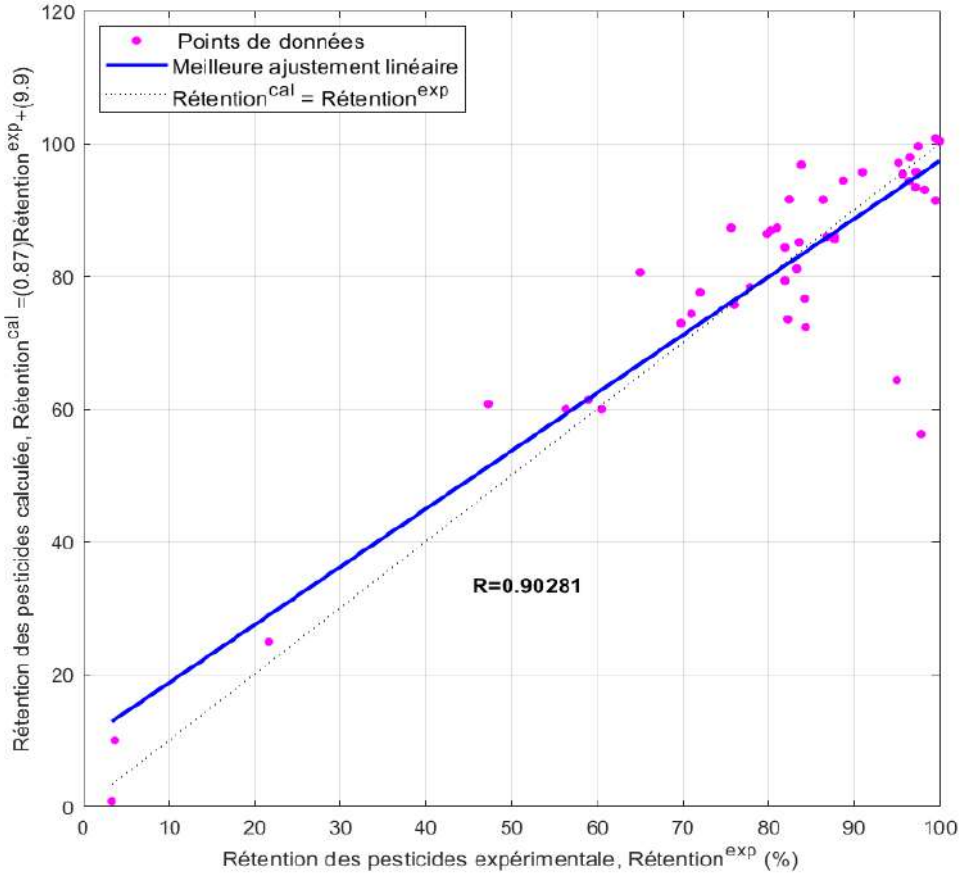
Algorithme d'apprentissage	**Couche entrée**	**Couche cachée**		**Couche sortie**	
	Nombre de neurone	**Nombre de neurone**	**Fonction d'activation**	**Nombre de neurone**	**Fonction d'activation**
BFGS quasi-Newton (trainbfg)	13	17	logistic	1	Identité

Selon la discussion précédente, le modèle neuronal optimal (QSPR-$RN_{optimal}$) a été développé dans le but de prédire la rétention des pesticides par les membranes d'OI/NF. L'intrigue et les paramètres de la régression linéaire sont franchement obtenus en utilisant la fonction MATLAB ***"postreg"***. Une valeur de R supérieure à 0.5 est généralement considérée comme satisfaisante, une valeur supérieure à 0.9 est excellente (Golbraikh et Tropsha 2002).

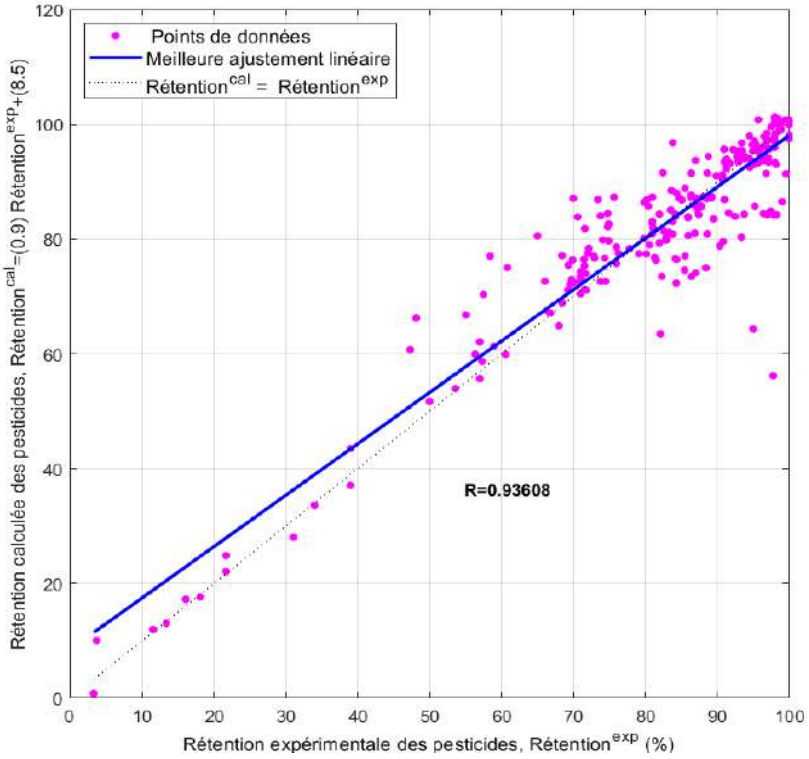
La figure IV.1 montre les courbes de régression linéaire des taux de rétentions calculés par le modèle neuronal "QSPR-$RN_{optimal}$" optimisé avec les taux de rétentions expérimentales avec des vecteurs de régression approchant de l'idéal ([α, β, R] = [0.9027, 7.9644, 0.9480] "phase d'apprentissage" ; [α, β, R] = [0.8742, 9.9374, 0.9028] "phase de test" et [α, β, R] = [0.8955, 8.4942, 0.9361] "phase totale"). Les coefficients des corrélations sont généralement considérés comme excellentes ($0.90 \leq R \leq 1.00$) pour la phase d'apprentissage, la phase de test et la phase totale. Le coefficient de corrélation est couramment considéré comme satisfaisant ($0.50 \leq R < 0.90$) pour toutes les phases, ce qui montre une bonne robustesse du modèle neuronal "QSAR-$RN_{optimal}$" établi et la possibilité de la prédiction des différents paramètres qui caractérisent la rétention des pesticides lors d'OI/NF.


Points de données
Meilleure ajustement linéaire
Rétention^cal = Rétention^exp
R=0.94804
Rétention des pesticides calculée, Rétention^cal =(0.9)Rétention^exp +(8) (%)
Rétention des pesticides expérimentale, Rétention^exp (%)

(a)


Points de données
Meilleure ajustement linéaire
Rétention^cal = Rétention^exp
R=0.90281
Rétention des pesticides calculée, Rétention^cal =(0.87)Rétention^exp+(9.9)
Rétention des pesticides expérimentale, Rétention^exp (%)

(b)



(c)

Figure IV.2. Comparaison entre les valeurs expérimentales et les valeurs calculées pour QSPR-RN$_{optimal}$: (a) la phase d'apprentissage, (b) la phase de test, (c) la phase totale.

La performance de modèle neuronal "QSPR-RN$_{optimal}$" optimisé a été évaluée en termes de coefficient de corrélation (R) et la racine de l'erreur quadratique moyenne (REQM), l'erreur absolue moyenne (EAM) et l'erreur de standard de prédiction (ESP). REQM, EAM et ESP ont été utilisées pour déterminer l'erreur de modélisation entre les valeurs expérimentales et calculées. Les erreurs sont définies comme suit (Faisal et al. 2015 ; Kolluri et al. 2015).

$$REQM = \sqrt{\frac{\sum_{i=1}^{n}(Y_{i,exp} - Y_{i,cal})^2}{n}} \qquad (IV.1)$$

$$EAM = \frac{1}{n}\sum_{i=1}^{n}\left|(y_{i,exp} - y_{i,cal})\right| \qquad (IV.2)$$

$$ESP(\%) = \frac{REQM}{Y_e} * 100 \qquad (IV.3)$$

Où n est le nombre total de données, $y_{i,exp}$ est la valeur expérimentale, $y_{i,cal}$ représente la valeur calculée à partir du modèle QSAR et Y_e est la valeur moyenne des données expérimentales.

Les vecteurs de régression linéaire pour le modèle neuronal "QSPR-RN$_{optimal}$" optimisé et les erreurs (REQM, EAM, ESP) sont présentés dans le tableau IV.2 et la figure IV.3. Il est prouvé que l'approche neuronale proposée donne des résultats satisfaisants avec des valeurs de vecteur de régression approchant à l'idéal [équation linéaire : $y^{cal} = \alpha y^{exp} + \beta$, ***α*=1** (pente), ***β*=0** (ordonnée à l'origine), ***R*=1** (coefficient de corrélation)] dans l'ajustement des profils des QSPR-RN$_{optimal}$.

Table IV.2. Vecteurs de régression linéaire issue de l'équation linéaire : [équation linéaire : $y^{cal} = \alpha y^{exp} + \beta$, ***α***(pente), ***β***(ordonnée à l'origine), ***R***(coefficient de corrélation)] et l'erreur quadratique moyenne ***(REQM***), l'erreur absolue moyenne ***(EAM)*** et l'erreur de standard de prédiction ***(ESP)***.

	α	**B**	**R**	**REQM**	**EAM**	**ESP**
Phase d'apprentissage	0.9027	7.9644	0.9480	5.9068	4.0147	7.2646
Phase de test	0.8742	9.9374	0.9028	9.7160	6.0323	12.3420
Totale	0.8955	8.4942	0.9361	6.8367	4.4164	8.4617

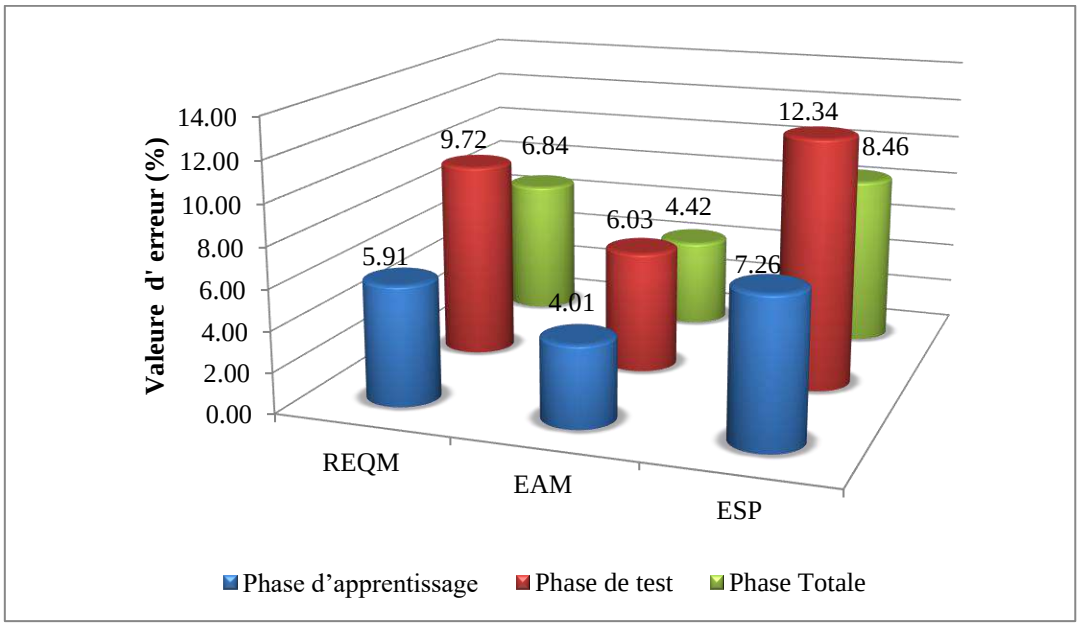


Figure IV.3. Performances des modèles neuronaux (QSAR-RN*1*, QSAR-RN*2* et QSAR-RN*3*) pour : (a) la phase d'apprentissage, (c) la phase de test, (d) la phase totale.

IV.2.2. Expressions mathématiques des modèles neuronaux optimisés

À partir du modèle neuronal "QSPR-RN$_{optimal}$" optimisé, représenté sur la figure IV.4, on peut exprimer l'assimilation de la rétention des pesticides par les membranes d'OI/NF par un modèle mathématique y incorporant toutes les entrées Ei (dc, LogD, moment dipolaire, Length, EqWidth, MWCO, SR(NaCl), angle de contact, potentiel zêta, pH, pression, taux de récupération et température) comme suit : Les sorties Zj de la couche cachée :

$$Z_j = f_H\left[\sum_{i=1}^{13} w_{ji}^I E_i + b_j^H\right] = \frac{1}{1 + exp\left(-\sum_{i=1}^{13} w_{ji}^I E_i + b_j^H\right)} \qquad (IV.2)$$

j = 1,2,….17

La sortie "rétention" :

$$Rétention\ (QSPR - RN_{optimal}) = f_0\left[\sum_{j=1}^{17} w_{1j}^H Z_j + B_1^0\right] = \sum_{j=1}^{17} w_{1j}^H Z_j + B_1^0 \qquad (IV.\mathbf{3})$$

La combinaison des équations (VI.2) et (VI.3) mène à la formule mathématique suivante qui décrit l'assimilation de rétention en prenant en compte toutes les entrées Ei (dc, LogD, moment dipolaire, Length, EqWidth, MWCO, SR(NaCl), angle de contact, potentiel zêta, pH, pression, taux de récupération et température) :

$$Rétention\ (QSPR - RN_{optimal}) = \sum_{j=1}^{17} w_{1j}^H \frac{1}{1 + exp\left(-\sum_{i=1}^{13} w_{ji}^I E_i + b_j^H\right)} + B_1^0 \qquad (IV.\mathbf{4})$$

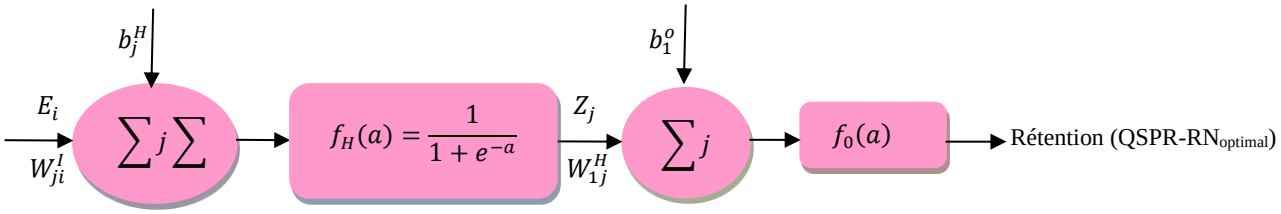


Figure IV.4. Représentation schématique du modèle neuronal "QSPR-RN$_{optimal}$" optimisé.

Ou W_{ji}^I représente le poids synaptique de la connexion du neurone i de la couche d'entrée au neurone j de la couche cachée (17 lignes x 13 colonnes); W_{1j}^H représente le poids synaptique de la connexion du neurone j de la couche cachée au neurone de la couche de sortie (17 lignes x 1 colonne); b_j^H représente les biais du neurone j de la couche cachée (17 lignes x 1 colonne) et b_1^0 représente les

biais du neurone j de la couche de sortie (1 ligne). Le matrice des poids et biais de modèle neuronal "QSPR-$RN_{optimal}$" optimisé est donné dans le tableau IV.3.

Tableau IV.3. Poids et biais du QSPR-RN$_{optimal}$.

Connexions Couche d'entrée-Couche cachée														**Connexions Couche cachée-Couche de sortie**	
Poids													**Biais**	**Poids**	**Biais**
$w_{j,1}^{I}$	$w_{j,2}^{I}$	$w_{j,3}^{I}$	$w_{j,4}^{I}$	$w_{j,5}^{I}$	$w_{j,6}^{I}$	$w_{j,7}^{I}$	$w_{j,8}^{I}$	$w_{j,9}^{I}$	$w_{j,10}^{I}$	$w_{j,11}^{I}$	$w_{j,12}^{I}$	$w_{j,13}^{I}$	b_{j}^{H}	$w_{1,j}^{H}$	b_{1}^{0}
1,5238	13,8715	1,1527	12,5184	0,7919	1,7719	0,8499	-0,0748	-6,3122	2,5234	-0,3184	-3,3907	-2,2513	4,8653	-4,0615	
0,7523	0,5306	0,7140	-0,0309	1,8254	-3,7405	-1,7585	-3,9732	5,7037	-1,3686	-11,1143	0,3043	2,8986	0,2111	-0,3623	
6,4243	-0,8876	4,4973	-0,0605	-3,6531	-1,5350	-0,6237	-3,3897	-2,8423	-4,9953	-1,1899	-6,0073	-7,1989	2,3263	-1,0948	
3,2075	17,0160	-2,4936	0,1568	-6,3325	3,3797	-1,0242	-9,8908	17,1479	2,2333	-4,4667	-17,7991	1,6009	-0,9591	15,8834	
6,9028	-24,4611	-0,5907	-0,0031	1,4146	-1,6145	1,8260	1,7841	4,1134	0,5797	-2,2436	-5,2433	1,1536	1,2459	-12,0426	
6,2258	-4,2353	4,4404	-0,3980	6,8124	0,7682	-1,2140	14,1196	-4,1254	-10,4855	-1,4871	-3,0571	-0,8875	1,7058	0,9222	
5,3667	24,8003	3,0280	-5,1707	1,7753	1,3335	1,0108	1,1097	7,4040	-5,2165	7,8429	-3,0927	-1,4681	-1,1617	3,7402	
8,0007	-18,9606	-1,5302	-8,9263	8,8894	-1,0785	-5,8778	1,3382	-10,1631	6,9935	-19,5288	-6,2312	-0,5486	6,6778	0,3770	
5,5269	15,4327	-2,2830	2,5698	-2,5350	-5,0149	-5,2439	-1,1546	1,1545	0,3202	0,5338	4,2230	-3,1488	-3,4065	0,2246	-3.8275
-0,5349	-0,2274	-7,4285	9,3117	1,6225	-0,4945	8,4938	-5,2236	-0,7915	-2,0422	4,7101	1,3054	-4,7790	5,7218	7,4736	
0,1611	-0,6780	6,7817	18,4092	0,8012	7,0503	4,5219	8,7985	-1,4386	4,7292	-13,1190	-4,5839	-6,8073	3,0173	-0,6707	
3,7412	-0,5217	-18,0582	7,4618	-0,1747	-13,1420	11,8017	6,2276	3,1390	-2,6575	-4,3281	2,4920	3,3623	-3,0879	-0,3884	
2,2525	-1,1635	7,0023	-7,7718	-3,8436	-6,2867	0,0009	1,1071	-0,1563	7,9119	1,4411	5,1905	8,4166	4,1431	-2,7067	
-17,0712	-4,5638	-0,1708	3,4795	-1,5308	6,9977	-6,9455	-0,2846	-3,4848	5,8005	0,8034	-1,5240	-0,4122	16,4597	0,5287	
2,4266	-3,5374	49,9303	1,9109	-2,0567	-5,2698	-3,5334	-1,9751	9,2111	0,0216	-11,2541	-3,2538	5,7151	-2,4113	-0,2413	
-21,4171	-4,2582	5,4868	-0,7874	-4,5055	-5,5500	1,0879	2,2804	5,8122	-3,6269	-2,9858	-8,1735	-5,6723	5,4382	-0,3055	
-6,8048	5,4438	-21,8798	0,5485	-5,9328	4,7463	0,1036	-3,2804	-2,3406	1,0644	1,0823	4,6985	-1,7012	2,0824	0,7311	

L'équations mathématiques, dérivant le modèle neuronal "QSPR-RN$_{optimal}$" optimisé, contient juste le degré exigé de complexité et incluent dans leurs formules les caractéristiques pertinentes de la rétention des pesticides par les membranes d'OI/NF à savoir les propriétés des pesticides, les caractéristiques de la membrane et les conditions opératoires de filtration. Ils peuvent donc facilement être appliqués dans des boucles de contrôle du processus de la rétention des pesticides par les membranes d'OI/NF.

L'architecture des réseaux de neurones optimaux obtenus est donnée sur la figure IV.5, ci-après.

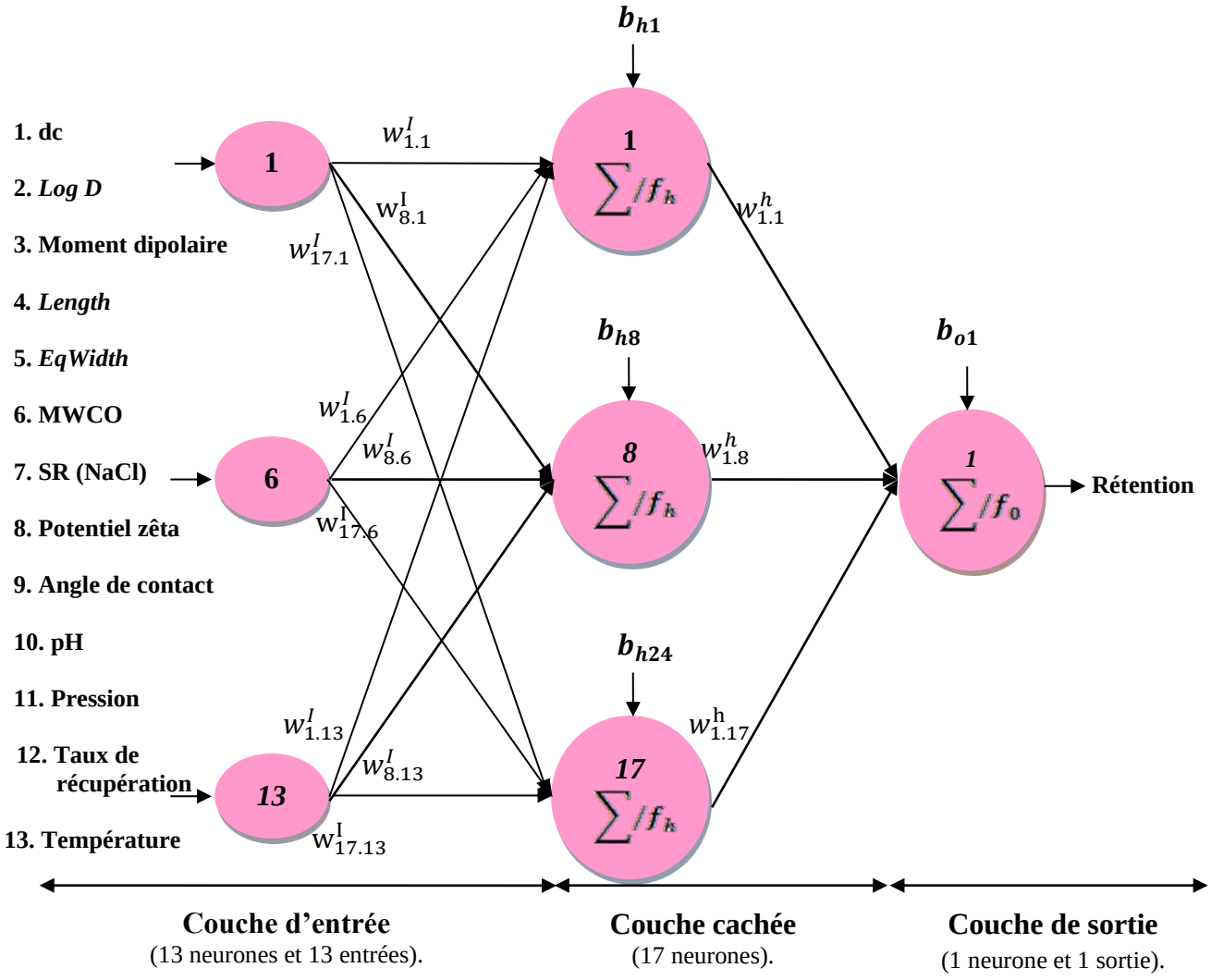


Figure IV.5. Architecture finale du réseau de neurone optimal de type feedforward pour la prédiction de la rétention des pesticides par les membranes d'OI/NF.

IV.3. Analyse de sensibilité par la méthode de "poids"

Pour voir la contribution ainsi que le profil de variation de chaque variable d'entrée (dc, LogD, Dipole moment, Lenght, EqWidth, MWCO, SR(NaCl), angle de contact, potentiels zêta, pH,

pression, taux de récupération et température) sur la sortie de modèle neuronal "QSPR-$RN_{optimal}$" optimisé, On a procédé à une analyse de sensibilité par méthode de "poids"

Nous avons appliqué cette méthode à modèle neuronal "QSPR-$RN_{optimal}$" optimisé en utilisant leurs poids respectifs illustrés dans les tableaux IV.3. Les résultats de contributions sont montrés sur la figure IV.6. Les variables les plus pertinentes qui peuvent influer sur la rétention de pesticides par les membranes d'OI/NF sont : log D, moment dipolaire, dc, pression, angle de contact, length, taux de récupération, MWCO, potentiel zêta, pH, température, SR(NaCl), Eqwidth. Toutes les entrées exercent des contributions importantes ($IR > 5\%$) sur la rétention des pesticides lors d'OI/NF.

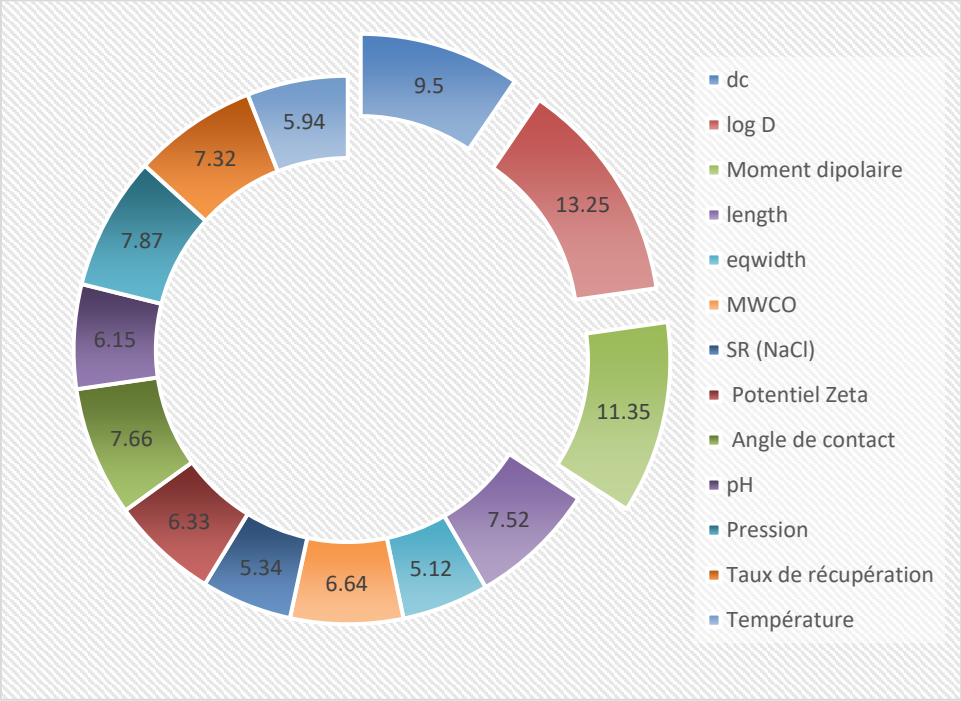
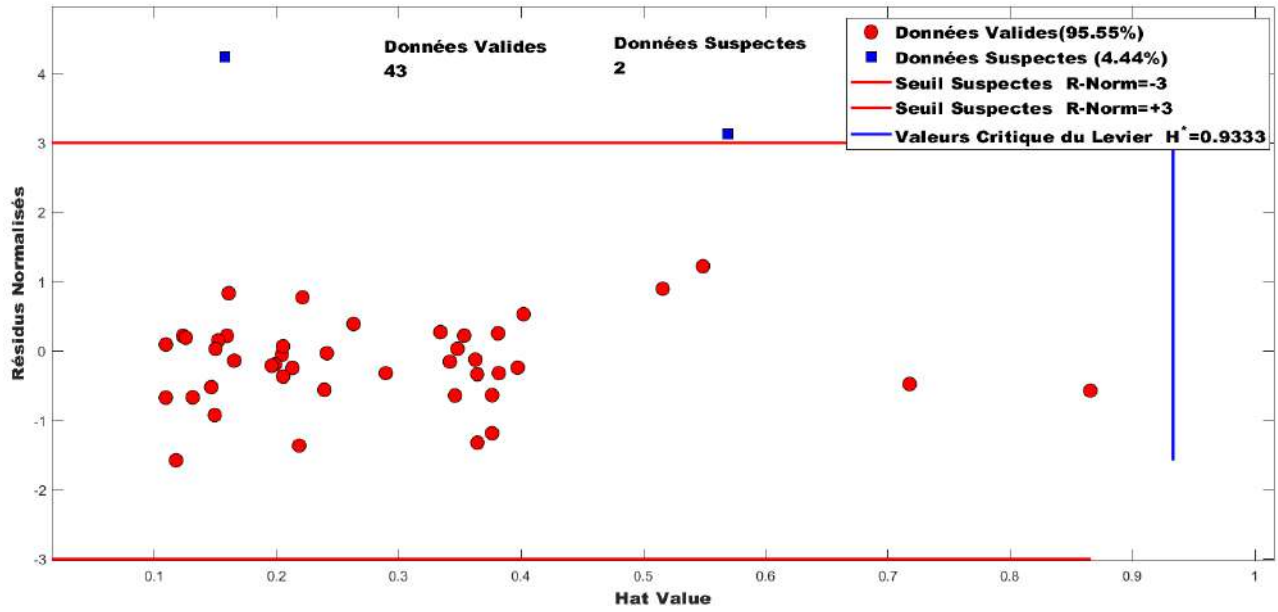


Figure IV.6. Présentation des pourcentages de l'importance relative (*IR*) de chaque variable d'entrée sur la rétention des pesticides lors d'OI/NF.

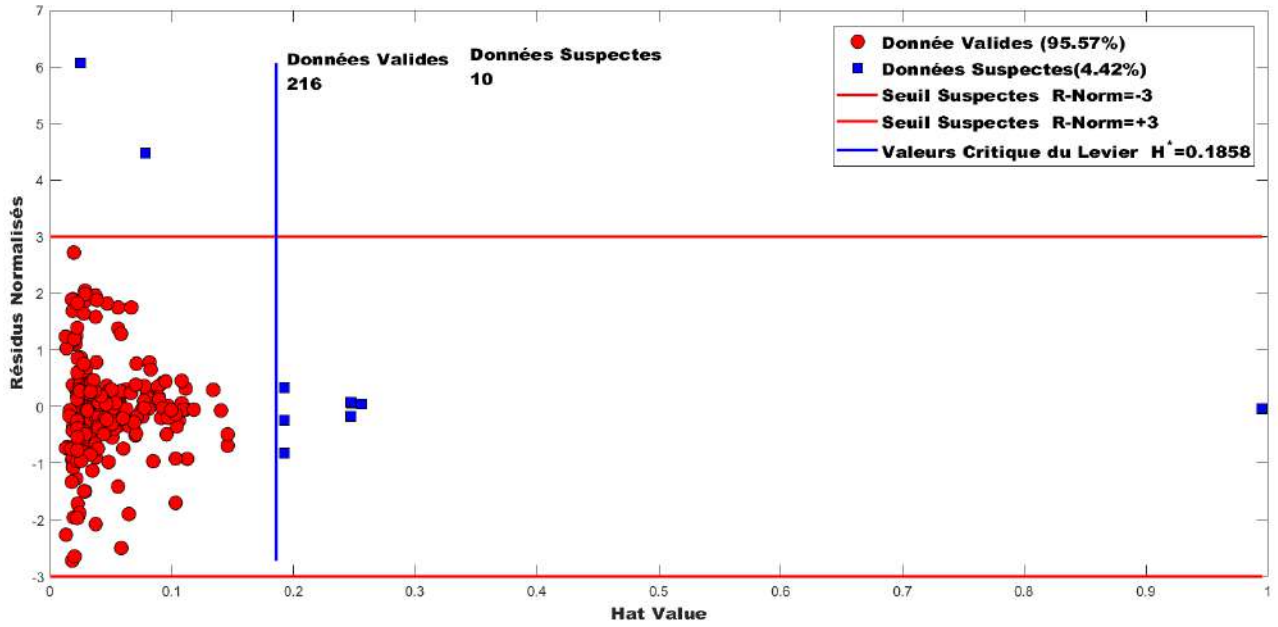
IV.4. Diagnostic des points aberrants et domaine d'applicabilité du modèle neuronal optimisé

La figure III.7 Montre le diagramme de William pour le domaine d'applicabilité pour le modèle neuronal optimal "QSPR-$RN_{optimal}$". Dans ce diagramme l'existence de **43/45 (95.55%)** de points de données valides et 2/45 (4.44%) de points de données suspectes pour la phase test et **216/226 (95.57%)** de points de données valides et 10/226(4.42%) de points de données suspectes pour la phase totale. La valeur critique de levier est $H^* = \frac{3(n+1)}{m} = \frac{3(13+1)}{45} = 0.9333$ pour la phase de test et $H^* = \frac{3(n+1)}{m} = \frac{3(13+1)}{226} = 0.9333$ pour la phase totale. Cela indique que le développement du modèle

QSPR-$RN_{optimal}$ et ses prédictions sont effectués dans le domaine d'applicabilité. Ce qui conduit à un modèle neural optimal statistiquement valide. Ainsi on peut affirmer qu'il existe des points "Good High Leverage" pour la phase validation.



(a)



(b)

Figure IV.7. Diagnostic des données potentiellement suspectes et domaine d'applicabilité du du modèle neuronal optimal : (a) pour la phase test et (b) pour la phase totale.

IV.5. Conclusion

Ce présent chapitre a donné une raison de plus pour l'utilisation généralisée des modèles QSPR-RN pour la modélisation de la rétention des pesticides par les membranes d'osmose inverse et de nanofiltration. Le but de la présente étude était de développer le modèle QSPR-RN capables de résumer les interactions entre les caractéristiques de la membrane, les conditions d'opératoire de filtration et les propriétés physico-chimiques des pesticides

Le modèle QSAR-$RN_{optimal}$ développé dans ce travail a donné des erreurs faibles. Les erreurs absolues moyennes étaient 4.4164% pour phase totale, 4.0147% pour la phase apprentissage et 6.0323% pour phase test. Les racines des erreurs quadratiques moyennes étaient 6.8367% pour phase totale, 5.9068% pour phase apprentissage et 9.7160% pour la phase test. Les erreurs standards de prédictions étaient 8.4617% pour phase totale, 7.2646% pour phase apprentissage et 12.3420% pour phase test.

L'analyse de sensibilité par la méthode de "poids" a révélé que les variables les plus importantes qui influent sur la prédiction de la rétention de pesticides sont : log D, moment dipolaire, dc, pression, angle de contact, length, taux de récupération, MWCO, potentiel zêta, pH, température, SR(NaCl), Eqwidth. Toutes les entrées exercent des contributions plus importantes ($IR > 5\%$) sur la rétention des pesticides lors d'OI/NF.

Analyse diagnostic des points aberrants et domaine d'applicabilité du modèle neuronal optimisé a montré que le développement du modèle QSPR-$RN_{optimal}$ et ses prédictions sont effectués dans le domaine d'applicabilité. Ce qui conduit à un modèle neural optimal statistiquement valide. Ainsi on peut affirmer qu'il existe des points "Good High Leverage" pour la phase de test.

Conclusion générale et perspectives

Les réseaux de neurones possèdent une propriété remarquable dont l'apport est manifesté pour la modélisation des processus non linéaires : ce sont des approximateurs universels parcimonieux. Les modèles des réseaux neurones artificiels (RNA) ont finalement été appliqués pour la prévision de séparation des micropolluantes par les membranes d'osmose inverse et de nanofiltration qui ont été abordés par d'autres chercheurs d'une façon réductible dont la taille de la base de données et les paramètres calculés ont donné des résultats encourageants.

L'objectif de ce mémoire a porté sur l'utilisation des réseaux de neurones pour la modélisation du processus de la rétention des pesticides par les membranes d'osmose inverse et de nanofiltration. Un modèle neuronal optimal de type feedforward ($QSAR\text{-}RN_{optimal}$) caractérisé par une structure (13 neurones pour $QSAR\text{-}RN_{optimal}$ et un neurone dans la couche de sortie) sont construits dans le but de prédire la rétention des pesticides lors d'osmose inverse et de nanofiltration. Le modèle QSAR-RNoptimal développé dans ce travail a donné des erreurs faibles. Les erreurs absolues moyennes étaient 4.4164% pour phase totale, 4.0147% pour la phase apprentissage et 6.0323% pour phase test. Les racines des erreurs quadratiques moyennes étaient 6.8367% pour phase totale, 5.9068% pour phase apprentissage et 9.7160% pour la phase test. Les erreurs standards de prédictions étaient 8.4617% pour phase totale, 7.2646% pour phase apprentissage et 12.3420% pour phase test. Ce qui montre une bonne robustesse du modèle neuronal "$QSAR\text{-}RN_{optimal}$" établi et la possibilité de la prédiction des différents paramètres qui caractérisent la rétention des pesticides lors d'OI/NF.

L'analyse de sensibilité par la méthode de "poids" a révélé que les variables les plus importantes qui influent sur la prédiction de la rétention de pesticides sont : log D, moment dipolaire, dc, pression, angle de contact, length, taux de récupération, MWCO, potentiel zêta, pH, température, SR(NaCl), Eqwidth. Toutes les entrées exercent des contributions plus importantes ($IR > 5\%$) sur la rétention des pesticides lors d'OI/NF.

Analyse diagnostic des points aberrants et domaine d'applicabilité du modèle neuronal optimisé a montré que le développement du modèle $QSPR\text{-}RN_{optimal}$ et ses prédictions sont effectués dans le domaine d'applicabilité. Ce qui conduit à un modèle neural optimal statistiquement valide. Ainsi on peut affirmer qu'il existe des points "Good High Leverage" pour la phase de test.

À la lumière des résultats obtenus, on peut tirer les conclusions suivantes :

- Les réseaux de neurones sont des outils statistiques, qui permettent d'ajuster des fonctions non linéaires très générales à des ensembles de points comme toute méthode statistique ;
- Les réseaux de neurones fonctionnent comme une boîte noire et les paramètres du réseau (poids et biais) n'ont pas de signification physique. L'application des réseaux neuronaux ne développe pas directement de la théorie de description du système, mais rend faisable la modélisation et la simulation des procédés d'osmose inverse et de nanofiltration ;
- La modélisation par réseau de neurones est une méthode qui est basée seulement sur des données expérimentales et/ou semi-expérimentales (déduites des modèles analytiques, empiriques, numériques, …), donc la qualité et la quantité de données décident de l'exactitude et la robustesse du modèle ;
- Les réseaux de neurones peuvent être très robustes et performants même en présence de peu de données expérimentales à condition que ces dernières soient statistiquement bien réparties dans le domaine d'entrée du modèle neuronal, ce qui est le cas avec le plan composite centré ;
- Les formules mathématiques extraites des modèles neuronaux peuvent trouver des applications intéressantes dans la prédiction, l'optimisation et même le contrôle des procédés d'osmose inverse et de nanofiltration.

De ce travail de ce mémoire ont émergé de nombreuses pistes de recherche qui feront l'objet de futurs travaux.

- ✓ Les perspectives concernent plus particulièrement la modélisation de la rétention des pesticides par les membranes d'OI/NF, mais en utilisant d'autres indicateurs de porosité, on propose de faire une étude avec le rejet de sel SR($CaCl_2$).
- ✓ Les perspectives concernent plus particulièrement la modélisation d'autres processus (microfiltration, ultrafiltration) pour des fins de prédiction.
- ✓ Nous pouvons aussi étendre l'idée de ce travail à d'autres méthodes de l'intelligence artificielle (Support vecteur machine, Réseaux de neurones multiples,…).

Références Bibliographiques

Agenson K. O., Oh J. I. et Urase T. (2003). "Retention of a wide variety of organic pollutants by different nanofiltration/reverse osmosis membranes : controlling parameters of process". Journal of Membrane Science. 225 (1-2) : 91-103.

Altalyana N., Jonesa B., Bradda J., Nghiemb L. D. et Alyazichia Y. M. (2006)."Removal of volatile organic compounds (VOCs) from groundwater byreverse osmosis and nanofiltration". Journal of Water Process Engineering. 9 : 9-21

Ammi Y. (2016) "Modélisation des phénomènes de transport lors d'osmose inverse " Thèse de doctorat. Université Yahia Feres. Médéa. Algérie.

Ammi Y., Khaouane L. et Hanini S. (2015)."Prediction of the rejection of organic compounds (neutral and ionic) by nanofiltration and reverse osmosis membranes using neural networks".Korean Journal of Chemical Engineering.**32** (11) : 2300-2310.

Ammi Y., Khaouane L. et Hanini S. (2018)."A Model Based on Bootstrapped Neural Networks for Modeling the Removal of Organic Compounds by Nanofiltration and Reverse Osmosis Membranes" Arabian Journal for Science and Engineering. 43 (11): 6271-6284.

Ammi Y., Khaouane L. et Hanini S (2020). "A Comparison of Models "Neural Networks and Multiple Linear Regressions" to Describe the Rejection of Micropollutants by Membranes" Journal of chemists and chemical engineers of Croatia. 69 (3-4) : 111–127.

Ammi Y., Khaouane L. et Hanini S (2021). "Stacked neural networks for predicting the membranes performance by treating the pharmaceutical active compounds" Neural Computing and Applications. **33** (12429–12444).

Ammi Y., Hanini S et Khaouane L. (2021). "An artificial intelligence approach for modeling the rejection of anti-inflammatory drugs by nanofiltration and reverse osmosis membranes using kernel support vector machine and neural networks" Comptes Rendus. Chimie. **24** (2) : 243-254.

Anselme C. et Jacobs E. P. (1996). "Ultrafiltration, In : Water Treatment Membrane Process". Chapitre 10, McGraw-Hill : 88.

Bellona C. et Drewes J. E. (2005). "The role of membrane surface charge and solute physico-chemical properties in the rejection of organic acids by NF membranes". Journal of Membrane Science. 249 (1) : 227-234.

Bellona C., Drewes J. E., Xu P. et Amy G. (2004). "Factors affecting the rejection of organic solutes during NF/RO treatment - a literature review". Water Research. 38 (12) : 2795-2809.

Berg P., Hagmeyer G. et Gimbel R. (1997). "Removal of pesticides and other micropollutants by nanofiltration". Desalination. 113 (2-3) : 205-208.

Boussahel R., Baudu M., Montiel A. (2022)."Influence de la matière organique et inorganique de l'eau sur l'élimination des pesticides par nanofiltration " Revue des sciences de l'eau Journal of Water Science. 5 (4) 2002.

Chabane M. (2015)."Étude de la résistance au colmatage et de la rétention des composes toxiques par membranes". Thèse de doctorat. Université d'Aboubekr Belkaid de Tlemcen. Tlemcen. Algérie.

Chouachi , Kaddouri (2021). "Étude de l'unité de dessalement de l'eau de mer MTM-Mostaganem" mémoire de master.Université Abdel Hamid Ibn Badis.Mostaganem Algérie.

Cornelissen E. R., Sijbers P., Van den Berkmortel H., Koning J., De Wit A., De Nil F. et Van Impe J. F. (2001). "Reuse of leachate waste-water using MEMBIOR technology and reverse osmosis". Membrane Technology. 136 : 6-9.

Desmukh S. S. et Childress A. E. (2001). "Zeta potential of commercial RO membranes: influence of source water type and chemistry". Desalination. 140 (1) : 87-95.

Dolar D., Ana V., Danijela A. et Kresimir., K. (2011). "Effect of water matrices on removal of veterinary pharmaceuticals by nanofiltration and reverse osmosis membranes". Journal of Environmental Sciences. 23 (8) : 1299–1307.

Girard B., Fukumoto L. R. (2000). "Membrane processing of fruit juices and beverages: a review". Critical Reviews in Biotechnology. 20 (2) : 109-175.

Hamani (2016). Récupération et séparation d'ions métalliques par la méthode de membranes liquides émulsionnées " mémoire de master. Université A. Mira. Bejaia. Algérie.
https://culturesciences.chimie.ens.fr/thematiques/chimie-des-materiaux/chimie-des-surfaces/les-techniques-membranaires-differences-et (Consulté le 02.02.2022).

https://culturesciences.chimie.ens.fr/thematiques/chimie-des-materiaux/chimie-des-surfaces/les-techniques-membranaires-differences-et (Consulté le 21.02.2022).

https://www.linseis.com/fr/grandeurs-mesurees/angle-de-contact/ (Consulté le 21.02.2022).

http://www.chemspider.com/Chemical-Structure.6886.html?rid=0ae6e03a-77d9-40b1-a7cd-44816ea08f94&page_num=0) (Consulté le 21.02.2022).

https://fr.wikipedia.org/wiki/Bio-informatique) (Consulté le 21.02.2022).

https://www.lebigdata.fr/reseau-de-neurones-artificiels definition#:~:text=(apprentissage%20automatique).,L'histoire%20des%20r%C3%A9seaux%20de%20neurones%20artificiels,et%20le%20math%C3%A9maticien%20Walter%20Pitts) (Consulté le 27.02.2022).

https://scilogs.fr/intelligence-mecanique/reprenons-bases-neurone-artificiel-neurone-biologique/). (Consulté le 27.02.2022).

(https://scilogs.fr/intelligence-mecanique/reprenons-bases-neurone-artificiel-neurone-biologique/). (Consulté le 02.03.2022).

https://scilogs.fr/intelligence-mecanique/reprenons-bases-neurone-artificiel-neurone-biologique (Consulté le 05.03.2022).

http://www.statsoft.fr/concepts-statistiques/reseaux-de-neurones-automatises/reseaux-de-neurones-automatises.htm#.YjJNDtXMLIU (Consulté le 11.03.2022).

http://www.chemspider.com (Consulté le 12.04.2022).

Jean-Marc B, Catherine J (2002) " Les procédés membranaires pour le traitement de l'eau ".Direction de l'Espace Rural et de la Forêt

Kimura K., Toshima S., Amy G. et Watanabe Y. (2004). "Rejection of neutral endocrine disrupting compounds (EDCs) and pharmaceutical active compounds (PhACs) by RO membranes". Journal of Membrane Science. 245 (1-2) : 71-87.

Kiso Y., Kltao T., Jinnob K. et Miyae M. (1992)."The effects of molecular width on permeation of organic solute through cellulose acetate reverse osmosis membranes". Journal of Membrane Science. 74 : 95-103.

Kiso Y., Kon T., Kitao T. et Nishimura K. (2001). "Rejection properties of alkyl phthalates with nanofiltration membranes". Journal of Membrane Science. 182 (1-2) : 205-214

Kiso Y., Muroshige K., Oguchi T., Yamada T., Hhirose M., Ohara T. et Shintani T. (2010). "Effect of molecular shape on rejection of uncharged organic compounds by nanofiltration membranes and on calculated pore radii". Journal of Membrane Science. 358 (1-2) : 101–113

Kosutic K., Kastelan-Kunst L. et Kunst B. (2000). "Porosity of some commercial reverse osmosis and nanofiltration polyamide thin-film composite membranes". Journal of Membrane Science. 168 (1-2) : 101-108.

Lanteri Y. (2009). "Transport a travers des membranes de nanofiltration : caractérisation des propriétés électronique". Thèse de doctorat. Université de Franche-COMTE. Besançon. France.

Lin S. H. et Yang C. R. (2004). "Chemical and physical pretreatment of chemical mechanical polishing wastewater from semiconductor fabrication". Journal of Hazardous Materials. 108 (1-2) : 103-109

Linares R.V., Yangali-Quintanilla V., Li Z. et Amy G. (2011). "Rejection of micropollutants by clean and fouled forward osmosis membrane". Water Research. 45 (20) : 6737 -6744

Nechad (2009) " Elaboration d'une membrane de filtration d'eau à base de polyéthylène basse densité et de gypse de bouzggaia" mémoire de magister.Université Hassiba ben bouali. Chlef. Algérie.

Nghiem L. D., Schäfer A. I. et Elimelech M., (2004). "Removal of natural hormones by nanofiltration membranes : measurement, modelling, and mechanisms". Environmental Science & Technology . 38 (6) : 1888–1896.

Norberg D., Hong S., Taylor J. et Zhao Y. (2007). "Surface characterization and performance evaluation of commercial fouling resistant low-pressure RO membranes". Desalination. 202 (1-3) : 45-52.

Ozaki H. et Li H. (2002). "Rejection of organic compounds by ultra-low pressure reverse osmosis membrane". Water Research. 36 (1) : 123-130.

Sadmani A. H. M. A., Andrews R.C. et Bagley M. D. (2014)."Nanofiltration of pharmaceutically active and endocrine disrupting compounds as a function of compound interactions with DOM fractions and cations in natural water". Separation and Purification Technology. 122 : 462–471.

Sagne C. (2008). "Etude des mécanismes de transfert de molécules organiques en osmose inverse". Thèse de doctorat. École nationale supérieure de chimie de Montpellier. Montpellier. France

Santos J. L. C., de Beukelaar P., Vankelecom I. F. J., Velizarov S. et Crespo J. G. (2006). "Effect of solute geometry and orientation on the rejection of uncharged compounds by nanofiltration". Separation and Purification Technology. 50 (1) : 122–131.

Schäfer A. I., Nghiem L. D. et Waite T. D. (2003). "Removal of the natural hormone estrone from aqueous solutions using nanofiltration and reverse osmosis". Environmental Science & Technology. 37 (1) : 182-188.

Sridhar S., Kale, A., Khan A. A. (2002). "Reverse osmosis of edible vegetable oil industry effluent". Journal of Membrane Science. 205 (1-2) : 83-90.

Taheran M., Brar S. K., Verma M., SurampalliR. Y., Zhang T. C. et Valero J. R (2016). "Membrane processes for removal of pharmaceutically active compounds (PhACs) from water and wastewaters". Science of The Total Environment. 547 : 60-77.

Van der Bruggen B., Schaep J., Maes W., Wilms D. et Vandecasteele C. (1998). "Nanofiltration as a treatment method for the removal of pesticides from ground waters". Desalination. 117: 139-147

Verliefde A. R. D., Cornelissen E. R., Heijman S. G. J., Verberk J .Q. J. C., Amy G. L., Van der Bruggen B. et van Dijk J. C.(2008). "The role of electrostatic interactions on the rejection of organic solutes in aqueous solutions with nanofiltration". Journal of Membrane Science. 322 (1) : 52–66.

Yangali Quintanilla V. (2010). "Rejection of emerging organic contaminants by nanofiltration and reverse osmosis membranes effects of fouling, modelling and water reuse". Thèse de doctorat. Université de Delft de technologie. Delft. Netherlands.

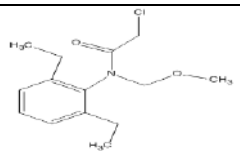
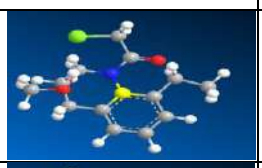
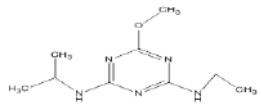
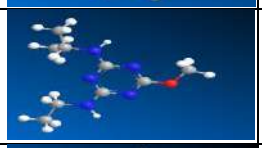
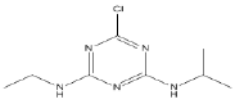
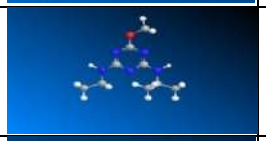
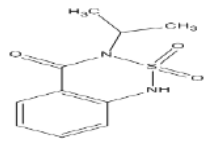
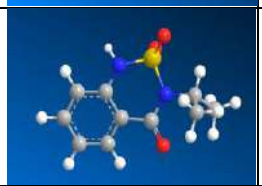
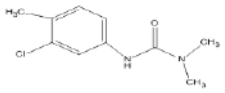
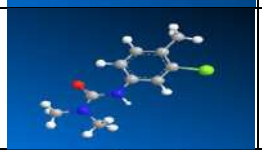
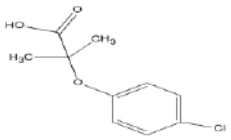
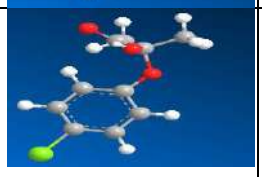
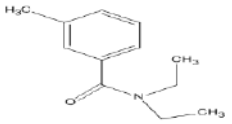
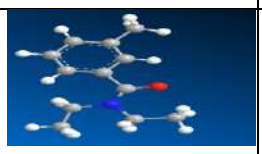
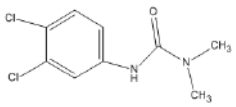
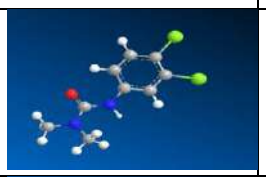
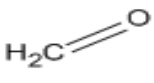
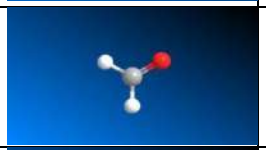
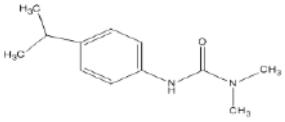
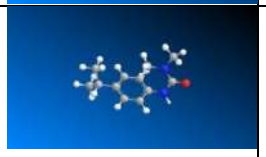
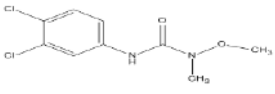
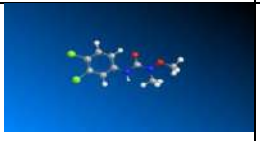
Yangali-Quintanilla V., Verliefde A., Kime T. U., Sadmani A., Kennedy M. et Amy G. (2009). "Artificial neural network models based on QSAR for predicting rejection of neutral organic compounds by polyamide nanofiltration and reverse osmosis membranes". Journal of Memebrane Science. 342 (1-2) : 251-262.

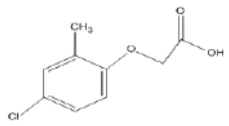
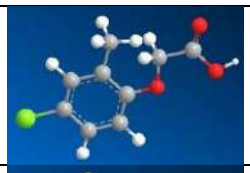
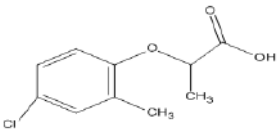
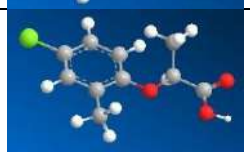
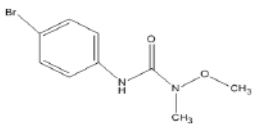
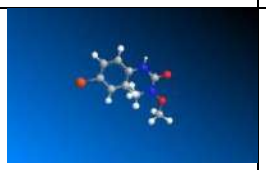
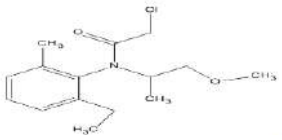
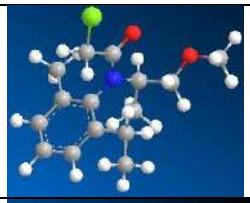
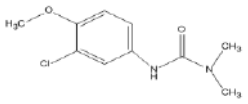
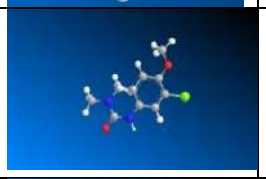
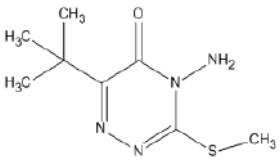
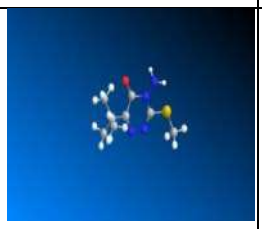
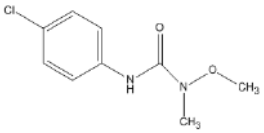
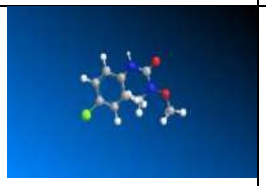
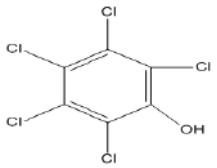
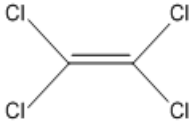
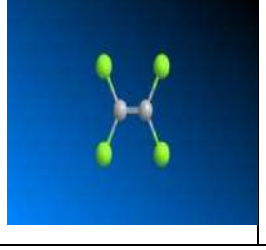
Yangali-Quintanilla V., Kim T. U., Kennedy M. et Amy G. (2008). "Modeling of RO/NF membrane rejection of PhACs and organic compounds : a statistical analysis". Drinking Water Engineering and Science. 1 (1) : 7–15.

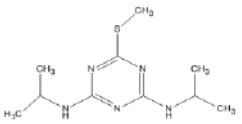
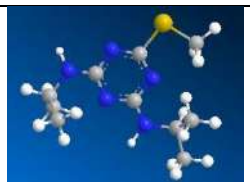
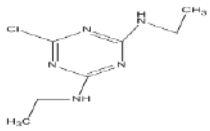
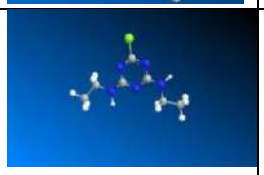
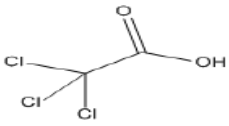
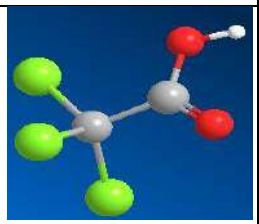
Yangali-Quintanilla V., Sadmani A., McConville M., Kennedy M. et Amy G. (2010b). "A QSAR model for predicting rejection of emerging contaminants (pharmaceuticals, endocrine disruptors) by nanofiltration membranes". Water Research. 44 (2) : 373 – 384.

Annexe A

Tableau 1. Liste de 23 pesticides utilisées pour développer les modèles QSPR-RN.

#	Nom de pesticides	Formule brute	MW (g/mol)	Structure chimique 2D	Structure chimique 3D	Classe de pesticides
1	**Alachlor**	$C_{14}H_{20}ClNO_2$	269.77			Herbicide
2	**Atraton**	$C_9H_{17}N_5O$	211.26			Herbicide
3	**Atrazine**	$C_8H_{14}ClN_5$	215.68			Herbicide
4	**Bentazon**	$C_{10}H_{12}N_2O_3S$	240.28			Herbicide
5	**Chlorotoluron**	$C_{10}H_{13}ClN_2O$	212.68			Herbicide
6	**Clofibric acid**	$C_{10}H_{11}ClO_3$	214.64			Pesticides
7	**DEET**	$C_{12}H_{17}NO$	191.27			Insecticide
8	**Diuron**	$C_9H_{10}CL_2N_2O$	233.09			Herbicide
9	**Formaldehyde**	CH_2O	30.02			Insecticide Fongicide
10	**Isoproturon**	$C_{12}H_{18}N_2O$	206.28			Herbicide
11	**Linuron**	$C_9H_{10}CL_2N_2O_2$	249.09			Herbicide

12	**MCPA**	**$C_9H_9ClO_3$**	200.62			Fongicide Herbicide
13	**MCPP**	**$C_{10}H_{11}ClO_3$**	214.65			Herbicide
14	**Metobromuron**	**$C_9H_{11}BrN_2O_2$**	259.1			Herbicide
15	**Metolachlor**	**$C_{15}H_{22}ClNO_2$**	283.79			Herbicide
16	**Metoxuron**	**$C_{10}H_{13}ClN_2O_2$**	228.67			Herbicide
17	**Metribuzin**	**$C_8H_{14}N_4OS$**	214.29			Herbicide
18	**Monolinuron**	**$C_9H_{11}ClN_2O_2$**	214.05			Herbicide
19	**Pentachlorophenol**	**C_6HCL_5O**	266.34			Fongicide Herbicide
20	**Perchloroethene**	**C_2CL_4**	165.83			Fongicide

21	**Prometryn**	**$C_{10}H_{19}N_5S$**	241.36			Herbicide
22	**Simazin**	**$C_7H_{12}CIN_5$**	201.08			Herbicide
23	**Trichloroacetic acid**	**$C_2HCL_3O_2$**	163.39			Herbicide

Annexe B

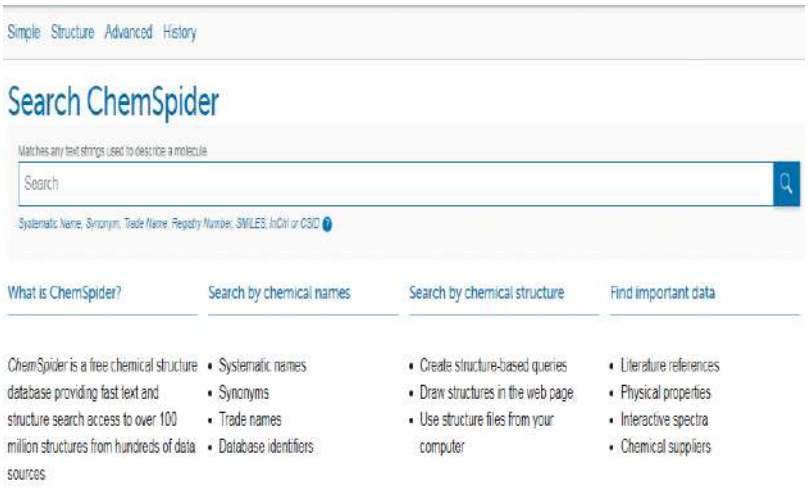
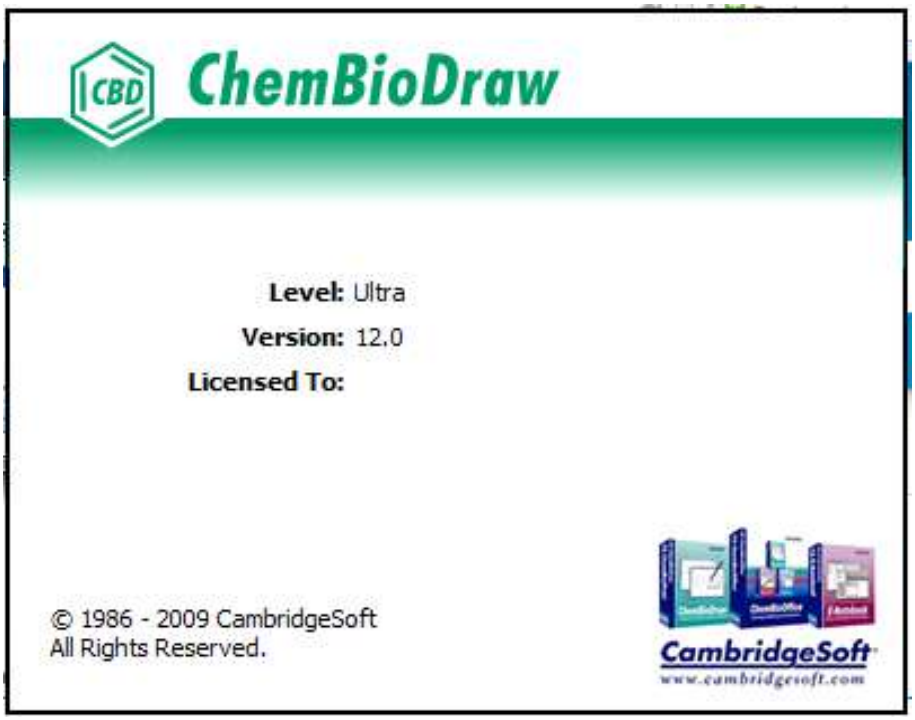


Figure 1. Capture d'écran du site intitulé Chemspider.



(a)



(b)

Figure 2. Capture d'écran des logiciels : (a) ChemBioDraw et (b) ChemBio3D.

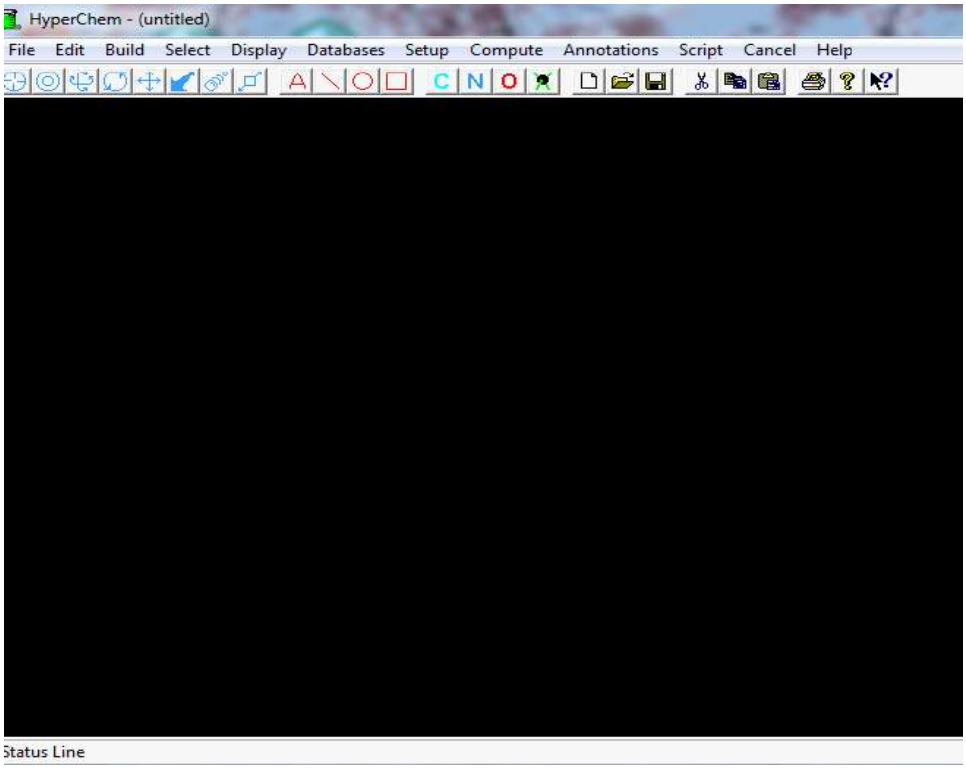


Figure 3. Capture d'écran de logiciel Hyperchem



Figure 4. Capture d'écran de logiciel STATISTICA

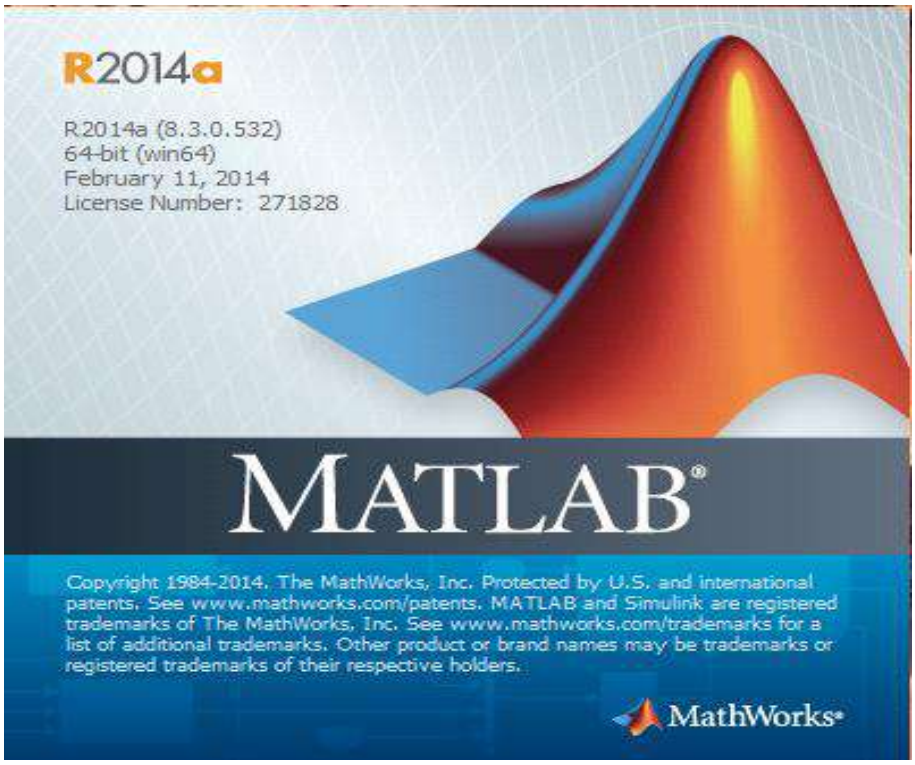


Figure 5. Capture d'écran de logiciel MATLAB.

Annexe C

Tableau 1. Matrice des coefficients de corrélation pour la Base des données.

Nom	dc (g/mole)	LogD	Moment dipolaire (Debey)	Length (nm)	Eqwidth (nm)	MWCO (Dalton)	SR (NaCl)	Potentiel zêta (mv)	Angle de contact (°)	pH	Pression (KPa)	Taux de récupération (%)	Température (C°)	Rétention (%)
dc(g/mol)	1.0000	0.1106	0.1523	0.2666	0.6986	0.2584	-0.0331	0.1234	-0.1618	-0.2208	-0.0530	0.2520	0.0356	0.08822
LogD	0.1106	1.0000	-0.1766	0.7153	0.0727	0.0175	0.0753	0.1094	-0.0395	-0.0651	-0.0955	0.0936	0.1093	0.0218
Moment dipolaire (Debey)	0.1523	-0.1766	1.0000	0.1822	-0.2111	0.1676	0.1321	-0.5822	-0.0040	0.2780	0.1201	0.1557	-0.3653	-0.0111
Length (nm)	0.2666	0.7153	0.1822	1.0000	0.0568	0.1934	0.1061	-0.1993	-0.1057	0.0480	-0.0767	0.1985	-0.0700	-0.0854
Eqwidth (nm)	0.6986	0.0727	-0.2117	0.5685	1.0000	0.1106	0.0799	0.1099	-0.0682	-0.2867	-0.2702	0.3078	0.3239	-0.0943
MWCO (Dalton)	0.2584	0.0175	0.1676	0.1934	0.1106	1.0000	-0.3131	-0.4465	-0.5691	-0.1133	0.0128	0.0276	-0.2885	-0.3336
SR (Nacl)	-0.0331	0.0753	0.1321	0.1061	0.0799	-0.3131	1.0000	-0.2345	0.3571	0.2422	-0.2471	0.2828	-0.0058	-0.1646
Potentiel zêta(mv)	-0.1234	0.0194	-0.5822	-0.1993	0.1099	-0.4465	-0.2345	1.0000	0.1468	-0.3438	0.1013	-0.3007	0.3766	0.2520
Angle de contact (°)	-0.1618	-0.0395	-0.0040	-0.1057	-0.0682	-0.5691	0.3571	0.1468	1.0000	0.1771	-0.1322	0.1432	0.1285	0.2192
pH	-0.2208	-0.0651	0.2780	0.0480	-0.2867	-0.1133	0.2422	0.3438	0.1771	1.0000	0.0421	0.2502	-0.3534	0.0111
Pression (KPa)	-0.0530	-0.0955	0.1201	-0.0767	-0.2702	0.0128	-0.2471	0.1013	-0.1322	0.0421	1.0000	-0.1288	-0.5776	0.01397
Taux de récupération (%)	0.2520	0.0936	0.1557	0.1985	0.3078	0.0276	0.2828	-0.3007	0.1432	0.2502	-0.1288	1.0000	-0.2669	-0.1916
Température (C°)	0.0356	0.1093	-0.3653	-0.0700	0.3239	-0.2885	-0.0058	0.3766	0.1285	-0.3534	-0.5776	-0.2669	1.0000	0.1048
Rétention (%)	0.0822	0.0218	-0.0111	-0.0854	-0.0943	0.3336	-0.1646	0.2520	0.2192	0.0114	0.1397	-0.1916	0.1048	1.0000