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



UNIVERSITE YAHIA FARES DE MEDEA

Faculté des Sciences et de la Technologie

Département de Génie de Procédés et Environnement

MEMOIRE DE MASTER

En vue de l'obtention du Diplôme de Master en Génie des Procédés

Option : Procèdes pharmaceutiques

Thème

Modélisation de la rétention des rejets pharmaceutiques lors d'osmose inverse et nanofiltration

Etudié par :

- ✓ **Mlle KHALDI Aicha**
- ✓ **Mlle SOUIHEL Takia**

Dirigé par :

*** Dr. KHAOUANE Latifa**

*** Dr. AMMI Yamina**

Promotion 2015-2016

Remerciement

On remercie ALLAH le tout puissant de nous avoir donné la santé et la volonté d'entamer et de terminer ce mémoire.

Ce travail ne serait pas aussi riche et n'aurait pas pu voir le jour sans l'aide et l'encadrement de Dr KHAOUANE Latifa, enseignante à l'UYFM et Dr AMMI Yamina, on les remercie pour la qualité de leurs encadrement exceptionnel, pour leurs patiences, leurs rigueurs et leurs disponibilités durant notre préparation de ce mémoire.

Nous tenons également à remercier messieurs les membres de jury pour l'honneur qu'ils nous ont fait en acceptant de siéger à notre soutenance.

Nos remerciements s'adressent également à tous nos professeurs pour leurs générosités et la grande patience dont ils ont su faire preuve malgré leurs charges académiques et professionnelles.

Nos profonds remerciements vont également à toutes les personnes qui nous ont aidés et soutenue de prés ou de loin.

Dédicaces

***A** l'homme de ma vie, mon exemple éternel, mon soutien moral et source de joie et de bonheur, celui qui s'est toujours sacrifié pour me voir réussir, Que ALLAH leur procure bonne santé et longue vie, à toi **mon père**.*

***A** la lumière de mes jours, la source de mes efforts, la flamme de mon cœur, ma vie et mon bonheur ; **maman** que j'adore.*

***A** mes chères frères **Yacine, Hamza, Mohamed et Mnaour** et bien sur a mes sœurs **Amel, Cherifa, Asia et Zhor**. **A** mes nièces et mes neveux. A toute ma famille je dédie ce travail dont le grand plaisir leurs revient en premier lieu pour **l**eurs conseils, aides, et encouragements.*

***A**ux personnes qui m'ont toujours aidé et encouragé, qui étaient toujours à mes côtés, et qui m'ont accompagnaient durant mon chemin d'études supérieures,*

*A toi **Aicha** et à mes aimables amies, collègues d'étude, et à tous ceux qui ont contribué de près ou de loin pour que ce projet soit possible, je vous dis merci.*

Takia

Dédicaces

Je dédie ce modeste travail :

*A **mes parents**. Aucun hommage ne pourrait être à la hauteur de l'amour dont ils ne cessent de me combler. Qu'ALLAH leur procure bonne santé et longue vie.*

*A mes chères frères **Amine**, **Riadh** et **Mohammed**.*

*A celui qui m'a soutenue tout au long de ce projet : mon fiancé **Mohammed**, et bien sur à la personne qui ma toujours aidé et encouragé ma tante **Nadjia**, sans oublié mes beaux-parents que j'aime.*

*A toute ma famille et mes amis. A mon binôme **Takia**.*

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

Aicha

ملخص

نمذجة الاحتفاظ بالنفايات الصيدلانية في التناضح العكسي والترشيح الدقيق جدا

الهدف من هذا العمل هو نمذجة قدرة الاحتفاظ بالجزيئات العضوية عن طريق أغشية التناضح العكسي والترشيح الدقيق بعد القيام بتوحيد المقاييس باستخدام نموذجين: الشبكات العصبية الاصطناعية (QSAR-RNA) ومنحنى الانحدار الخطي المتعدد (QSAR-RLM). المقارنة بين ثلاث قواعد للبيانات الموحدة المقاييس (عادية، مرجعية، مختزلة) بينت تفوق القاعدة المرجعية. المقارنة بين النموذجين أجريت من أجل تأكيد أداء وجودة التنبؤ. النموذج (QSAR-RNA) للقاعدة المرجعية أعطى أفضل أداء من حيث (*REQM=0.113%, EAM=0.08%11, ESP=11.0351% et R=0.9153) (REQM*, l'*EAM*, l'*ESP* et *R*) لمجموع نقاط البيانات.

كلمات البحث: قدرة الاحتفاظ ، النفايات الصيدلانية، التناضح العكسي، الترشيح الدقيق جدا، الشبكات العصبية الاصطناعية، منحنى الانحدار الخطي المتعدد.

Résumé

Modélisation de la rétention des rejets pharmaceutiques lors d'osmose inverse et de nanofiltration

L'objectif de ce travail est de modéliser la rétention des molécules organiques par les membranes d'osmose inverse et nanofiltration en utilisant après une normalisation des données deux modèles QSAR: réseaux de neurones artificiels (QSAR-RNA) et régression linéaire multiple (QSAR-RLM). La comparaison entre les trois bases normalisées (simple, référentielle et réduite) a démontré la supériorité de la base référentielle. Une comparaison entre les deux modèles a été faite afin de constater leurs performances et leurs qualités de prédictions. Le modèle QSAR-RNA pour la base référentielle a donné les meilleures performances en termes de *REQM*, l'*EAM*, l'*ESP* et *R* (*REQM*=0.113%, *EAM*=0.0811%, *ESP*=11.0351% et *R*=0.9153) pour la base totale.

Mots clés: Modélisation, Rétention, Rejets pharmaceutiques, Osmose inverse, Nanofiltration, Réseaux de neurones artificiels, Régression linéaire multiple.

Summary

Modeling the retention of pharmaceuticals wastes in reverse osmosis and nanofiltration

The objective of this work is to model the retention of molecular organic by reverse osmosis and nanofiltration membranes using after data normalization two QSAR models: artificial neural networks and multiple linear regressions. A comparison between the three normalized bases (simple, referential, reduced) was proved the superiority of the referential base. A comparison between the two models was made to find the performance and quality of the prediction. The model of QSAR-ANN for the referential base was gave the best performance in terms of *RMSE, MEA, SEP* and *R (RMSE=* 0.113%, *MEA= 0.0811%, SEP=11.0351%, R* =0.9153) to the total base.

Keywords: retention, pharmaceuticals wastes, Reverse Osmosis, Nanofiltration, Artificial Neural Networks, Multiple linear regressions.

Table des matières

Table des matières i
Abréviations iii
Nomenclature et symboles v
Liste des tableaux viii
Liste des figures ix

Introduction générale 1

Chapitre I. Description de la technologie membranaire

I.1. Introduction 04
I.2. Généralité 04
I.2.1. Définition de la membrane 05
I.2.2. Classification des membranes 05
I.2.2.1. Classification d'après la morphologie 05
I.2.2.2. Classification d'après la nature chimique 06
I.2.2.3. Classification d'après la géométrie de préparation des membranes 07
I.2.2.4. 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 13
I.2.5.3. Taux de récupération 14
I.3. Interactions et propriétés de séparation 14
I.3.1. Encombrement stériques (effet de tamisage) 14
I.3.1.1. État physique de la membrane 14
I.3.1.2. Taille moléculaire 17
I.3.2. Répulsion électrostatique (effet de charge) 18
I.3.2.1. Charge de la surface 18
I.3.2.2. Charge de soluté 19
I.3.2.3. Influence du pH 19
I.3.3. Interaction hydrophobe/adsorption 19
I.3.3.1. Hydrophobicité de la membrane 20
I.3.3.2. Hydrophobicité des solutés 21
I.4. Applications des procédés d'OI et de NF 22
I.5. Avantages et inconvénients des procédés d'OI et de NF 22
I.6. Conclusion 23

Chapitre II. Les réseaux de neurones

II.1. Introduction 24
II.2. Méthodes d'apprentissages statistiques 24
II.2.1. Régression linéaire multiple "RLM" 24

II.2.2. Réseaux de neurones artificiels (RNA) 25
II.2.2.1. Neurone biologique 27
II. 2.2.2. Neurone formel 27
II.2.2.3. Architecture d'un réseau de neurones 30
II.2.2.4. Perceptron multicouche 31
II.3. Conception d'un modèle neuronal 33
II.3.1. Détermination des entrées/sorties du réseau de neurones 33
II.3.2. Construction de la base de données et choix des échantillons 33
II.3.3. Définition de la structure du réseau 34
II.3.3.1. Nombre de couches cachées 34
II.3.3.2. Nombre de neurones cachés 35
II.3.4. Apprentissage 35
II.3.5. Évaluation des performances du réseau de neurones 39
II.4. Revue des travaux de la prédiction de la rétention par les modèles QSAR 39
II.5. Conclusion 40

Chapitre III. Modélisation de la rétention des molécules organiques lors l'osmose inverse et nanofiltration

III.1. Introduction 41
III.2. matériels et méthodes 41
III.2.1. Procédure de normalisation 41
III.2.1.1. Collecte des données 42
III.2.1. 2. Normalisation de la base de données 43
III.2.1.3. Analyse statistique 44
III.2.1.4. Développement des modèles QSAR 46
III.2.1.4.1. La régression linéaire multiple 46
III.2.1.4.2. Les réseaux de neurones artificiels 46
III.2.1.5. Outils informatiques utilises 47
III.3.1. Calcule des erreurs 49
III.2.4. Analyse de sensibilité 49
III.3. Résultats et discussion 52
III.3.1. Performance des modèles 52
III.3.2. Expressions mathématique des modèles neuronaux optimisés (modèle RN) 59
III.3.2.1.Analyse de sensibilité 60
III.3.3. Modèle RLM 61
III.3.4. Comparaison entre les modèles RLM et RNA 64
III.3.5. Comparaison entre les algorithmes de retro-propagation 66
III.4. Conclusion 68

Conclusion générale et perspectives 69
Références bibliographiques 71
Annexe A 83
Annexe B 94
Annexe C 97

Abréviations

AC	**A**ngle de **C**ontact
ART	**A**daptive **R**esonance **T**heory
BCF	**F**acteur de **B**io**C**oncentration
BD	**B**ase de **Do**nnées
BFGS	**B**royden-**F**letcher-**G**oldfarb-**S**hanno
Cal	**C**alculée
2D	**D**eux **D**imensions
EAM	**L**'erreur de **P**rédiction du **M**odèle
ESP	**L**'erreur **S**tandard de **P**rédiction
exp	**e**xpérimentale
I	**I**nduction
LM	**L**evenberg-**M**arquardt
max	**m**aximum
MD	**M**oment **D**ipolaire
min	**m**inimum
MF	**M**icro**f**iltration
MSE	**M**ean **S**quared **E**rror **(**l'erreur quadratique moyenne**)**
MO	**M**olécules **O**rganiques
MWCO	**M**olecular **W**eight **C**ut-**O**ff (poids moléculaire de coupure)
NF	**N**ano**f**iltration
OI	**O**smose **I**nverse
PHE	**P**lan de **H**elmholtz **E**xterne

PM	**P**oids **M**oléculaires
PMC	**P**erceptron **M**ulti-**C**ouche
Ptm	**P**ression **t**rans**m**embranaire
PZ	**P**otentiel **Z**êta
QSAR	**Q**uantitative **S**tructure-**A**ctivity **R**elations
QSPR	**Q**uantitative **S**tructure-**P**roperty **R**elationship
Rét	**Rét**ention
RLM	**R**égression Linéaire Multiple
RN	**R**éseaux de **N**eurones
RNA	**R**éseaux de **N**eurones **A**rtificiels
RN1	**R**éseaux de **N**eurones de la base de données simple
RN2	**R**éseaux de **N**eurones de la base de données référence
RN3	**R**éseaux de **N**eurones de la base de données réduite
RNRB	**R**éseau de **N**eurones formé avec l'algorithme de **R**égularisation **B**ayésienne
RNFF	**R**éseau de **N**eurones de type **F**eed **F**orward
REQM	**R**acine de **L**'erreur **Q**uadratique **M**oyenne
RNOEP	**R**éseau de **N**eurones formé avec l'algorithme d'**O**ptimisation par **E**ssaim de **P**articules
Sc	**S**euil de **c**oupure
SR	**R**étention des **S**els
STD	standard deviations (déviation standard **)**
SD	**S**olubilisation-**D**iffusion
3D	**T**rois **D**imensions
TDS	**T**otal **D**issolved **S**olids
UF	**U**ltra**f**iltration

Nomenclatures et symboles

Lecture	Désignation	Unité
a_0	Constante représentant l'effet moyen général	(-)
a_i	Coefficients de régression linéaire	(-)
b_1^0	Biais du neurone de la couche de sortie	(-)
b_j^H	Biais du neurone *j* de la couche cachée	(-)
C_i	Concentration initial du soluté	g/l
C_o	Concentration du composé dans la phase octanol	g/l
C_p	Concentration dans le perméat	g/l
C_w	Concentration du composé syndiqué dans la phase aqueuse	g/l
E_i	$i^{ème}$ entrée du modèle	(-)
Ir	Importance relative	%
$Log\ D$	Coefficients de distribution octanol-eau	(-)
$Log\ k_{0w}$	Coefficients de partage octanol-eau	(-)
n	Nombre d'expérience	(-)
P	Pression	Bar ou Pa
pH	Mesure de l'acidité ou la basicité du bouillon de fermentation	(-)
pHi	*pH* isoélectrique	(-)
Q_f	Débit d'alimentation	m^3/s
Q_p	Débit du perméat	m^3/s
T	Température	°C
tr	Taux de rejet	%
S	Surface	m^2
R	Coefficient de corrélation	(-)
r	Taux de récupération	%
w_i	Poids de connexion	(-)
w_{ji}^I	Poids de connexions entrée –cachée	(-)
w_{1j}^H	Poids de connexions cachée- sortie	(-)
x	Variables d'entrée	(-)

x_i	Position actuelle de la i^{ème} particule	(-)
x_{max}	Position maximale seuil	(-)
x_{min}	Position minimale seuil	(-)
$x_{réf}$	Valeur référentielle normalisée du paramètre x	(-)
$x_{réd}$	Valeur réduite normalisée du paramèter e x	(-)
y	Variables de sortie (rétention)	%
$y_{i,cal}$	valeur prédite de la sortie	%
y_e	moyenne des Données expérimentales de la sortie	%
$y_{i,exp}$	valeur expérimentale de la sortie	%
Z_j	j^{ème} sortie de la couche cachée	(-)

Lettres Grecques

Lecture	Désignation	Unité
$\Delta\pi$	Différence de la pression osmotique de part et d'autre de la membrane	Bar ou Pa
α	Pente de la droite du meilleur ajustement linéaire	(-)
β	Ordonnée à l'origine de la droite du meilleur ajustement linéaire	(-)
Θ	Angle de contact	Degré(^{0})
ψd	Potentiel régnant au niveau du PHE	mv
ζ	Potentiel zêta	mv
χ	constante de Hansch (constant de substitut lipoïdique)	(-)
σ	Constante de Hammett	(-)
ε_i	Erreur du modèle	(-)

Exposants

Lecture	Désignation
H	Cachée (Hidden)
I	Entrée (Input)
O	Sortie (Output)

Indices

Lecture	Désignation
H	Cachée (Hidden)
i	Elément i d'un vecteur
i,exp	Elément i d'un vecteur de valeurs expérimentales
i,cal	Elément i d'un vecteur de valeurs prédites
ij	Croisement de la ligne i avec la colonne j d'une matrice
j	Elément j d'un vecteur
max	Maximum
min	Minimum
O	Sortie (Output)

Liste des tableaux

Chapitre II. Revue critique sur La modélisation de la rétention des MO lors d'OI/NF par le modèles QSAR

Tableau II.1 Fonctions de transfert usuelles, $a=f(n)$ 29

Tableau II.2 Résumé des caractéristiques des algorithmes d'apprentissage de rétro-propagation implantés dans MATLAB® 37

Tableau II.3 Vue d'ensemble des travaux antérieurs sur les modèles de prédiction de la rétention des molécules organiques lors d'osmose inverse et de nanofiltration. 39

Chapitre III. Modélisation de la rétention des molécules organiques lors l'osmose inverse et nanofiltration

Tableau III.1 Analyse statistique des données de la BD simple 45

Tableau III.2 Analyse statistique des données de la BD référence 45

Tableau III.3 Analyse statistique des données de la BD réduite 46

Tableau III.4 Structure des modèles neuronaux optimisés 52

Tableau III.5 Vecteurs de régression linéaire issue de l'équation linéaire : $Rét(cal) = \alpha Rét(exp) + \beta$, [$\alpha$ (pente), β (ordonnée à l'origine), R (coefficient de corrélation)], MSE pour QSAR-RN1, QSAR-RN2 et QSAR-RN3. 56

Tableau III.6 Les erreurs calculées pour la valeur moyenne de la base d'apprentissage, test et totale pour les trois modèles neuronaux (RN1, RN2 et RN3). 57

Tableau III.7 Les erreurs calculées pour la valeur moyenne de la base d'apprentissage, test et totale pour le modèle QSAR-MLR 64

Annexe A

Tableau A.1 Les molécules organiques et leurs formules et structures 83

Annexe C

Tableau C.1 Matrice de corrélations 97

Liste des figures

Chapitre I. Description de la technologie membranaire

Figure I.1	Membrane sélective	5
Figure I.2	Structure d'une membrane composite (type Film Tec)	6
Figure I.3	Représentation schématique des différentes structures membranaires	6
Figure I.4	Module tubulaire	8
Figure I.5	Module à fibres creuses	8
Figure I.6	Les modules plans	9
Figure I.7	Module spirale	10
Figure I.8	Gamme de rétention de différentes particules par les procédés membranaires	11
Figure I.9	Principe des phénomènes d'osmose et d'OI	12
Figure I.10	Principe de phénomène de nanofiltration	12
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)	16
Figure I.12	Angle de contact	20

Chapitre II. Les réseaux de neurones

Figure II.1	Méthodologie du calcul de ε_i	25
Figure II.2	Réseau de neurones artificiel à n variables d'entrée et une sortie	26
Figure II.3	Schéma d'une synapse	27
Figure II.4	Neurone formel	28
Figure II.5	Réseaux de neurones les plus utilisés	30
Figure II.6	Exemple d'architecture d'un *PMC* à n variables d'entrée, une couche cachée avec quatre neurones cachés et une couche de sortie	31
Figure II.7	Algorithme de rétropropagation du gradient d'erreur	37

Chapitre III. Modélisation de la rétention des molécules organiques lors l'osmose inverse et nanofiltration

Figure III.1	Organigramme de la procédure de normalisation	42
Figure III.2	Organigramme pour le développement et l'optimisation d'un RNFF	47
Figure III.3	Représentation des % de division des données par MATLABR2010b	48
Figure III.4	Représentation des % de division des données par STATISTICA.	48
Figure III.5	Organigramme de la méthode de "poids"	51
Figure III.6	Comparaison entre les valeurs expérimentales et calculées par les modèles neuronaux optimisés pour l'ensemble de données de QSAR-RN1 : (a) base d'apprentissage, (b) base de test, (c) base totale	53
Figure III.7	Comparaison entre les valeurs expérimentales et calculées par les modèles neuronaux optimisés pour l'ensemble de données de QSAR-RN2 : (a) base d'apprentissage, (b) base de test, (c) base totale	54
Figure III.8	Comparaison entre les valeurs expérimentales et calculées par les modèles neuronaux optimisés pour l'ensemble de données de QSAR-RN3 : (a) base d'apprentissage, (b) base de test, (c) base totale	55
Figure III.9	Performances des modèles neuronaux (QSAR-RN1, QSAR-RN2, QSAR-RN3)	58
Figure III.10	Représentation schématique du QSAR-RN2 optimisés	59
Figure III.11	Architecture finale du réseau de neurone optimal de type feedforward pour la prédiction de la rétention des molécules organiques (neutres et ioniques) par les membranes d'OI/NF.	60
Figure III.12	Présentation des % de l'importance relative (Ir) de chaque variable d'entrée sur la rétention de QSAR-RN2.	61
Figure III.13	Comparaison entre les valeurs expérimentales et calculées par les modèles neuronaux optimisés pour l'ensemble de données de QSAR-MLR : (a) base d'apprentissage, (b) base de test, (c) base totale	63
Figure III.14	Performances des modèles neuronaux (QSAR-RN2, QSAR-MLR). (a) : la Racine de l'Erreur Quadratique Moyenne (REQM%), (b) : l'Erreur Absolue Moyenne (EAM%) et(c) : l'Erreur Standard de Prédiction (ESP%)	65
Figure III.15	Comparaison entre les valeurs expérimentales et calculées par les modèles neuronaux optimisés pour l'ensemble de données de QSAR-	67

RN2 : (a) base d'apprentissage, (b) base validation, (c) base test, (d) base totale

Annexe B

Figure B.1 Capture d'écran de logiciel Chemspider 94
Figure B.2 Capture d'écran de logiciel ACD/Labs 94
Figure B.3 Capture d'écran de logiciel ChemAxon 95
Figure B.4 Capture d'écran des logiciels STATISTICA et MATLAB 95

Introduction générale

L'eau est un élément vital, précieux et rare. Son rôle économique et social est très important (Metaiche 2006). La quantité d'eau est abondante sur notre planète, elle représente 1380 millions de km^3. Elle est essentiellement constituée d'eau de mer (97.2%) et de glace (2.15%) inutilisables directement. Sur les 0.65% d'eau douce restante, l'eau douce, facilement disponible (lacs, fleuves, certaines eaux souterraines), ne représente que 0.07% de la ressource totale soit environ un million de km^3, mais, la répartition de cette eau est très inégale d'un point de vue géographique. En effet, dix pays se partagent 60% des réserves d'eau douce dont le Brésil qui à lui seul en possède 12% et vingt-huit autres principalement en Afrique et au Moyen-Orient sont au contraire confrontés à une pénurie chronique d'eau douce (Triki 2014). Les ressources naturelles (conventionnelles) en eau douce sont limitées (et épuisées pour certaines régions surtout dans les pays à climat aride et semi-aride). Elles sont inexistantes ou insuffisantes vis-à-vis de la croissance démographique ou de la production industrielle, et le recours vers les ressources non conventionnelles est primordial, dont la réutilisation d'eau et le dessalement de l'eau de mer et de l'eau saumâtre constituent une solution adéquate du problème de l'insuffisance des ressources naturelles. Le recyclage des eaux usées apparaît dés à présent comme une nécessité (Roorda 2004 ; Petrinic et *al*. 2015). La réutilisation des eaux usées est donc « *une action volontaire et planifiée qui permet de fournir une source supplémentaire d'eau pour différents usages comblant les déficits hydriques tout en protégeant le milieu environnant* » (Ecosse 2001). Dans la plupart des pays développés, la majeure partie des eaux usées et des effluents industriels est traitée afin de protéger la qualité du milieu récepteur.

Les procédés de séparation membranaires regroupent un grand nombre de techniques permettant de réaliser des séparations en phase liquide ou en phase gazeuse sous l'action de diverses forces de transfert (mécanique, électrique ou chimique). Dans le cas de la filtration en phase liquide réalisée au moyen de procédés baro-membranaires, c'est-à-dire utilisant un gradient de pression comme force motrice, 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 répandues et de nombreuses études indiquent qu'il est possible de traiter différents types d'effluents en retenant à la fois les molécules inorganiques et organiques. La compréhension fondamentale et quantitative des mécanismes de transfert et des performances des membranes d'osmose inverse et de nanofiltration sont nécessaires à la sélection, au dimensionnement, au fonctionnement et au contrôle des systèmes industriels. D'autre part, le rôle complexe et encore mal cerné de nombreuses interactions qui peuvent s'établir entre les solutés, l'eau et la membrane. De nombreux modèles de transferts existent (thermodynamique irréversible, solubilisation-diffusion ou de pores), 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. Le comportement du procédé d'osmose inverse et de nanofiltration pour les solutions complexe plus difficile à décrire et prédire, du fait du grand nombre de mécanismes mis en jeu. De plus, peu de travaux font le lien entre les mécanismes révélés par l'étude de solutions-modèles et les performances de traitement de rejets industriels. Ils ne se concentrent que sur l'un ou l'autre de ces aspects.

Pour surmonter ces problèmes, des tentatives ont été faites pour proposer des approches de modélisation permettant de diminuer le nombre d'expériences à effectuer pour déterminer les paramètres optimaux pour le traitement d'un effluent industriel. Si elle représente bien le procédé, les influences des paramètres opératoires pourront être évaluées sans aucune expérience et n'auront qu'à être validées sur le site. De plus, la modélisation peut également permettre de valider les hypothèses simplificatrices émises lors de la formulation mathématique des phénomènes de transfert. Pour cela, il faut chercher d'autres méthodes de modélisation telles que les sciences cognitives.

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 (Bouamar et Ladjal 2007). 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, comparé à d'autres approximateurs universels tels que les régressions linéaires ou polynomiales, est qu'ils sont

des approximateurs non-linéaires parcimonieux, permettant une modélisation de procédés non-linéaires quel que soit le nombre de paramètres indépendants utilisés (Dreyfus et *al*. 2002).

Bien que les réseaux de neurones artificiels (RNA) et les régressions linéaires multiples (RLM) 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 et RLM comme moyen de calcul des phénomènes de transports par d'osmose inverse et de nanofiltration lors de la rétention des molécules organiques (Libotean et *al*. 2008 ; Yangali-Quintanilla et *al*. 2008, 2009a et 2010a ; Shahmansouri et al. 2013 ; Sadmani et *al*. 2014a et 2014b).

L'objectif de ce travail est l'application des RN de type feedforward et RLM pour modéliser la rétention des molécules organiques (neutres et ioniques) par les membranes d'OI et de NF. Ainsi, nous avons organisé ce mémoire en trois chapitres :

- Le premier chapitre concerne les procédés de séparation par les membranes d'osmose inverse et de nanofiltration.
- Le deuxième chapitre est dédié à la présentation de la théorie des réseaux de neurones.
- Le troisième chapitre prétend la méthodologie ainsi que les résultats obtenus lors de la modélisation de la rétention des molécules organiques par les membranes d'osmose inverse et de nanofiltration, en utilisant les modèles de réseaux de neurones et régression linéaire multiple.

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
Description de la Technologie membranaire

I.1. Introduction

Parmi les procédés membranaires, l'osmose inverse (OI) et la nanofiltration (NF) sont les procédés qui permettent les plus forts taux de rétentions en sels et en molécules organiques présents en utilisant des membranes. Le rôle de la membrane est d'agir comme une barrière sélective qui permet le passage de certains composés et qui retient certains autres. Ses performances dépendent du type de matériau membranaire et de ses propriétés structurelles et de charge, de la nature de la force motrice (gradient de pression, de concentration, de température, de potentiel électrique, de pression partielle) et de la configuration du module (Jacob 2011).

Cette partie est focalisée sur la définition des membranes et leurs classifications en suite, en donnant leurs propriétés ainsi on parlera sur l'influence des paramètres opératoire sur ces membranes. À la fin on donnant les différents stades d'application des procédés d'OI et de NF avec les avantages et les inconvénients dernier des ces procédés.

I.2. Généralités

I.2.1. Définition de la membrane

Une membrane peut être définie comme une barrière sélective ou semi-perméable de quelques nanomètres à quelques millimètres d'épaisseur séparant deux compartiments (Lanteri 2009). Ce dernier permettant de circuler une solution sous l'effet d'une force motrice (pression, concentration, champ électrique) en arrêtant certaines molécules et en laissant passer d'autres (Figure I.1). La sélection peut s'effectuer par :

- ✓ la taille des pores ;
- ✓ les critères d'affinité entre molécules ou ions et la membrane (Chikhi 2008).

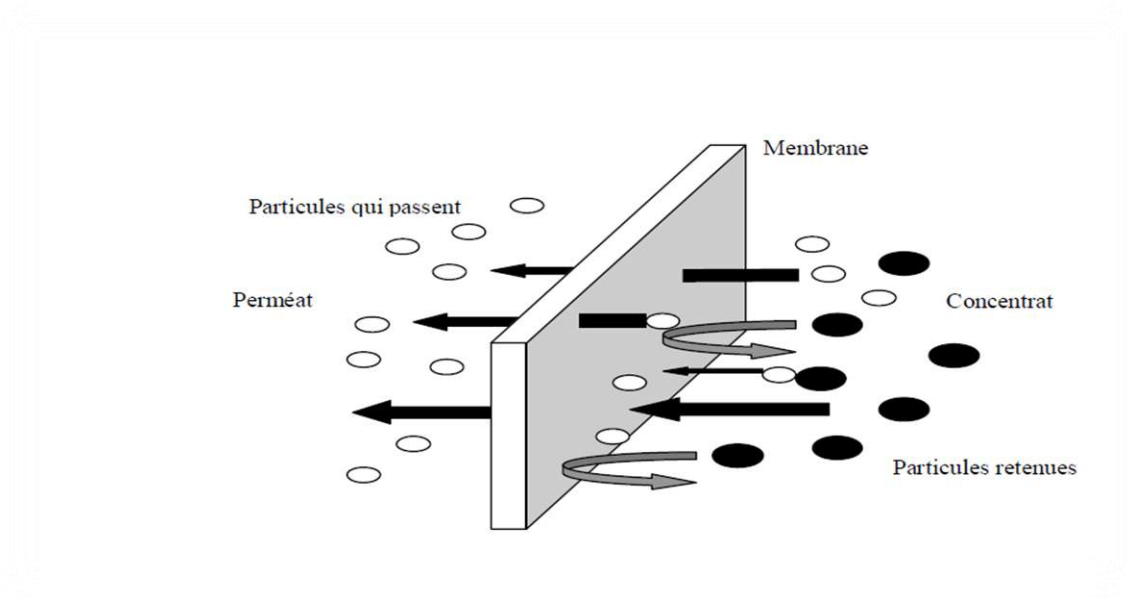


Figure I.1. Membrane sélective (Chikhi 2008).

I.2.2. Classification des membranes

I.2.2.1.Classification d'après la morphologie

Les membranes peuvent être classées en fonction de leur structure (Figure I.3) :

- ✓ Membranes à structure symétrique : ces membranes denses ou poreuses ont la même structure sur toute leur épaisseur (structure isotrope).
- ✓ Membranes à structure asymétrique : la structure de la membrane varie d'une couche à une autre (membranes anisotropes) ; on peut distinguer deux sous-types de membranes asymétriques :
  - Membranes préparées à partir du même matériau.
  - Membranes composites : ce sont des membranes constituées principalement de deux couches :
    - ➢ peau : une couche de très faible épaisseur en contact direct avec l'eau à traiter. Cette couche donne à la membrane sa sélectivité
    - ➢ couche support : une couche plus épaisse, d'une perméabilité beaucoup plus grande, qui supporte la peau et confère à la membrane sa résistance mécanique (voir figure I.2).

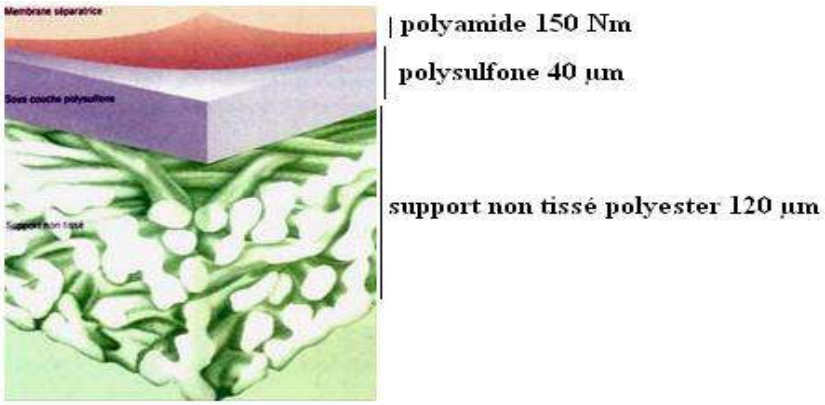


Figure I.2. Structure d'une membrane composite (type Film Tec) (Metaiche 2014).

La plupart des membranes commerciales de MF, UF et de NF sont des membranes à structure asymétrique car ce type de structure permet d'avoir des perméabilités plus élevées (Chikhi 2008).

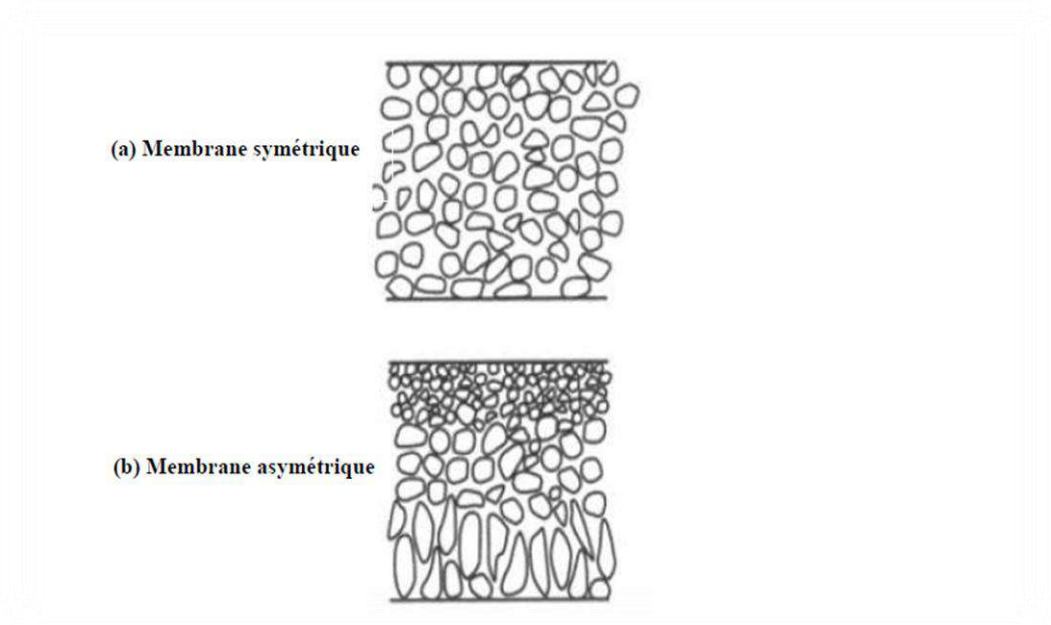


Figure I.3. Représentation schématique des différentes structures membranaires (Bouranene 2008).

I.2.2.2. Classification d'après la nature chimique

a. Membranes organiques

Elles sont fabriquées, pour la plupart d'entre elles, à partir de polymères organiques (acétate de cellulose, polysulfones, polyamides, etc). Les qualités de ces matériaux leur confèrent une grande adaptabilité aux différentes applications (Metaiche 2014).

b. Membranes minérales ou inorganiques

Ces membranes sont composées de corps entièrement minéraux (matières céramiques, métal fritté, verre). Leur arrivée a permis de travailler dans des conditions extrêmes de température et d'agression chimique, ce qui a ouvert de nouvelles voies dans la séparation par membrane.

Les membranes dites dynamiques sont fabriquées par l'utilisateur à partir de tubes poreux dont le diamètre des pores est compris entre 0.5 et 5 µm. Une dispersion colloïdale de zirconium associée à des copolymères est introduite à l'intérieur des tubes. Certaines de ces membranes peuvent trouver des applications en eaux usées (Metaiche 2014).

I.2.2.3. Classification d'après la géométrie de préparation des membranes

Les modules supportent les membranes, 4 grands types de modules sont commercialisés :

- Les modules tubulaires ;
- Les modules fibres creuses ;
- Les modules plans ;
- Les modules spiraux.

a. Module tubulaire

Les membranes se présentent sous forme des tubes regroupés en parallèle dans un module (faisceaux des tubes de 4 à 25mm de diamètre intérieur) (Figure I.4). Le perméat est récupéré dans l'enveloppe du module, à l'extérieur des tubes. Ces systèmes sont aussi très flexibles pour le nettoyage, peu portes au colmatage et fonctionnent généralement en régime d'écoulement turbulent. Leurs principaux défauts sont le coût et l'encombrement. Ce type de module est surtout utilisé pour des membranes minérales (Arzate 2008).

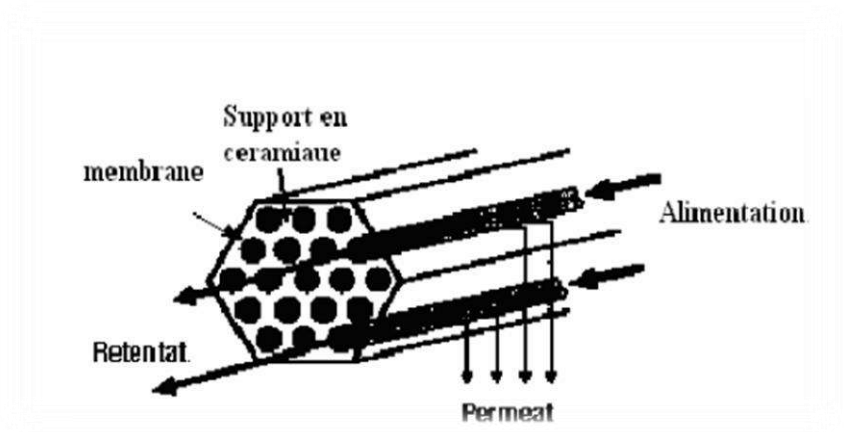


Figure I.4. Module tubulaire (Metaiche 2014).

b. Modules fibres creuses

Ce type de module se présente comme un ensemble des fibres creuses regroupées dans une enveloppe (Figure I.5). La peau est localissée à l'intérieur où à l'extérieur de chaque fibre selon que le perméat est collecté à l'extérieur ou à l'interieur des fibres. Ces systèmes sont très compacts peu coûteux, porticuliérement adaptés pour des fluides peu visqueux présentant des faibles risques de colmatage, une pré-filtration est donc fortement recommandée (Arzate 2008).

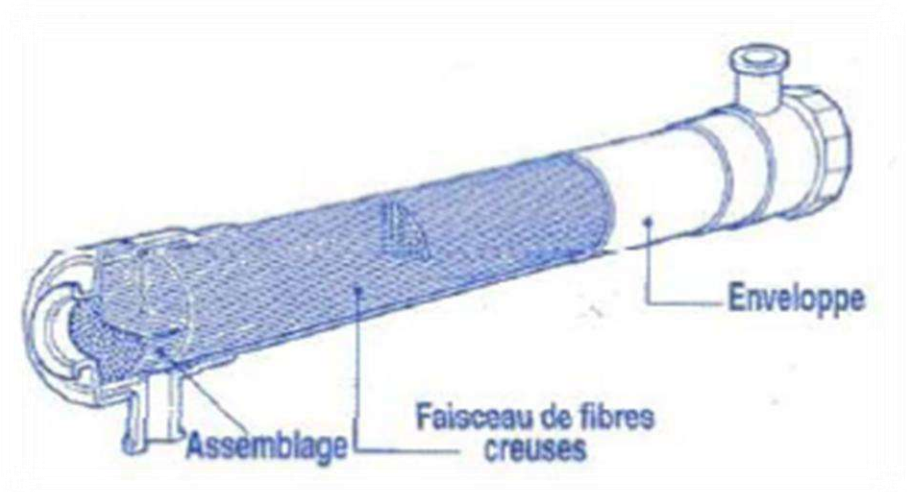


Figure I.5. Module à fibres creuses (Arzate 2008).

c. Les modules plans

Le système utilise des membranes sous forme de feuilles. Il s'agit du type de configuration le plus simple (Figure I.6) La surface filtrante reste toutefois faible par rapport à l'encombrement

(Bouranene 2008). Ces systèmes sont cependant peu adaptés aux débits élevés, se colmatent facilement et nécessitent généralement une pré-filtration (Arzate 2008).

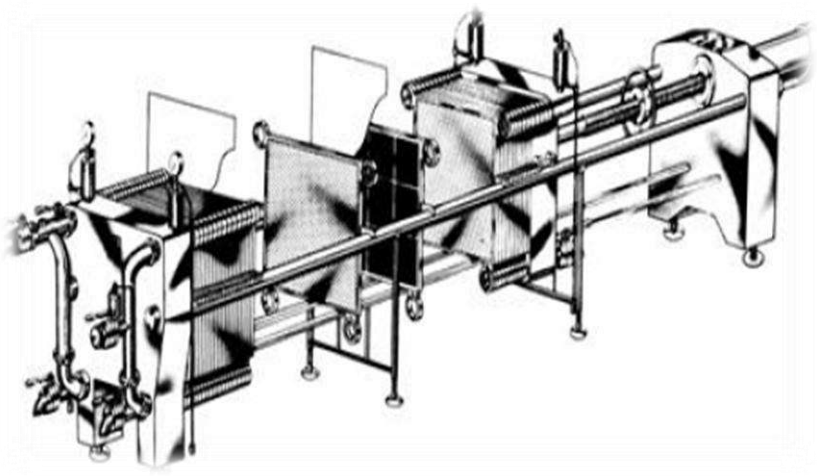


Figure I.6. Les modules plans (Arzate 2008).

d. Les modules spiraux

Ces modules sont les plus utilisés pour la nanofiltration et l'osmose inverse. Ils sont composés d'un tube perforé, sur lequel sont enroulés successivement une membrane, une grille fine en plastique, une membrane et une grille en plastique plus grossière (Voir figure I.7). Le retentât circulé dans la grille grossière alors que le perméat est acheminé par la grille fine jusqu'au tube central. Cet arrangement est très compact et convient pour le traitement à haute pression. Le nettoyage et la désinfection sont cependant peu aisés et le débit d'opération est limité (Arzate 2008).

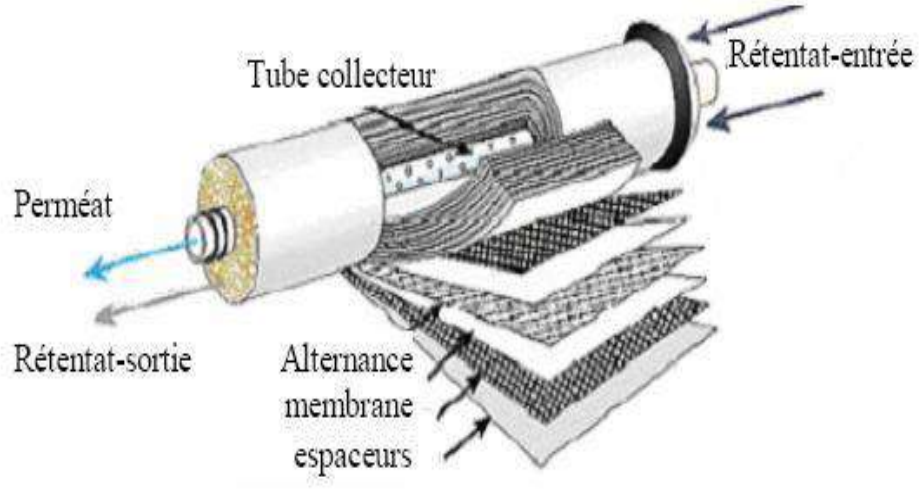


Figure I.7. Module spirale (Danis 2003).

I.2.2.4. Classification d'après le mécanisme de séparation

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

- ✓ **Membranes poreuses** (MF, UF, NF), où les effets de tamisage et de friction jouent des rôles important 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, UF) ;
  - Mésopores – 2 à 50 nm (UF) ;
  - Micropores – plus petits que 2 nm (NF).
- ✓ **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 (Aplet et Buckley 1996).
- ✓ **Membranes échangeuses des ions :** ces membranes sont considérées comme un type spécial de membranes non poreuses, dans des gels denses ayant une charge positive (échangeuses des anions) ou négative (échangeuses des cations) (Chikhi 2008).

La figure I.8 montre schématiquement la correspondance entre les opérations de filtration par membrane, la taille des pores et les composés retenus.

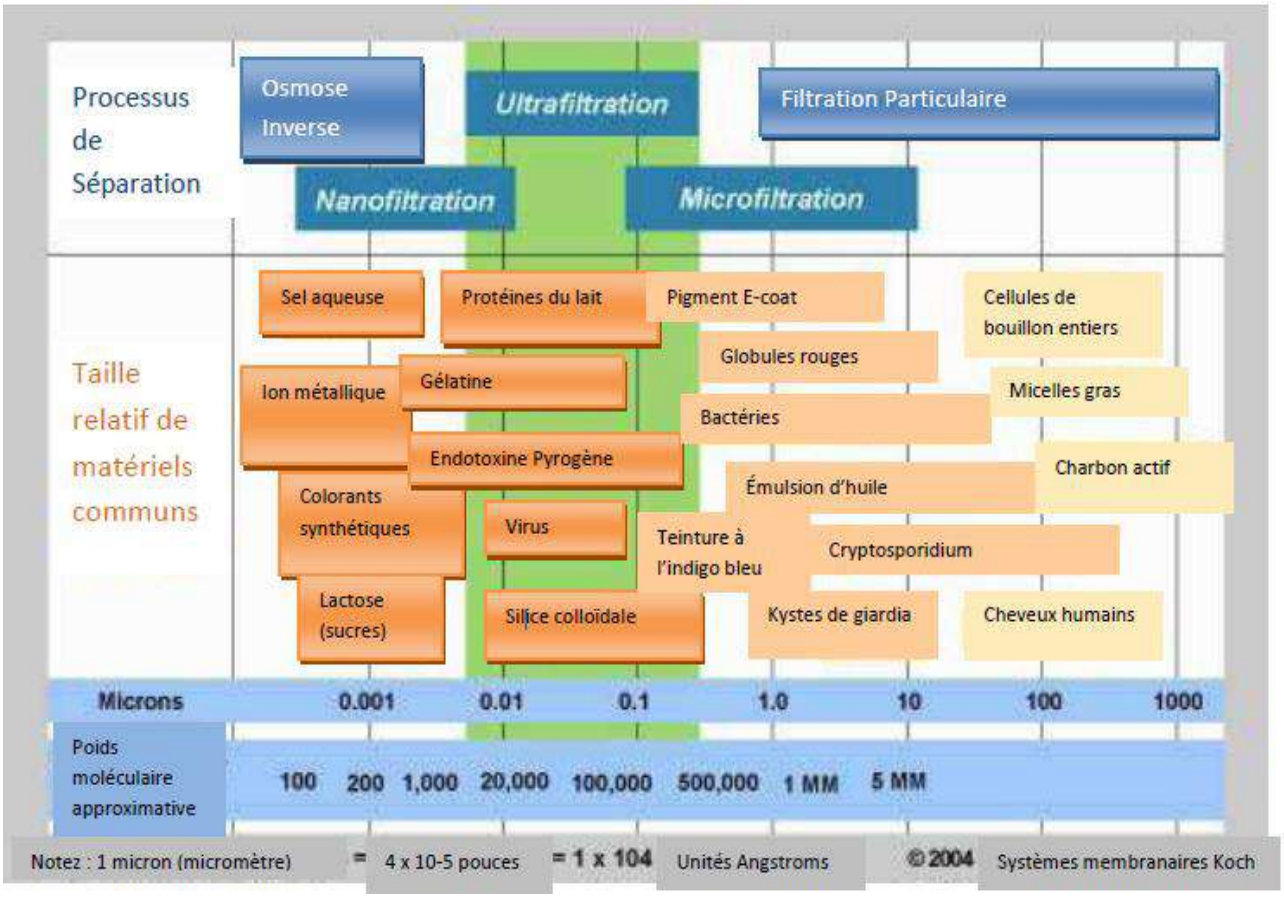


Figure I.8. Gamme de rétention de différentes particules par les procédés membranaires. (http://www.kochmembrane.com/sep_uf.htm)

I.2.3. Osmose inverse et nanofiltration

I.2.3.1. Osmose inverse

L'osmose inverse utilise des membranes denses qui laissent passer le solvant (l'eau) et arrêtent tous les sels.

Le phénomène d'osmose est un phénomène qui tend à équilibrer la concentration en solutés de part et d'autre d'une membrane semi-perméable. Le phénomène d'osmose est un phénomène naturel courant, notamment à travers les membranes cellulaires.

La membrane semi-perméable laissera passer le solvant (le soluté ne passe pas) pour équilibrer la concentration. La différence de concentration crée une pression, appelée Pression osmotique. Pour inverser le passage du solvant et augmenter la différence de concentration, il faut appliquer une pression supérieure à la pression osmotique (voir figure I.9) (Metaiche 2014).

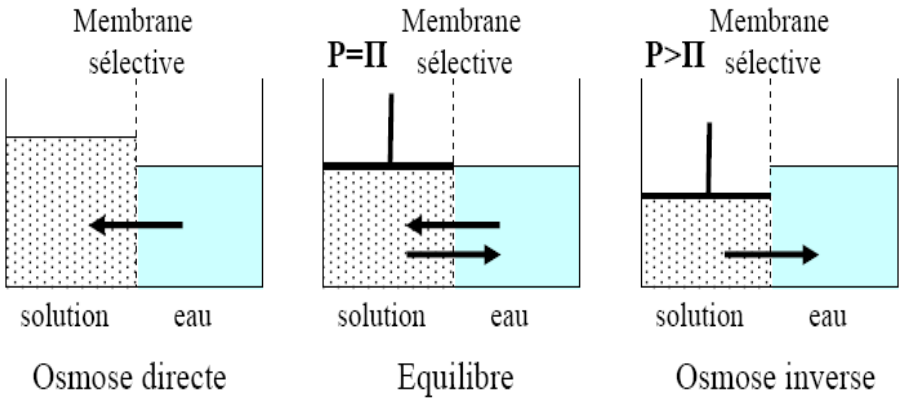


Figure I.9. Principe des phénomènes d'osmose et d'OI (Fell 1995).

I.2.3.2-Nanofiltration

Nanofiltration est le terme utilisé pour désigner une nouvelle technique séparative à membranes se situant entre l'osmose inverse et l'ultrafiltration. Elle permet la séparation de composants ayant une taille en solution voisine de celle du nanomètre (soit 10 Å) d'où son nom.

Les sels ionisés monovalents et les composés organiques non ionisés de masse molaire inférieure à environ 300 g/mol ne sont pas retenus par ce type de membrane. Les sels ionisés multivalents (calcium, magnésium, aluminium, sulfates...) et les composés organiques non ionisés de masse molaire supérieure à environ 300 g/mol sont, par contre, fortement retenus.

Les mécanismes de transfert sont intermédiaires entre ceux de l'osmose inverse et ceux de l'ultrafiltration (Figure I.10) (Metaiche 2014).

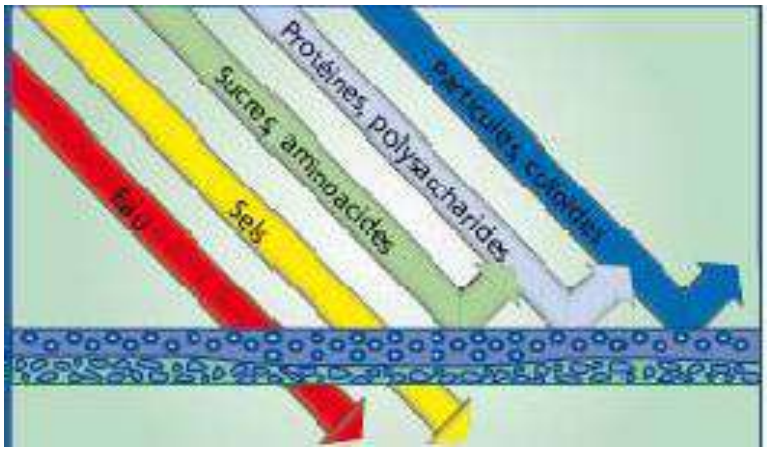


Figure I.10. Principe de phénomène de nanofiltration (Metaiche 2014).

I.2.4. Sélectivité d'une membrane

La sélectivité d'une membrane est définie comme la possibilité de laisser passer ou de retenir les solutés en fonction de leur poids moléculaire. En pratique, ce paramètre est déterminé en filtrant des solutions de produit étalons connus, souvent des protéines ou des polysaccharides (Chikhi 2008).

Le taux de rejet *tr* de la membrane pour le soluté est donné par la relation (Équation I.1) :

$$tr = \frac{C_i - C_p}{C_i} = 1 - \frac{C_p}{C_i} \quad \text{(I.1)}$$

Avec : C_i est la concentration initiale du soluté et C_p est la concentration dans le perméat.

I.2.5. Influence des paramètres opératoires

De nombreux paramètres ont des influences, parfois opposées, sur les performances du procédé d'osmose inverse. L'idéal est d'obtenir une densité de flux et un taux de rétention élevés, mais parfois les conditions de fonctionnement ont des effets antagonistes sur ces paramètres.

I.2.5.1. Pression

La pression de fonctionnement (Ptm) doit être supérieure à la pression osmotique. Ainsi, la pression efficace (Ptm-$\Delta\pi$) correspond à la pression qui conduit réellement à la production de perméat. Plus cette pression est élevée, plus le flux de perméat est important. (Padilla et Tavani 1999 ; Galambos et *al.* 2004 ; Lin et Yang 2004). De même, le taux de rétention augmente lorsque la pression augmente (Sridhar et *al*. 2002 ; Lin et Yang 2004). 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.

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 (Bilstad 1995), d'environ 3% par degré Celsius (Maurel 1998). Le taux de rétention diminue. En effet, la diffusivité de l'eau et des solutés augmente avec la température (Girard et Fukumoto 2000). Afin de prendre en compte la variation de la viscosité de l'eau, même si ce n'est pas le seul paramètre dépendant de la température, il est d'usage de ramener les densités de flux à une température de référence *(x)*

(Allègre et *al.* 2004). Par ailleurs, pour une pression donnée, le compactage de la membrane est d'autant plus important que la température est élevée (Maurel 1998). 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

Le taux de récupération d'eau est un terme de production qui relie le débit du perméat et le débit d'alimentation (Équation I.2). Ce facteur est fonction du temps (Absar 2012).

$$r = \frac{Q_P}{Q_f} \qquad \text{(I. 2)}$$

Avec : Q_P est le débit du perméat [m^3/s] et Q_f est le débit d'alimentation [m^3/s].

I.3. Interactions et propriétés de séparation

En ce qui concerne les molécules organiques, les propriétés de séparation ne sont pas facilement prédictibles. En effet, il a été montré qu'elles résultent d'interactions de diverses natures entre le soluté et la membrane : des effets stériques, des interactions électrostatiques, des liaisons hydrogènes et de Lifshitz-van der Waals (entre dipôles permanents ou induits). Par ailleurs, la solution dans laquelle les espèces sont solubilisées peut modifier ses propriétés d'interactions et celles de la membrane (Sagne 2008).

I.3.1. Encombrement stériques (effet de tamisage)

I.3.1.1. État physique de la membrane

Les propriétés physiques de la membrane comme le nombre de pores, leur 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'osmose inverse 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 defaults 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 former 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).

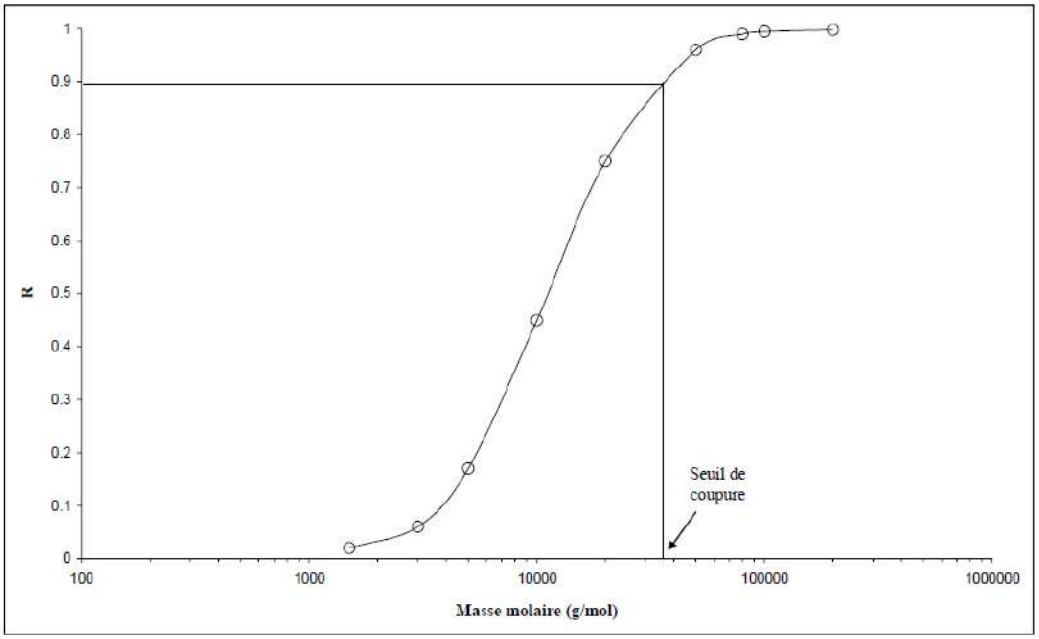


Figure I.11. Illustration du principe de détermination du seuil de coupure d'une membrane, la rétention en fonction de la masse molaire (défini ici pour un taux de rejet de 0.90) (Lanteri et *al.* 2009).

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 (Kiso et *al.* 2000), de polysaccharides et d'alcools (Kiso et *al.* 2001a ; Bellona et *al.* 2004). Étude d'Agenson et *al.* 2003 sur 36 composés organiques semi-volatils et volatils d'intérêt environnemental, de masse

molaire 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.

- **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 (Yangali-Quintanilla et *al*. 2009a et 2010a).

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

Pour les molécules non chargées, peu polaires, la masse molaire peut fournir une première information sur la rétention (Kiso et *al*. 2001b) avec de bonnes corrélations (Ozaki et Li 2002). Cependant, elle 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 et le taux de rétention en NaCl (Bellona et *al*. 2004). En effet, des pesticides ayant une section, et non une masse molaire, plus faible sont moins bien retenus (Berg et *al*. 1997).

D'autres paramètres ont été développés pour évaluer la taille des molécules et la corréler au taux de rétention : la longueur, la largeur, la profondeur et la largeur équivalente qui sont calculés à partir de la structure moléculaire.

- **La largeur moléculaire** (*Width*) est calculé à partir de la projection du plus petit cylindre enfermant la molécule (Van der Bruggen *et al.* 1999).

- **La longueur moléculaire** (*Length*) correspond à la distance entre les deux atomes les plus éloignés, en prenant en compte leur rayon de Van der Waals (Kiso et *al*. 2001a).
- **La profondeur moléculaire** (*Depth*) est calculée à partir de la surface S du plus petit rectangle englobant la projection de la molécule selon l'axe de la longueur ($\sqrt{S}/2$) (Kiso et *al*. 2001a).
- **La largeur équivalente moléculaire** (*Eqwidth*) est égale à la moitié de l'arrête d'un cube englobant la molécule (Kiso et *al*. 2001a).

I.3.2. Répulsion électrostatique (effet de charge)

Les interactions électrostatiques ont souvent été rapportées comme participant de manière importante au mécanisme de rétention. Elles s'établissent entre deux entités chargées : dans notre cas les solutés ionisés et la membrane, qui peut porter une charge (Sagne 2008).

I.3.2.1. Charge des membranes

La charge de la membrane est généralement quantifiée par mesure du potentiel zêta (ζ) (Childress et Desmukh 1998 ; Desmukh et Childress 2001 ; Ozaki et Li 2002 ; Ardhaoui et *al*. 2003 ; Schäfer et *al.* 2004 ; Bellona et Drewes 2005). Il peut être mesuré par potentiel d'écoulement (Childress et Desmukh 1998 ; Desmukh et Childress 2001 ; Ardhaoui et *al*. 2003 ; Schäfer et *al.* 2004), ou mobilité électrophorétique (Ozaki et Li 2002 ; Shim et *al*. 2002 ; Bellona et Drewes 2005).

Lorsque la phase solide et la phase liquide sont mises en mouvement tangentiel l'une par rapport à l'autre, la couche compacte de la double couche électrique reste solidaire de la surface du matériau. Dés lors, les couches compacte et diffuse glissent l'une par rapport à l'autre. La potentielle électrocinétique ou potentiel zêta (ζ) est défini comme le potentiel électrostatique au niveau du plan de cisaillement hydrodynamique entre les deux couches. Il est généralement admis que plan de cisaillement est pratiquement confondu avec le PHE (Hiemenz 1986). Le potentiel zêta (ζ) est donc considéré comme une bonne approximation du potentiel régnant au niveau du PHE (ψd) (Hunter 1996).

Le potentiel zêta peut s'annuler pour un certain *pH* que l'on appelle point isoélectrique (pie). La pie correspond au *pH* pour lequel la charge nette de la membrane (surface physique du matériau + couche compacte) est globalement nulle (LyKlema 1995).

I.3.2.2. Charge des solutés

Les membranes d'osmose inverse 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'électroneutralité implique que les composés chargés positivement le seront également (Shim et *al*. 2002). De plus, les ions sont entourés d'une sphère de solvatation plus importante que les molécules neutres, ce qui accentue leur encombrement stérique (Liew et *al*. 1995) et donc leur rétention. Les molécules ionisables sont donc en général mieux retenues sous leur forme ionisée (Berg et *al*. 1997 ; Ozaki et Li 2002). 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 charge opposée à celle de la membrane (Van der Bruggen et *al*. 1999 ; Hu et *al*. 2003). Pour ces dernières, l'attraction électrostatique favorise leur passage. Pour les plus grosses molécules les effets stériques restent prédominants (Van der Bruggen et *al*. 1999).

I.3.2.3. Influence du *pH*

La charge des membranes dépendent du *pH* de la solution avec laquelle elles sont en contact (Childress et Desmukh 1998, Desmukh et Childress 2001, Schäfer et *al*. 2004, Bellona et Drewes 2005). 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 (Kiso et *al*. 2001b, Agenson et *al*. 2003), 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*. Le *pKa* des couples mis en jeu et le *pH* de travail donnent donc une première indication de leur rétention (Dickson et *al*. 1975). Le *pH* a donc une action concomitante sur la charge de la membrane et des solutés.

I.3.3. Interaction hydrophobe/adsorption

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

I.3.3.1. Hydrophobicité de la membrane

La mesure de l'angle de contact Θ 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 (Ardhaoui et *al.* 2003). 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 (Gumi et *al.* 2003). Par l'utilisation de la méthode de la goutte déposée sur une surface solide (Figure I.12), deux cas peuvent être présentés (Sagne 2008) :

- ✓ Θ est supérieur à 90°, la membrane est dite hydrophobe ;
- ✓ Θ est inférieur à 90°, la membrane est dite hydrophile.

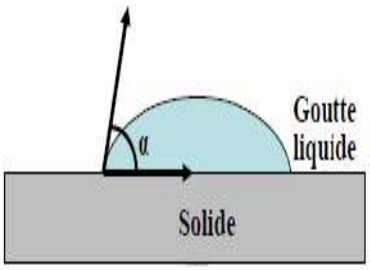


Figure I.12. Angle de contact (Sagne 2008).

La plupart des membranes d'OI sont hydrophiles avec des valeurs d'angle de contact généralement comprises entre 40 et 70° (Norberg et *al.* 2007). Des mesures d'angle de contact avec des solutions de pH différents permettent également d'avoir accès au *pHi*, correspondant au point d'inflexion de la courbe $\Theta = f(pH)$. Les résultats sont proches de ceux obtenus avec les mesures de potentiel zêta (Ardhaoui et *al.* 2003).

Une hystérésis peut être observée entre l'angle de contact de dépôt et de retrait de la goutte. En effet, lors du retrait, la force capillaire des pores augmente la résistance au mouvement de la goutte sur la surface de la membrane. Cette hystérésis est d'autant plus importante que la porosité et la rugosité sont élevées (Gumi et *al.* 2003), sa mesure donne donc une information directe sur la structure de la membrane. La variation de l'angle de contact au cours de la durée d'utilisation de la membrane peut également permettre d'évaluer l'adsorption (Van der Bruggen et *al.* 2002b) et/ou le colmatage (Ardhaoui et *al.* 2003).

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 coefficient de partage octanol-eau K_{ow}, plus souvent exprimé sous forme logarithmique *log* K_{ow} (Équation I.3). S'il est supérieur à 0, la molécule est hydrophobe; s'il est inférieur, la molécule est hydrophile. On parle également de molécules polaires/apolaires. Les molécules hydrophobes sont en général apolaires (Azizi 2011)

$$logK_{ow} = log\frac{C_o}{C_w} \quad \text{(I.3)}$$

Avec: C_o est la concentration du composé dans la phase octanol et C_w est la concentration du composé syndiqué dans la phase aqueuse.

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.4 et I.5) (Verliefde et *al*. 2008).

Pour les acides, (Verliefde et *al*. 2008) **:**

$$LogD_{(pH)}=logk_{ow}-log(1+10^{(pH-pKa)}) \quad \text{(I.4)}$$

Pour les basses, (Verliefde et *al*. 2008) **:**

$$LogD_{(pH)} = logk_{ow} - log(1 + 10^{(pKa-pH)}) \quad \text{(I.5)}$$

La répartition de charge peut également être exprimée par le moment dipolaire de la molécule. Celui-ci est la résultante du moment dipolaire permanent et du moment dipolaire induit. Ce dernier est proportionnel au champ électrique dans lequel il se trouve, par une constante de proportionnalité appelée polarisabilité (Van der Bruggen et *al*. 1999).

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

I.4.1. Applications des procédés d'OI

Cette technique est utilisée pour :

- ✓ Le dessalement des eaux de mer ;
- ✓ Le dessalement des eaux saumâtres ;
- ✓ La production d'eau ultra pure ;
- ✓ La concentration de solutions (concentration de jus de fruits par exemple) ;
- ✓ La production d'eau de process (Metaiche 2014).

I.4.1. Applications des procédés de NF

Les applications possibles sont nombreuses :

- ✓ La déminéralisation sélective (adoucissement des eaux) ;
- ✓ La concentration de composés organiques de faible masse molaire (antibiotiques) ;
- ✓ L'adoucissement des eaux (Metaiche 2014).

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

I.5.1. Avantages

- ✓ Un taux de conversion élevé (supérieur à 55%) ;
- ✓ Une consommation spécifique d'énergie basse (3 à 5 Kwh/m3) ;
- ✓ 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 agro-alimentaire ;
- ✓ Pas d'intervention de réactifs chimiques comme des agents d'extraction qui sont des sources de pollution (Christophe 2008, Aimar 2008, Aptel 2008, Audinos 2008).

I.5.1. Inconvénients

- ✓ 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 colmatage des
- ✓ Membranes (Christophe 2008, Aimar 2008, Aptel 2008, Audinos 2008).

I.6. Conclusion

À partir de ce chapitre on résume que la rétention des molécules organiques par des membranes d'osmose inverse et de nanofiltration dépend d'un grand nombre de facteurs liés à la fois à la membrane (structure physique, charge, hydrophobicité) et 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 stérique, électrostatique et dues au caractère hydrophobe/hydrophile.

Chapitre II
Les réseaux de neurones

II.1. Introduction

La modélisation de La rétention des molécules organiques (MO) par les membranes d'OI et NF gouvernée par leurs propriétés physicochimiques (poids moléculaire, solubilité, diffusivité, polarité, hydrophobicité, charge), les propriétés de la membrane (la perméabilité, la taille des pores, la rugosité, l'hydrophobicité, la charge), les conditions opératoire de procédés (le débit, la pression osmotique, la température, le pH et la composition de l'eau d'alimentation, qui a été des sujets de recherche pendant l'apparition de ces types de membranes (Yangali-Quintanilla 2010, Young 2009).

Pour donner une signification de problème rencontré lors de la séparation des MO par les procédés d'OI/NF, ils ont développé des modèles pour étudier la performance de ces procédés avec l'utilisation des analyses et des expériences qui ont été réalisés. Parmi ces modèles nous rencontrons le RNA et RLM. Mais il y un nombre réduit de travaux cités dans la revue de la littérature relative aux travaux sur la modélisation par les RNA et RLM dans la rétention des MO par les membranes d'OI /NF durant les années (2008-2015) (Azizi 2011).

Dans ce chapitre nous présenterons d'une manière générale les réseaux de neurones, l'architecture d'un réseau de neurones, les types d'apprentissages et les étapes les plus importantes de la conception d'un modèle neuronal. Nous terminerons par une revue des travaux de prédiction de la rétention des MO lors l'OI/NF.

II.2. Méthodes d'apprentissages statistiques

Au cours de la dernière décennie, les algorithmes d'apprentissage statistique ont suscité beaucoup d'intérêt dans le milieu académique et au sein d'entreprises de diverses industries. Ils ont été implantés avec succès pour l'accomplissement des tâches prédictives reliées à des processus statistiques observés pour lesquels on peut identifier plusieurs variables explicatives (Hamadache 2014). Cette étude se concentre sur une classe particulière de ces algorithmes : la régression linéaire multiple et les réseaux de neurones artificiels.

II.2.1. Régression linéaire multiple (RLM)

La régression linéaire multiple est la plus simple des méthodes de modélisation. Son but est de rechercher une relation linéaire mettant en jeu la variable y à modéliser et variables x_i Le nombre d'observations étant "n". L'équation linéaire de régression recherchée s'écrit :

$$y_i = a_0 + a_1 x_{i,j} + \cdots + \varepsilon_i \qquad \text{(II.1)}$$

Où y_i est la $i^{ème}$ observation de la variable y ; $x_{i,j}$ est la $i^{ème}$ observation de la $j^{ème}$ variable ; a_i représentent les coefficients du modèle et ε_i est l'erreur du modèle. ε_i est défini comme la différence (figure II.1) entre la valeur observée de la grandeur à modéliser pour cet élément $i(y_i)$ et l'estimation du modèle (y_i).

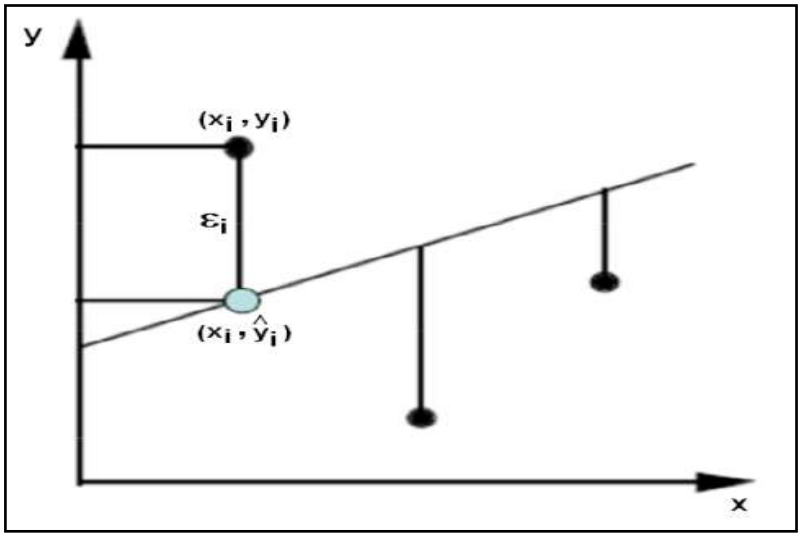


Figure II.1. Méthodologie du calcul de ε_i (Hamadache 2014).

L'apprentissage est réalisé par minimisation de la somme des $\sum_{i=1}^{n}(\varepsilon_i)^2$. Les coefficients du modèle mesurent l'influence de chacune des variables sur la grandeur étudiée (Hamadache 2014).

II.2.2. Réseaux de neurones artificiels (RNA)

L'objectif de cette partie est de donner un aperçu sur les réseaux de neurones artificiels (RNA). En effet, ces réseaux nécessitent la sélection des variables d'entrée xi ayant une influence sur la donnée "y" Tel qu'il est illustré sur la figure II.2. Ensuite, il s'agit de développer une fonction algébrique f, reliant l'évolution de la sortie à ces variables. Ceci sera développé à l'aide d'une modélisation, par régression non linéaire vis-à-vis des paramètres de la fonction estimée, utilisant les réseaux de neurones artificiels (RNA).

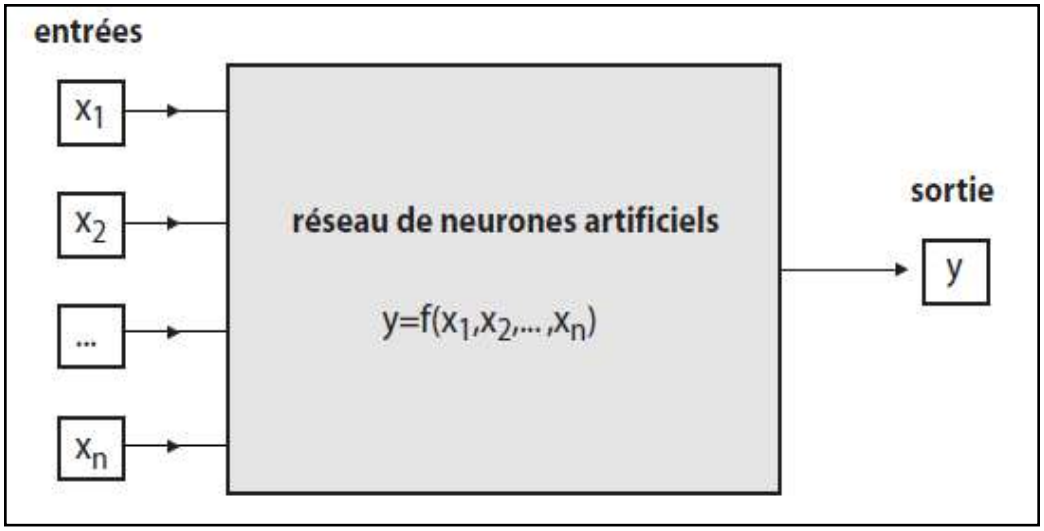


Figure II.2. Réseau de neurones artificiel à n variables d'entrée et une sortie (Hamadache 2014).

Les réseaux de neurones tirent leur puissance de modélisation de leur capacité à capter les dépendances complexes qui impliquent plusieurs variables simultanément. Les RNA présentent en effet une propriété à l'origine de leur succès : ce sont des approximateurs parcimonieux. En effet, Hornik et *al*. (1994) ont montré qu'à précision égale, les RNA nécessitent moins de paramètres ajustables (les poids des connexions) que les approximateurs universels couramment utilisés, comme par exemple les polynômes (Dreyfus et *al*. 2002). Plus précisément, le nombre de coefficients de régression nécessaires à l'obtention d'une précision donnée varie linéairement avec le nombre de variables dans le cas des RNA, alors qu'il varie exponentiellement dans le cas des régressions polynomiales. La parcimonie se révèle donc une propriété d'autant plus importante que le nombre de variables du modèle est grand. En pratique, dès qu'un phénomène à modéliser fait intervenir plus de deux variables, les RNA sont en général préférables aux autres méthodes de régression. C'est le cas du phénomène de la rétention, connus pour être complexes. Sa modélisation impose de considérer un grand nombre de variables.

D'une manière générale, les RNA tels que nous les utiliserons ici, permettent d'obtenir une approximation d'une fonction de régression algébrique reliant les variables d'entrée et la sortie. Une base de connaissances constituée de couples entrées-sortie (s) connue(s) est utilisée de manière à entraîner le RNA, c'est-à- dire à déterminer les meilleurs coefficients de régression nécessaires à l'estimation de la fonction régressive et à sa généralisation sur des données en aveugle (Hamadache 2014).

II.2.2.1. Le neurone biologique

Le cerveau humain est composé de cellules distinctes appelées neurones formant un ensemble dense d'environ 10 à 100 milliards d'unités interconnectées. Tel qu'il est illustré sur la figure II.3, le neurone est une cellule composée d'un corps cellulaire et d'un noyau. Le corps cellulaire se ramifie pour former ce que l'on nomme les dendrites. C'est par les dendrites que l'information est acheminée de l'extérieur vers le soma (corps du neurone) (Khaouane 2013).

L'information est traitée alors par le corps cellulaire. Si le potentiel d'action dépasse un certain seuil, le corps cellulaire répond par un stimulus. Le signal transmis par le neurone est Cheminé ensuite le long de l'axone (unique) pour être transmis aux autres neurones. La transmission entre deux neurones n'est pas directe. En fait, il existe un espace intercellulaire de quelques dizaines d'Angströms entre l'axone du neurone afférent et les dendrites du neurone efférent. La jonction entre deux neurones, comme schématisé sur la figure, s'appelle synapse (Tzanakou 2000).

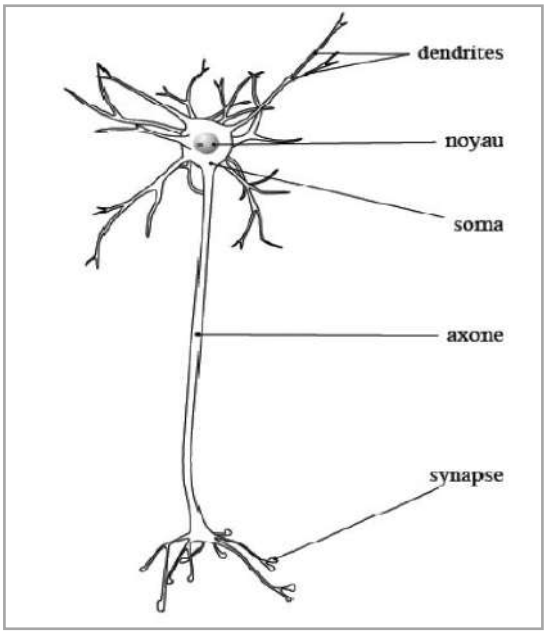


Figure II.3. Schéma d'une synapse (Talbi 2011).

II.2.2.2. neurone formel

Par analogie avec le neurone biologique, chaque neurone artificiel est un processeur élémentaire. Il reçoit un nombre variable d'entrées en provenance de neurones en amont. À chacune de ces entrées est associé un poids *w*, abréviation du mot poids (weight en anglais), représentatif de la force de la connexion. Chaque processeur élémentaire est doté d'une sortie unique qui se ramifie

ensuite pour alimenter un nombre variable de neurones en aval tel qu'il est illustré sur la figure II.4 (Haykin 1999). À chaque connexion est associé un poids.

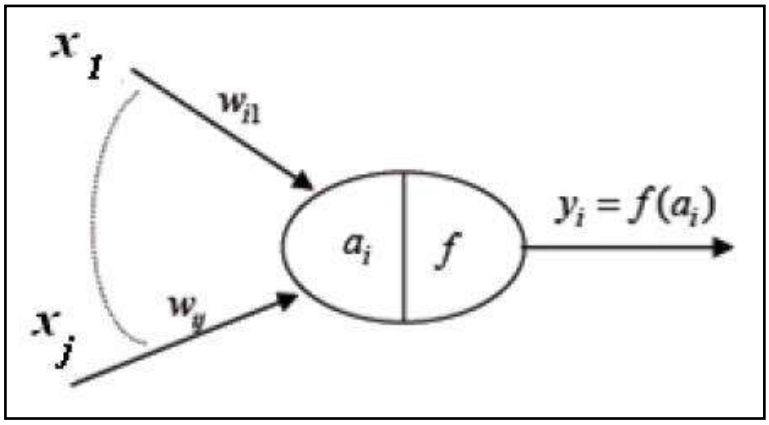


Figure II.4. Neurone formel (Talbi 2011).

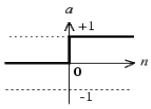
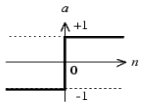
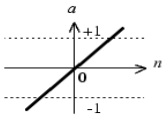
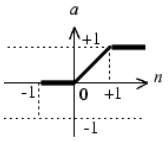
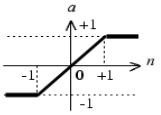
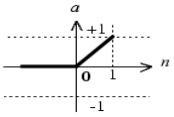
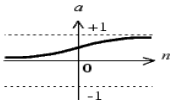
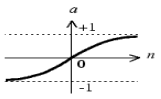
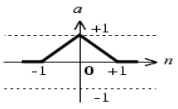
Le comportement du neurone artificiel se compose de deux phases (Haykin 1999) :

1. La première phase représente le prétraitement des données reçues en calculant la somme pondérée des entrées selon l'expression suivante :

$$a_i = \sum_i w_{ij}\, x_j \qquad \text{(II.5)}$$

2. Dans la deuxième phase, une fonction de transfert calcule la valeur de l'état interne du neurone. C'est cette valeur qui sera transmise aux neurones en aval. Plusieurs possibilités existent pour spécifier la nature de la fonction d'activation du modèle. Les fonctions mathématiques les plus souvent utilisées 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]. Ainsi, le neurone peut être défini mathématiquement comme étant une fonction algébrique, a priori non linéaire (suivant la fonction d'activation *f*) et bornée, avec des entrées *x* et paramétrée par les poids w_{ij} .

Tableau II.1. Fonctions de transfert usuelles, *a=f(n)* (Si-Moussa 2009).

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

II.2.2.3. Architecture d'un réseau de neurones

On distingue deux grandes familles de réseaux de neurones : les réseaux bouclés et les réseaux non bouclés (Figure II.5).

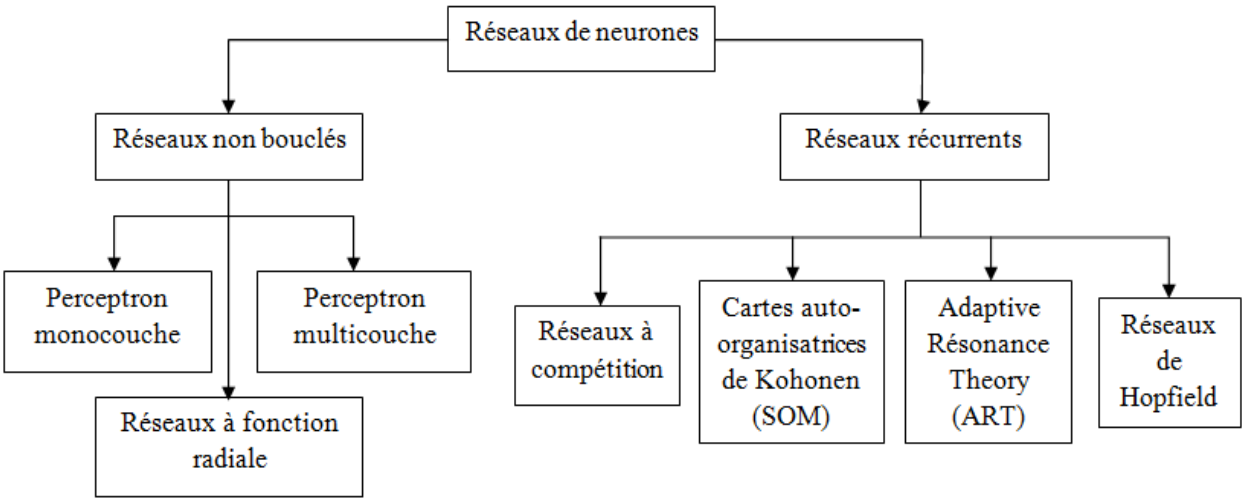


Figure II.5. Réseaux de neurones les plus utilisés (Talbi 2011).

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

- ✓ **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.
- ✓ **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 (Dreyfus 1998).

Dans notre étude, nous avons adopté le modèle le plus couramment utilisé, le Perceptron Multi- couches (PMC), illustré par la figure II.6 et dont les principes sont exposés dans ce qui suit.

II.2.2.4. Perceptron multicouche

Le PMC est un réseau de neurones non bouclé caractérisé par une structure bien particulière La figure II.6 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, dont la fonction d'activation est de la famille des courbes en "S" bornées. Ces neurones n'ont aucun lien avec l'extérieur et sont appelés neurones cachés (Rivals 1995).

✓ couche de sortie : elle est constituée par la valeur de sortie de la combinaison linéaire des fonctions sigmoïdes 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 (Parizeau 2004).

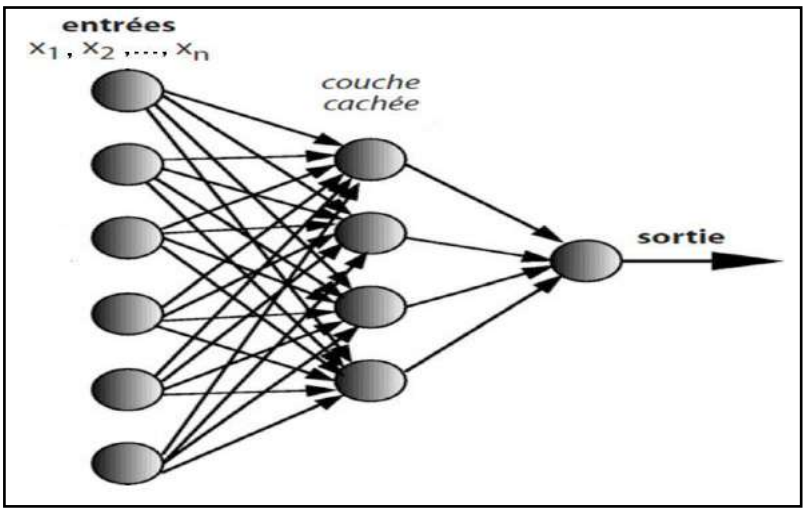


Figure II.6. Exemple d'architecture d'un PMC à n variables d'entrée, une couche cachée avec quatre neurones cachés et une couche de sortie (Hamadache 2014).

II.2.2.5. 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 (Hassoun 1995).

II.2.2.5.1. 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 son 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 (Hassoun 1995).

II.2.2.5.2. 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-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 (Hassoun 1995).

II.3. 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.3.1. 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.3.2. 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 et *al*. 1977). 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'entrée et de sortie sont normalisées dans un intervalle spécifique afin de donner à chaque paramètre la même influence statistique. Nous détaillerons l'ensemble des données utilisées dans cette étude et la méthodologie employée concernant ce découpage entre base d'apprentissage et base de validation.

II.3.3. 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 (Khaouane 2013).

II.3.3.1. 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 multicouches avec une seule couche cachée pourvue d'un nombre suffisant de neurones, peut approximer n'importe quelle fonction avec la précision souhaitée (Hornik 1991).

II.3.3.2. 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.3.4. 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.

La figure II.7 représente un exemple d'Algorithme de rétropropagation du gradient d'erreur

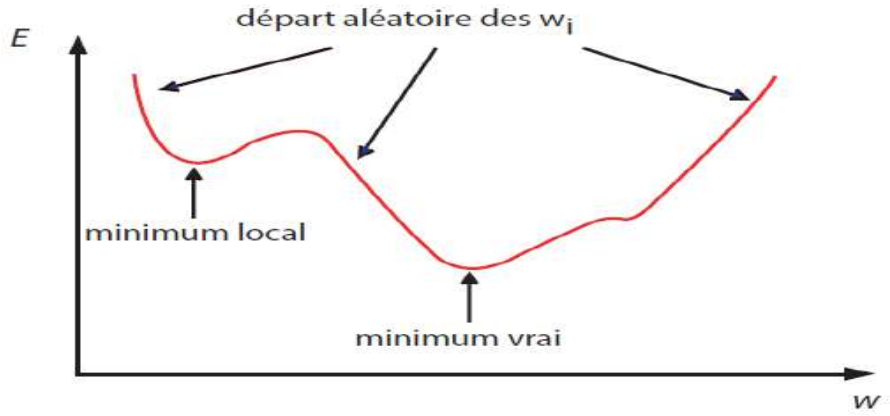


Figure II.7. Algorithme de rétropropagation du gradient d'erreur (Hamadache 2014).

Tableau II.2. Résumé des caractéristiques des algorithmes d'apprentissage de rétro-propagation implantés dans MATLAB® (Demuth et Beale 2000).

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 d'apprentissage variable (***traingda)***	Taux d'apprentissage variable. Plus rapide que ***traingd***, mais ne peut être utilisée que pour apprentissage en mode batch. Utilise des règles heuristiques pour la variation du taux d'apprentissage
Descente du gradient avec moment et taux d'apprentissage variable (***traingdx)***	Descente du gradient avec inertie (momentum) et taux d'apprentissage variable
Descente des gradients conjugués	
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.
Scaled conjugate gradient ***(trainscg)***	Le seul algorithme de gradients conjugués qui n'exige aucune recherche de ligne. Très bon algorithme d'apprentissage en général.
Quasi-Newton	
BFGS quasi-Newton ***(trainbfg)***	Il exige le stockage d'une approximation de la matrice Hessiènne et nécessite plus de calcul à chaque itération que les algorithmes de gradients conjugués, mais possède habituellement une convergence plus rapide.
Sécante à un pas ***(trainoss)***	Cet algorithme exige moins de stockage et calcul par époque que l'algorithme BFGS. On peut considérer cela un compromis entre des algorithmes de quasi-Newton complets et des algorithmes de gradients conjugués.
Levenberg-Marquardt	
Levenberg-Marquardt ***(trainlm)***	C'est l'algorithme d'apprentissage le plus rapide pour les réseaux de taille modérée. Il a la caractéristique de réduction de mémoire pour l'utilisation quand l'échantillon d'apprentissage est grand.
Régularisation	
Régularisation Bayésienne ***(trainbr)***	Mise à jour des poids et des biais selon l'optimisation de Levenberg-Marquardt. Il réduit au minimum une combinaison d'erreurs quadratiques et des poids et détermine ensuite la combinaison correcte afin de produire un réseau avec une capacité de généralisation meilleure. Ce processus est appelé la régularisation bayésienne. Il réduit la difficulté de déterminer l'architecture de réseau optimum.

II.3.5. Évaluation 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.4. Revue des travaux de la prédiction de la rétention par les modèles QSAR

Durant les dernières années la littérature scientifiques relative aux RNA s'est enrichie par plusieurs publications de tout sujet d'intérêt en génie chimique. Dans cette revue l'accent est mis sur les applications des RNA pour la modélisation, dans le but de prédiction de La rétention par les modèles QSAR. Le tableau II.3 résumé les travaux relatifs aux modèles de prédiction de la rétention des MO lors l'OI/NF.

Tableau II.3. Vue d'ensemble des travaux antérieurs sur les modèles de prédiction de la rétention des molécules organiques lors d'osmose inverse et de nanofiltration.

Nombre et nature de MO étudiés	Type de membrane	Taille de base de données	Méthode de Modélisation	Références
82 Molécules organiques	NF et OI	956	QSAR-RN	Ammi et *al*. 2015
17 Molécules organiques neutres et ioniques	NF	51	QSAR-RLM	Sadmani et *al*. 2014a
17 Molécules organiques neutres et ioniques	NF	51	QSAR-RLM	Sadmani et *al*. 2014b
67 Molécules organiques neutres	NF	255	QSAR-RN	Shahmansouri et al. 2013
66 Molécules organiques neutres et ioniques	NF	106	QSAR-RLM	Yangali-Quintanilla et *al*. 2010a
50 Molécules organiques neutres	OI et NF	161	QSAR-RN	Yangali-Quintanilla et *al*. 2009a
14 Molécules organiques neutres et ioniques	OI et NF	64	QSAR-RLM	Yangali-Quintanilla et *al*. 2008
50 Molécules organiques neutres	OI	144	QSAR-RN	Libotean et *al*. 2008

II.5. Conclusion

Ce chapitre introduit les notions de base sur les réseaux de neurones artificiels (RN). Après Une présentation du modèle biologique et mathématique du neurone, les différentes architectures des RN ainsi que leurs types d'apprentissage sont présentés. Dans un deuxième temps, nous avons exposé le type du RN utilisés dans notre étude ainsi que les différentes étapes de conception d'un réseau de neurones. Nous avons terminé ce chapitre par une revue des travaux de la modélisation de la rétention des MO lors d'OI/NF.

Chapitre III
Modélisation de la rétention des molécules organiques lors l'osmose inverse et nanofiltration

III.1. Introduction

Ce chapitre traite la modélisation de la rétention des MO lors l'OI/NF. Pour ce faire, on commencera par une première partie intitulée "Matériels et méthodes ". Cette dernière comportera les diverses étapes de l'élaboration de la base de données nécessaire au fonctionnement des méthodes d'apprentissage statistique qu'on utilisera, les techniques les plus courantes pour établir des modèles QSAR : la régression linéaire multiple (RLM ou MLR en anglais) et les réseaux de neurones artificiels (RNA), ainsi que les étapes permettant l'élaboration de deux modèles linéaire (par RLM) et non linéaire (par RNA) de la rétention des MO lors d'osmose inverse et nanonfiltration.

La seconde partie de ce chapitre se veut une application à l'élaboration de modèles QSAR pour la modélisation de la rétention des MO lors l'OI /NF. On y discutera des résultats obtenus et on présente les architectures optimales des modèles neurones. Ensuite après la présentation des résultats de comparaison des modèles en évaluant la rétention et en calculant les erreurs, nous exposant les résultats de l'analyse de sensibilité. Cette partie se termine par une comparaison entre les modèles neuronaux et les différents modèles QSAR.

III.2. Matériels et méthodes

III.2.1. Procédure de normalisation

Dans ce chapitre, une procédure basée sur le développement et l'optimisation de l'architecture d'un réseau de type Feedforward (RNFF) est décrite. Et que nous pouvons représenter le contenu de cette idée, comme décrit sur la figure III.1.

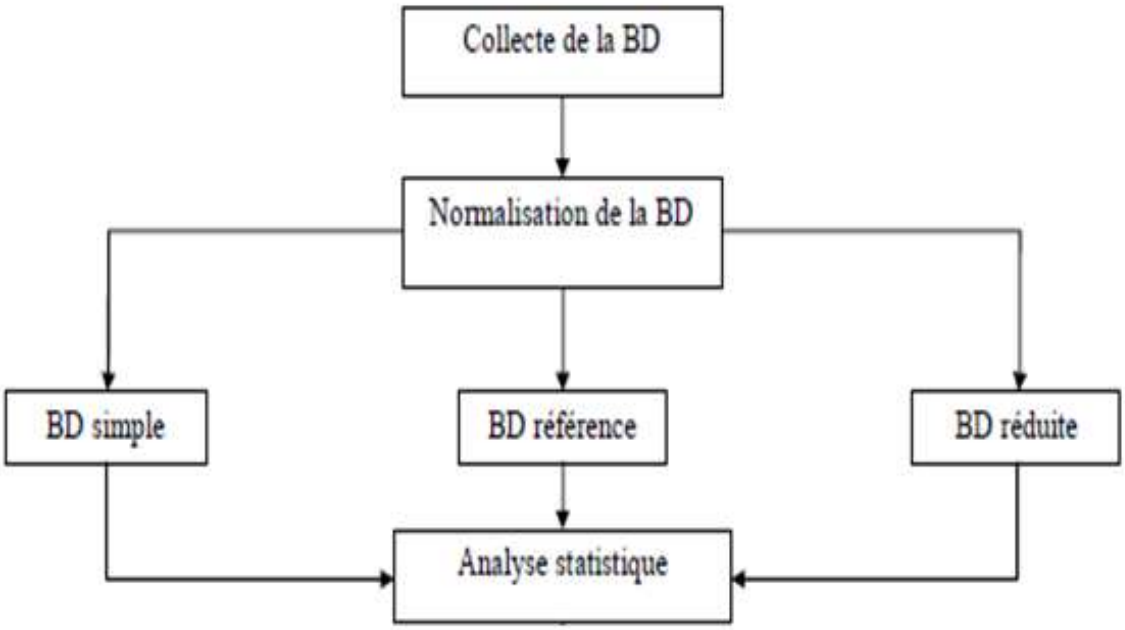


Figure III.1. Organigramme de la procédure de normalisation.

III.2.1.1. Collecte des données

Dans cette étude nous avons utilisé les données disponibles dans la littérature, [Verliefde (2008), N´uria Fontanals (2015), Yangali-Quintanillaa et *al.* (2009), Yangali (2010)]. On a utilisée le logiciel Chemspider (figure. B. 1(Annexe B)) pour calculer le poids moléculaire(PM), ainsi que la formule chimique de chaque molécule. Pour le logarithme du coefficient de partage octanol-eau ($log\ k_{ow}$), et le logarithme du coefficient de distribution (*logD*) on utilise le logiciel ACD/labs (Figure B.2 (Annexe B)), on a calculée le largueur moléculaire (*Width*), longueur moléculaire (*Length*), profondeur moléculaire (*Depth*) et la largeur équivalente moléculaire (*Eqwidth*) par le logiciel ChemAxon (Figure. B.3 (Annexe B)). Chembiooffice Draw (Figure. B.4 (a) (Annexe B)) pour dessiner la structure 2D de chaque molécule et chembiooffice 3D (Figure. B.4 (b) (Annexe B)) pour calculer le moment dipolaire (*MD*) et obtenir la structure 3D.

Les données expérimentales rassemblées pour notre étude ont été mentionnées selon les indicateurs de porosité qu'on les classe souvent en tenant compte de poids moléculaires de coupure (*MWCO*) et la rétention des sels (*SR*) pour $MgSO_4$ et NaCl. Notre base a une taille de 291 échantillons.

Les entrées de modèle sont : le poids moléculaires (*PM*), le logarithme du coefficient de partage octanol-eau ($log\ k_{ow}$), le logarithme du coefficient de distribution octanol-eau (*logD*), le moment dipolaire (*MD*), le largueur moléculaire (*Width*), longueur moléculaire (*Length*),

profondeur moléculaire (*Depth*) et la largeur équivalente moléculaire (*Eqwidth*), le poids moléculaire de coupure (*MWCO*), la rétention des sels (*SR*) pour $MgSO_4$ et NaCl, le potentiel zêta (*PZ*), l'angle de contact (*AC*), *pH*, la pression (*P*), le taux de récupération (*r)* et la température (*T*), la rétention (*Rét*) représente la sortie de la base de données.

Le choix des paramètres se justifie par l'intensité de la liaison qui existe entre eux. Cette intensité de liaison a été obtenue par le calcul des coefficients de corrélation de Pearson (*R*) :

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

Où z_i est la i^{ème} valeur du vecteur Z (*PM, log* K_{ow}*, logD, MD*, le largueur moléculaire (*Width*), longueur moléculaire (*Length*), profondeur moléculaire (*Depth*) et la largeur équivalente moléculaire (*Eqwidth*) , *MWCO, SR, PZ, AC, pH, P, r, T,* Rétention); y_i est la i^{ème} valeur du vecteur Y (*PM, log* K_{ow}*, logD, MD*, le largueur moléculaire (*Width*), longueur moléculaire (*Length*), profondeur moléculaire (*Depth*) et la largeur équivalente moléculaire (*Eqwidth*) , *MWCO, SR, PZ, AC, pH, P, r, T,* Rétention); $\bar{z}$ et $\bar{y}$ sont respectivement les moyennes arithmétiques des vecteurs Z et Y .

Tableau.C.1(Annexe C) regroupe les coefficients de corrélation entre Z (*PM, log* K_{ow}*, logD, MD*, le largueur moléculaire (*Width*), longueur moléculaire (*Length*), profondeur moléculaire (*Depth)* et la largeur équivalente moléculaire (*Eqwidth*) , *MWCO, SR, PZ, AC, pH, P, r, T,* Rétention), On remarque que le degré de corrélation est élevé entre *log* K_{ow} *et logD* ($\pm 0.25 \leq R \leq \pm 0.75$) par conséquent il existe une relation entre ces deux paramètres qu'on a mentionnée dans les équations (I.4 et I.5), tandis qu'entre le largueur moléculaire (*Width*), longueur moléculaire (*Length*), profondeur moléculaire (*Depth*) et la largeur équivalente moléculaire (*Eqwidth*) est modéré ($\pm 0.75 \leq R < \pm 1$) ; montrant ainsi l'existence de liaison linéaire entre les paramètres précédents .

III.2.1.2. Normalisation de la base de données

Tout d'abord, il a fallu normaliser toutes les variables d'entrée et la variable de sortie. En effet, si des entrées ont des grandeurs très différentes, celles qui sont petites n'auront pas d'influence lors de l'apprentissage. Avec cette normalisation, l'apprentissage sera plus efficace et très rapide et ce en empêchant la fonction de transfert de devenir saturée. Les entrées et les sorties sont normalisées à un intervalle de [-1, 1].

La normalisation des modèles de données permet de vérifier la robustesse de leur conception pour améliorer la modélisation (et donc obtenir une meilleure représentation) et faciliter

la mémorisation des données en évitant la redondance et les problèmes sous-jacents de mise à jour ou de cohérence. La normalisation s'applique à toutes les entités et aux relations porteuses de propriétés. On a effectuée deux méthodes de normalisation pour notre base, la première est appelée la normalisation référentielles qui suit la loi suivant (III.2) :

$$x_{réf} = \frac{x_{exp}}{x_{max}} \quad \text{(III.2)}$$

Avec $x_{réf}$ est la valeur normalisée du paramètre$x$.

x_{exp} , x_{max} sont respectivement les valeurs réelles et maximales de x.

La deuxième méthode est dit réduite, et elle suit la loi qu'il est illustré dans le modèle suivant (III.3):

$$\boldsymbol{x_{réd}} = \frac{\boldsymbol{x_{exp} - x_{min}}}{\boldsymbol{x_{max} - x_{min}}} \quad \text{(III.3)}$$

Avec $x_{réd}$ est la valeur normalisée du paramètre$x$.

x_{exp} , x_{max} , x_{min} sont respectivement les valeurs réelles, maximales et minimales de$x$.

A la fin de cette étape on avait trois bases de données pour faire les calcules et choisir les meilleurs résultats.

III.2.1.3. Analyse statistique

L'analyse statistique des données d'entrée et de sortie des trois base de données fait regrouper les valeurs minimales (min), maximales (max), moyennes (moyenne) (équation III.4) et l'écart type ou bien ce qu'on appelle la déviation standard (STD) (équation III.5) est indiquée dans les tableaux III.1, III.2, III.3 pour la BD simple, BD référence et la BD réduite respectivement.

$$\bar{x} = moyenne = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{(III.4)}$$

$$STD = \left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{1/2} \quad \text{(III.5)}$$

Tableau III.1. Analyse statistique des données de la BD simple.

	BD simple			
	min	**max**	**moyenne**	**STD**
***MW* (g/mol)**	30,026	361,819	204,4517	67,0553
LogD	-7,08	6,13	0,7147	2,1583
dipolar Moment *(*De*)*	0	15,03	3,4219	2,6416
***Length* (nm)**	0,13	2,11	1,1929	0,3574
***Eqwidth* (nm)**	0,17	1,74	0,358	0,2077
Zeta Potential (mv)	-48	20,38	-18,8973	12,8885
Contact angle	26,1	63,2	45,3866	7,7792
pH	2,5	10	7,9416	0,5366
Pressure (Kpa)	270	1306	458,5189	165,6662
Recovery	3	90	29,0447	31,0674
***T* (°C)**	13	25	21,7904	3,4841
SR	0,96	1	0,9768	0,0107
***MWCO (*Da)**	100	300	238,488	64,0631
NaCl	0,3	0,997	0,6713	0,2144
Rétention	0	100	73,5151	27,3387

Tableau III.2. Analyse statistique des données de la BD référence.

	BD référence			
	min	**max**	**moyenne**	**STD**
***MW* (g/mol)**	0,082	1	0,5634	0,1859
LogD	-1,155	1	0,1149	0,351
dipolar Moment *(*De*)*	0	1	0,2266	0,1767
***Length* (nm)**	0,061	1	0,5651	0,1695
***Eqwidth* (nm)**	0,0977	1	0,2058	0,1194
Zeta Potential (mv)	-2,355	1	-0,9277	0,632
Contact angle	0,413	1	0,7181	0,1231
pH	0,25	1	0,7942	0,0537
Pressure (kpa)	0,2067	1	0,3511	0,1269
Recovery	0,0333	1	0,3227	0,3452
***T* (°C)**	0,52	1	0,8716	0,1394
SR	0,96	1	0,9768	0,0107
***MWCO (*Da)**	0,3333	1	0,795	0,2135
NaCl	0,3009	1	0,6733	0,2151
Rétention	0	1	0,7352	0,2734

Tableau III.3. Analyse statistique des données de la BD réduite.

	BD réduite			
	min	**max**	**moyenne**	**STD**
***MW* (g/mol)**	0	1	0,5257	0,2021
LogD	0	1	0,5952	0,158
dipolar Moment *(*De*)*	0	1	0,2277	0,1758
***Length* (nm)**	0	1	0,5368	0,1805
***Eqwidth* (nm)**	0	1	0,6957	0,2323
Zeta Potential (mv)	0	1	0,4256	0,1885
Contact angle	0	1	0,5199	0,2097
pH	0	1	0,7255	0,0715
Pressure (kpa)	0	1	0,182	0,1599
Recovery	0	1	0,2994	0,3571
***T* (°C)**	0	1	0,7325	0,2903
SR	0	1	0,4192	0,2665
***MWCO (*Da)**	0	1	0,6924	0,3203
NaCl	0	1	0,5327	0,3077
Rétention	0	1	0,7352	0,2734

III.2.1.4. Développement des modèles QSAR

Cette étape concerne la création d'une relation statistiquement significative de la propriété à modéliser. Beaucoup de méthodes de développement de modèle sont largement utilisées. Les deux méthodes utilisées pour l'établissement des modèles QSAR sont la régression linéaire multiple (RLM) et les réseaux de neurones artificiels (RNA).

III.2.1.4.1. Régression linéaire multiple

La régression linéaire multiple est un outil statistique dont l'objectif est d'établir une relation linéaire entre une propriété d'intérêt (rétention par exemple) et un ensemble de variables (variables d'entrées dans notre cas) et ce suivant l'équation III.6 suivante :

$$Rétention = a_0 + \sum_{i=1}^{n} a_i x_i \qquad \text{(III.6)}$$

Où x_i représentent les sorties, a_i leur coefficient et a_0 est une constante (appelé intercepte).

III.2.1.4.2. Réseaux de neurones artificiels

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). L'utilisation de la méthode

RNA est nécessaire et est devenue une technique de modélisation importante. Elle est largement utilisée pour les études de QSAR non linéaires et est un outil puissant pour construire des modèles prédictifs. Le type de réseaux de neurones retenu pour ce travail est le perceptron multicouche (PMC). Le perceptron multicouche est un des réseaux de neurones les plus utilisés pour des problèmes d'approximation, de classification et de prédiction. La Figure III.2 est illustrée Organigramme pour le développement et l'optimisation d'un RNFF.

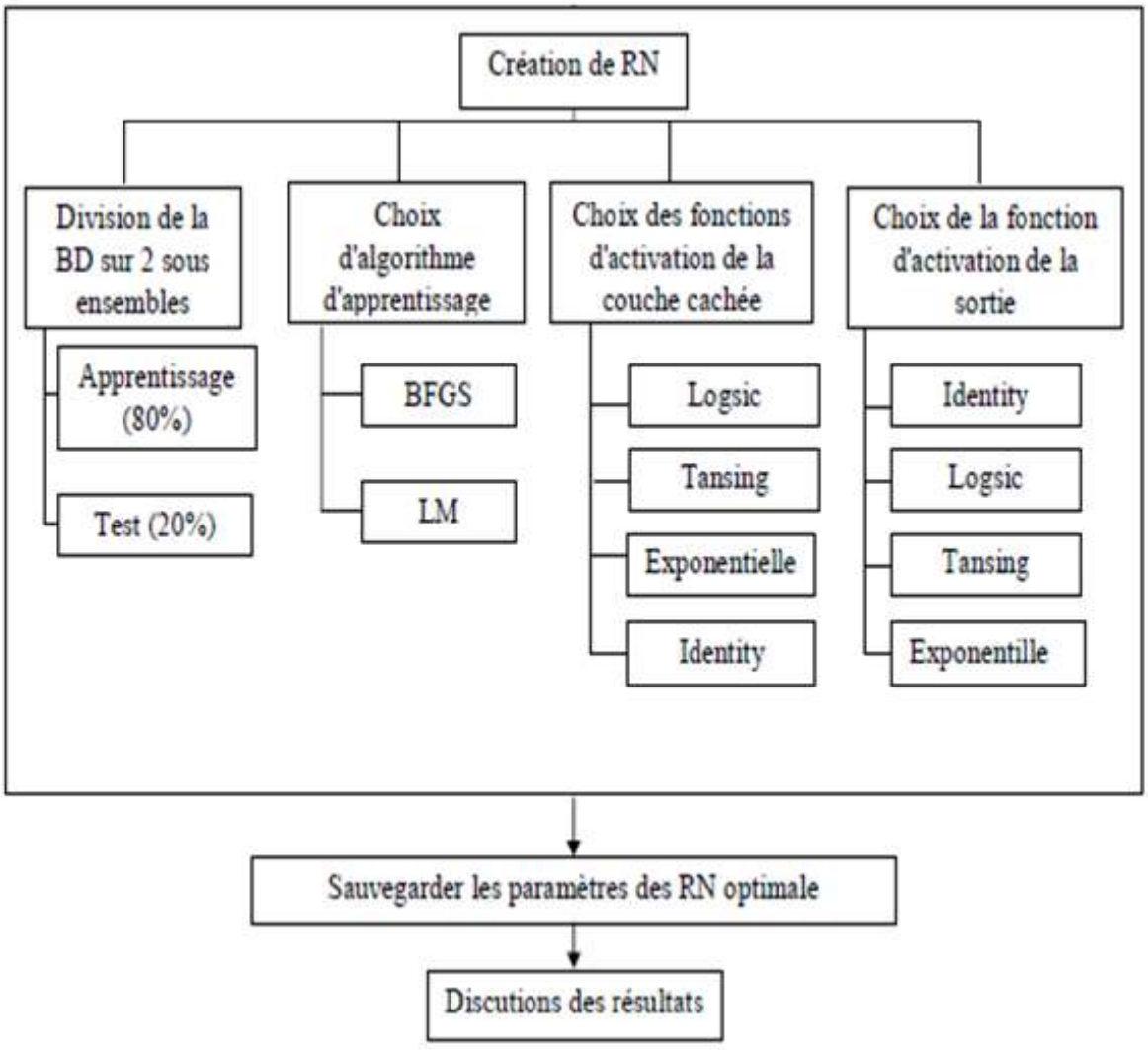


Figure III.2. Organigramme pour le développement et l'optimisation d'un RNFF.

III.2.1.5. Outils informatiques utilisés

Au cours de ce travail, nous avons eu à utiliser plusieurs logiciels (voir capture d'écran sur l'annexe B (figure. B. 5)) :

- STATISTICA (STATISTICA version 7.0);
- MATLAB R2010b.

Le logiciel MATLAB R2010b a été fonctionné par l'algorithme d'apprentissage *nftool* (Levenberg-Marquardt) et a été utilisé pour la conception et l'apprentissage des modèles neuronaux. Chaque base a été divisée en base d'apprentissage (80%), et base de validation (10%), et base de test (10%). La figure III.3 est illustrée la division des données selon MATLAB R2010b.

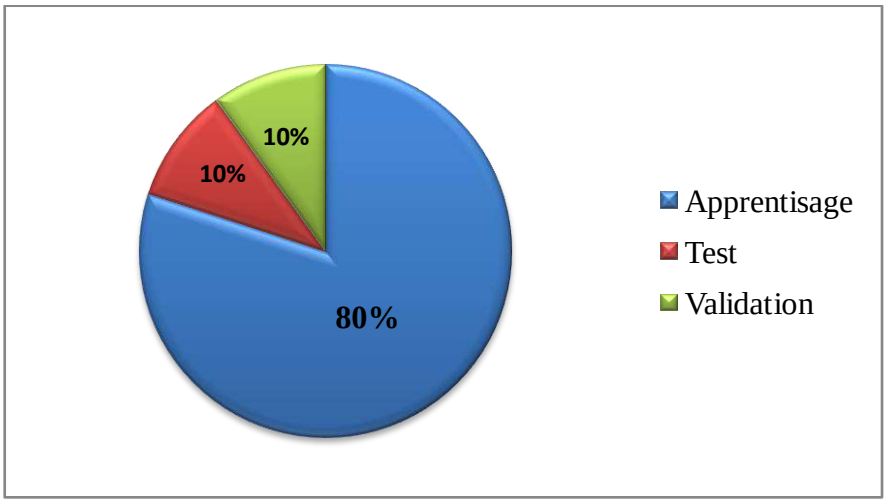


Figure III.3. Représentation des % de division des données par MATLAB R2010b.

Le logiciel STATISTICA a été utilisé pour l'élaboration du RNFF décrit ci-dessus. Ce logiciel pour optimiser l'architecture des trois modèles neuronaux (BD simple, BD référence, BD réduite) employés pour la rétention des MO par les membranes d'OI et NF. Pour ce logiciel on a divisée les trois bases en deux sous ensembles suit, 80% pour la base d'apprentissage et 20% pour la base de test comme c'est montré dans la figure III.4.

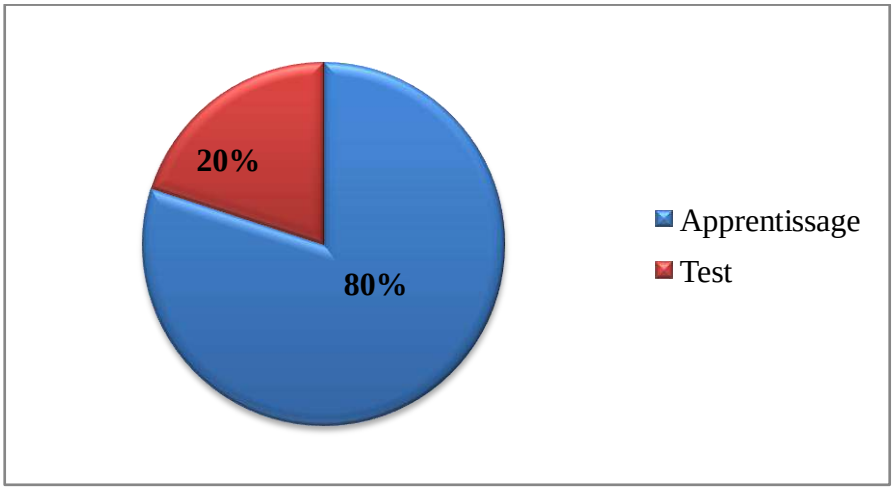


Figure III.4. Représentation des % de division des données par STATISTICA.

L'algorithme d'apprentissage utilisé dans notre étude est BFGS quasi-Newton *(trianbfg)*. Les trois modèles de RFFF sont caractérisés par la même structure générale (une couche d'entrée, une couche cachée, et une couche de sortie) : une couche d'entrée avec 14 neurones, une couche cachée avec un nombre de neurones actifs optimisés selon lors de la phase d'apprentissage, et une couche de sortie avec un neurone générant la valeur estimée de la rétention (*Rét*). Le nombre de neurones cachés a été varié de 3 à 25 neurones. Les fonctions d'activation de la couche cachée sont : la tangente hyperbolique sigmoïde (*Tansig*), le sigmoïde logarithmique *(logtsic*), exponentielle et *identity*. Pour la sortie on a utilisé les mêmes fonctions d'activation de la couche cachée. On a sauvegardé les résultats obtenus paramètres de neurones optimal et on fait la discussion.

III.2.3. Calcul des erreurs

Afin d'évaluer l'aptitude et la précision de prédiction des modèles neuronaux (RNOEP et RNRB) et du modèle polynomial on a utilisé la racine de l'erreur quadratique moyenne (*REQM*), L'erreur standard de prédiction (*ESP*) et l'erreur de prédiction du modèle (*EAM*) (Kumar *et al.* 2011; Lu *et al.* 2011; Zafar *et al.* 2012a; Zafar *et al.* 2012b).

$$REQM(\%) = \sqrt{\frac{\sum_{i=1}^{n} (y_{i,exp} - y_{i,cal})^2}{n}} \quad \text{(III.7)}$$

$$ESP(\%) = \frac{REQM}{y_e} * 100 \quad \text{(III.8)}$$

$$EAM(\%) = \frac{100}{n} \sum_{i=1}^{n} (|y_{i,exp} - y_{i,cal}|) \quad \text{(III.9)}$$

Où $y_{i,exp}$ est la valeur expérimentale ; $y_{i,cal}$ représente la valeur prédite ; y_e est la moyenne des données expérimentales ; n est le nombre d'expériences.

III.2.4. Analyse de sensibilité

Pour voir la contribution ainsi que le profil de variation de chaque variable d'entrée (*PM, logD, MD,* La largeur, *MWCO, SR, PZ, AC, pH, P, r, T*) sur la sortie de chaque réseau (la rétention (*R%*)) On a procédé à une analyse de sensibilité par la méthode de "Poids" (en anglais 'Weights' method) (Khaouane 2013).

III.2.4.1. Méthode de "Poids"

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, qui se compose de quatre étapes, est schématisé sur l'organigramme De la figure III.5.

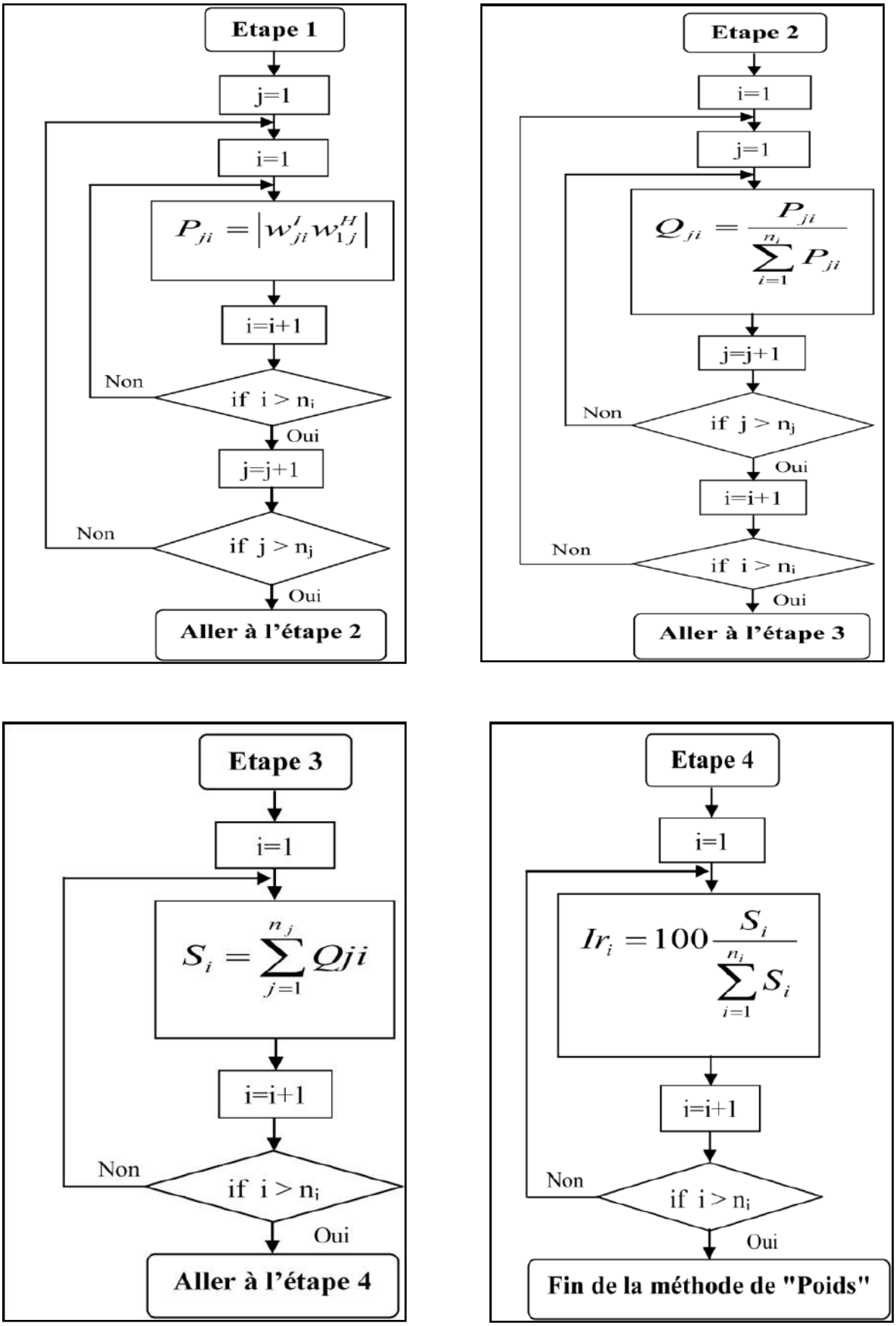


Figure III.5. Organigramme de la méthode de "poids".

III.3. Résultats et discussion

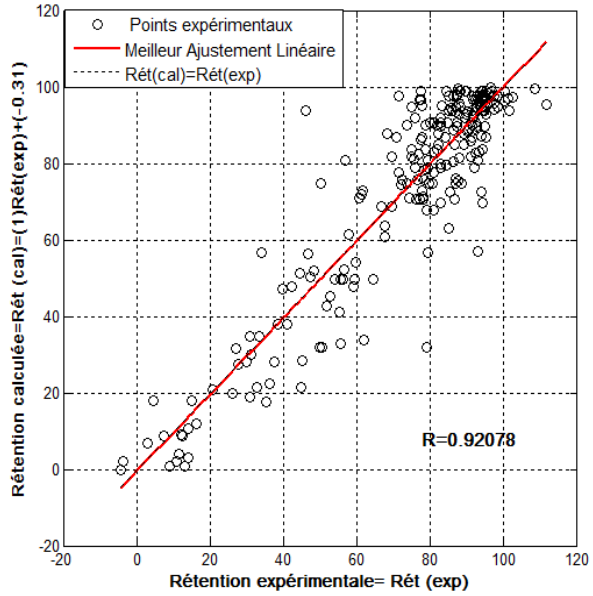
III.3.1. Performance des modèles

Selon la discussion précédente, trois modèles neuronaux ont été développés dans le but de Prédire la rétention des molécules organiques par les membranes d'OI/NF. Les structures obtenues des modèles neuronaux optimisés sont représentes dans le tableau III.4, une seul couche cachée était suffisante pour prévoir avec précision le profil de la rétention pour les trois modèle neuronaux.

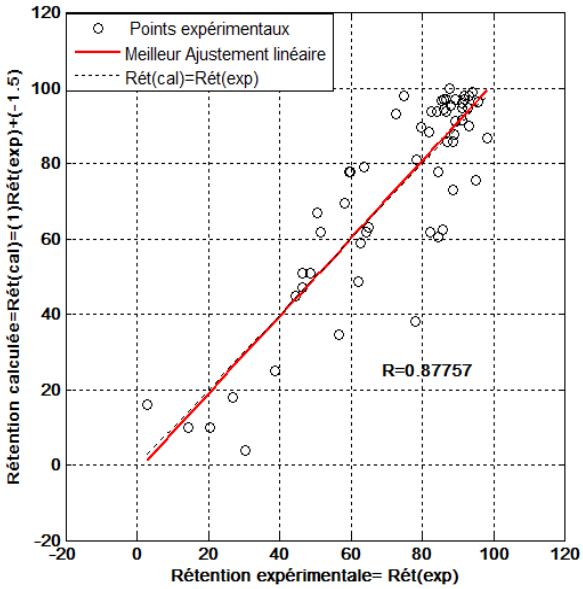
Tableau III.4. Structure des modèles neuronaux optimisés.

Modèles neuronaux	**Algorithme d'apprentissage**	**Couche d'entrée**	**Couche cachée**		**Couche de sortie**	
		Nombre de neurone	**Nombre de neurone**	**Fonction d'activation**	**Nombre de neurone**	**Fonction d'activation**
QSAR-RN1	BFGS	14	7	Tanh	1	Identité
QSAR-RN2	BFGS	14	19	Tanh	1	Logistic
QSAR-RN3	BFGS	14	7	Logistic	1	Exponential

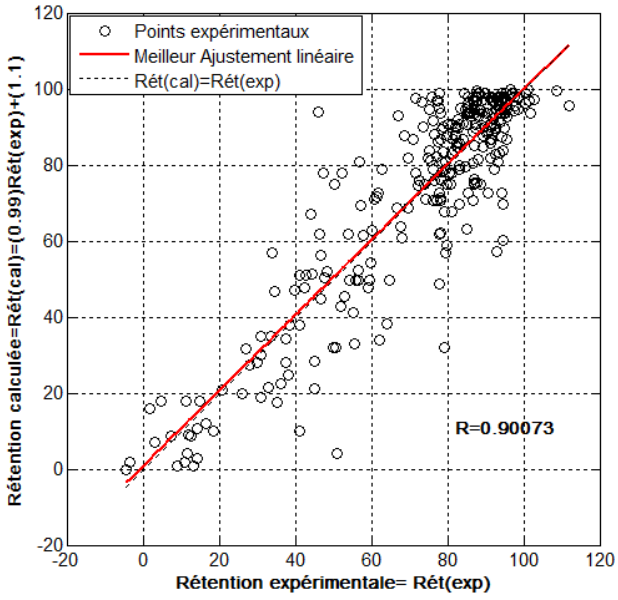
Les figures III.6, III.7 et III.8 montrent la comparaison entre les valeurs expérimentales et les valeurs calculées par les modèles neuronaux optimisés pour la totalité de la base de données, la base d'apprentissage et test pour RN1, RN2 et RN3 respectivement.



(a)



(b)



(c)

Figure III.6. Comparaison entre les valeurs expérimentales et calculées par les modèles neuronaux optimisés pour l'ensemble de données de QSAR-RN1 : **(a)** base d'apprentissage, **(b)** base de test, **(c)** base totale.

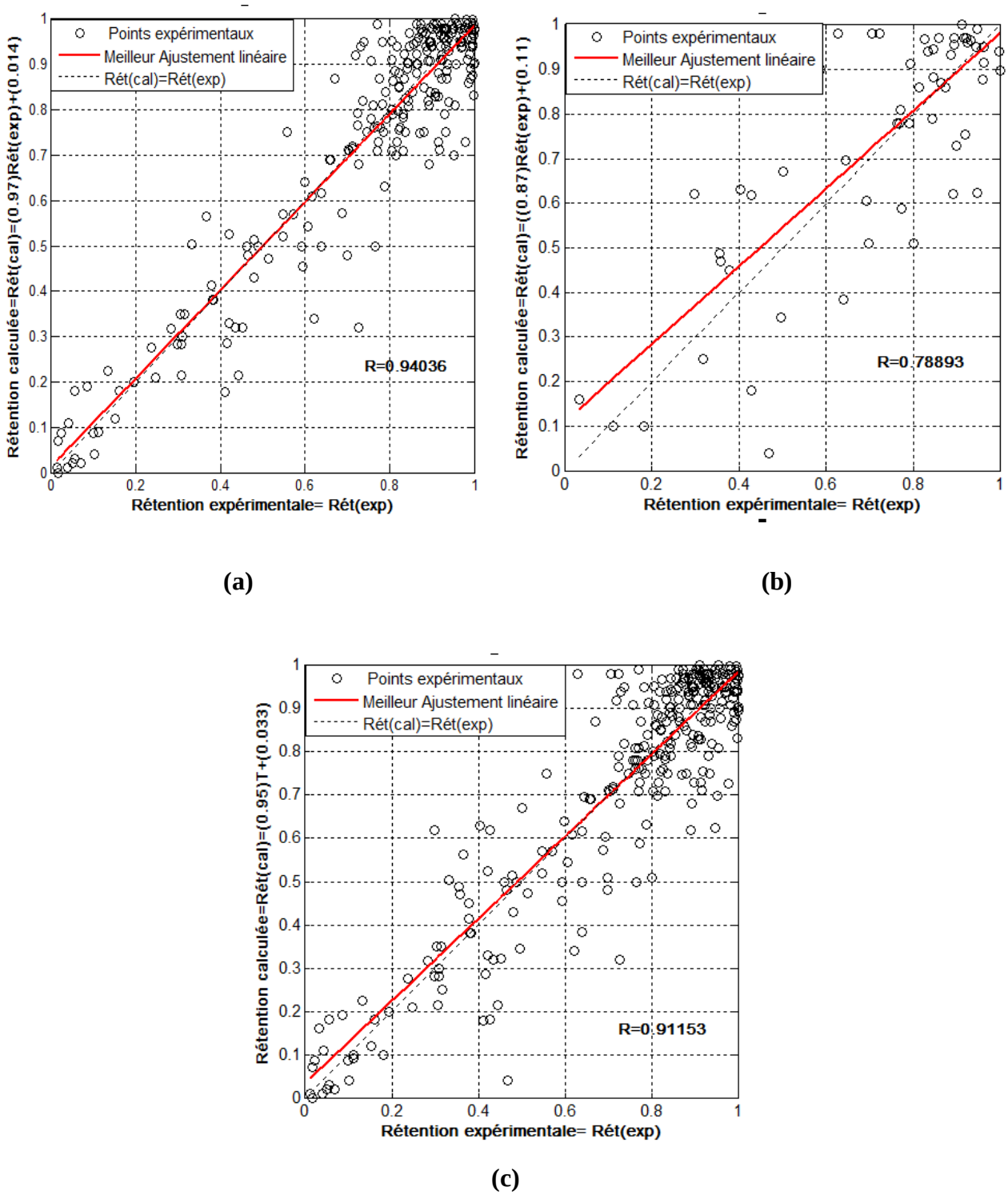
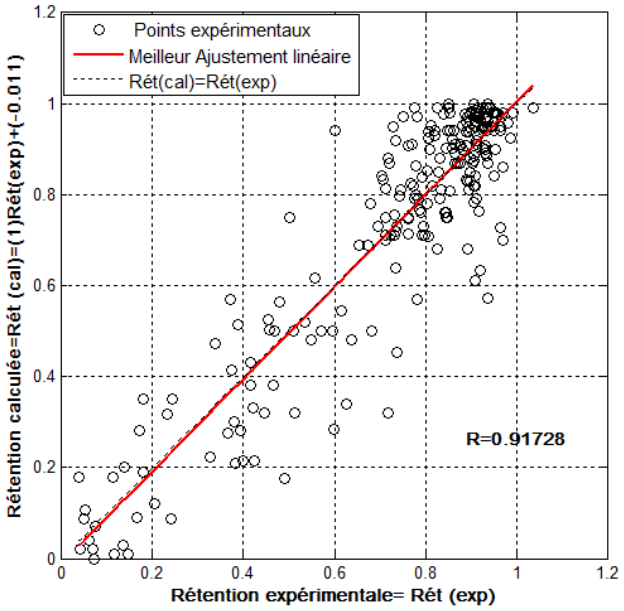
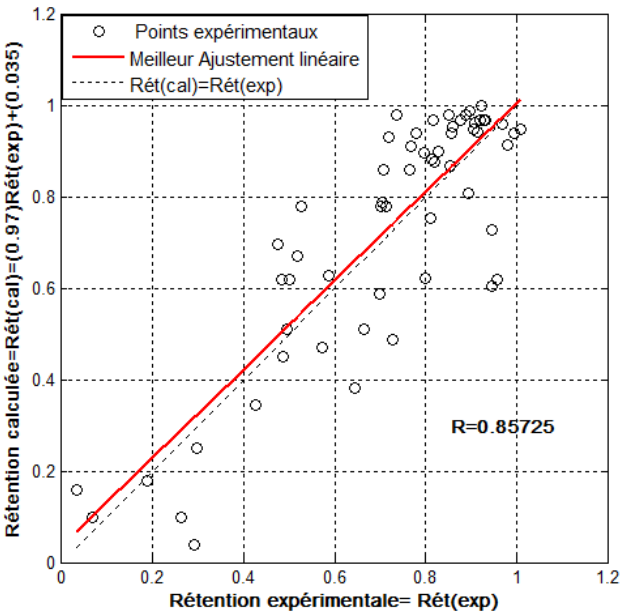


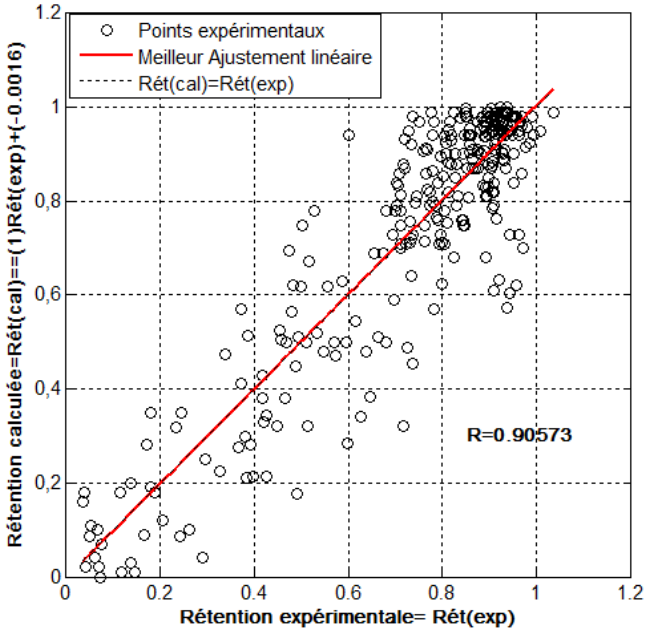
Figure III.7. Comparaison entre les valeurs expérimentales et calculées par les modèles neuronaux optimisés pour l'ensemble de données de QSAR-RN2 : **(a)** base d'apprentissage, **(b)** base de test, **(c)** base totale.



(a)



(b)



(c)

Figure III.8. Comparaison entre les valeurs expérimentales et calculées par les modèles neuronaux optimisés pour l'ensemble de données de QSAR-RN3 : **(a)** base d'apprentissage, **(b)** base de test, **(c)** base totale.

Le tableau III.5 rassemble les vecteurs de régression linéaire ainsi que l'erreur quadratique moyenne (en anglais mean squared error (MSE)) pour les trois modèles neuronaux (QSAR-RN1, QSAR-RN2 et QSAR-RN3). Ce tableau montre que les paramètres de régression les plus favorables sont ceux obtenus pour la base de donnée de référence. Il est prouvé que l'approche neuronale proposée donne des résultants satisfaisants avec des valeurs des vecteurs de régression approchant l'idéal [i.e. α=1 (pente), β=0 (ordonnée à l'origine), R=1 (coefficient de corrélation)] .Les vecteurs de régression ont été obtenus en utilisant la fonction MATLAB "postreg", le MSE a été obtenu en utilisant le programme STATISTICA.

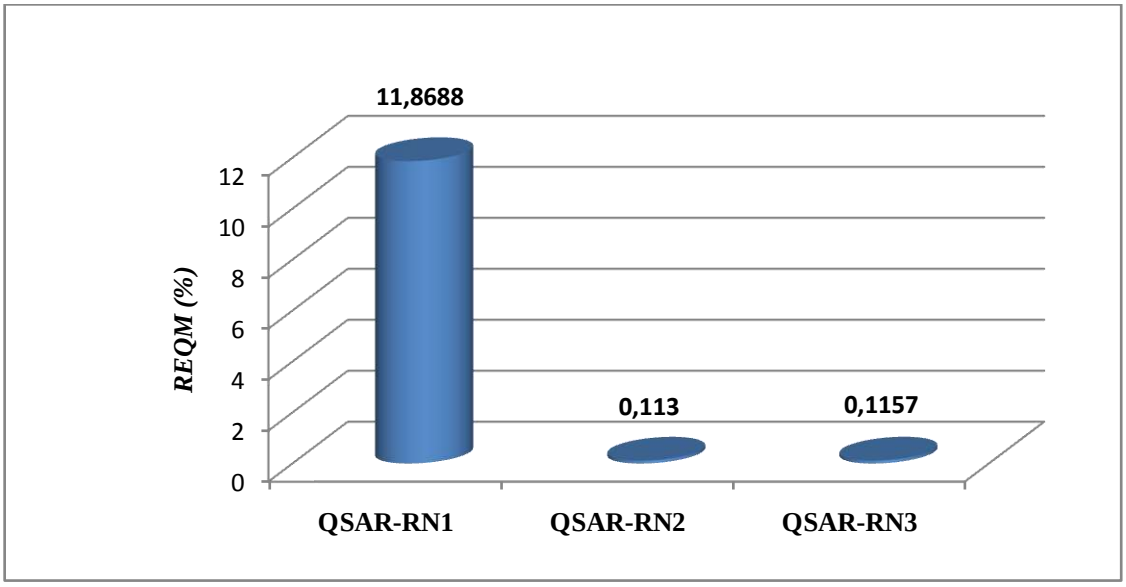
Tableau III.5. Vecteurs de régression linéaire issue de l'équation linéaire :
$Rét(cal) = \alpha Rét(exp) + \beta$, [$\alpha$ (pente), β (ordonnée à l'origine), R (coefficient de corrélation)], *MSE* pour QSAR-RN1, QSAR-RN2 et QSAR-RN3.

		α	β	R	MSE
QSAR-RN1	**Apprentissage**	1.0046	-0.3091	0.9208	0.004526
	Test	1.0291	-1.5262	0.8776	0.00822
	Totale	0.9919	1.1273	0.90073	-
QSAR-RN2	**Apprentissage**	0.971	0.0143	0.9404	0.004413
	Test	0.8722	0.1099	0.7889	0.014296
	Totale	0.9513	0.0333	0.9115	-
QSAR-RN3	**Apprentissage**	1.0142	-0.112	0.9173	0.005977
	Test	0.9725	0.0351	0.8572	0.009576
	Totale	1.0055	-0.0016	0.9057	-

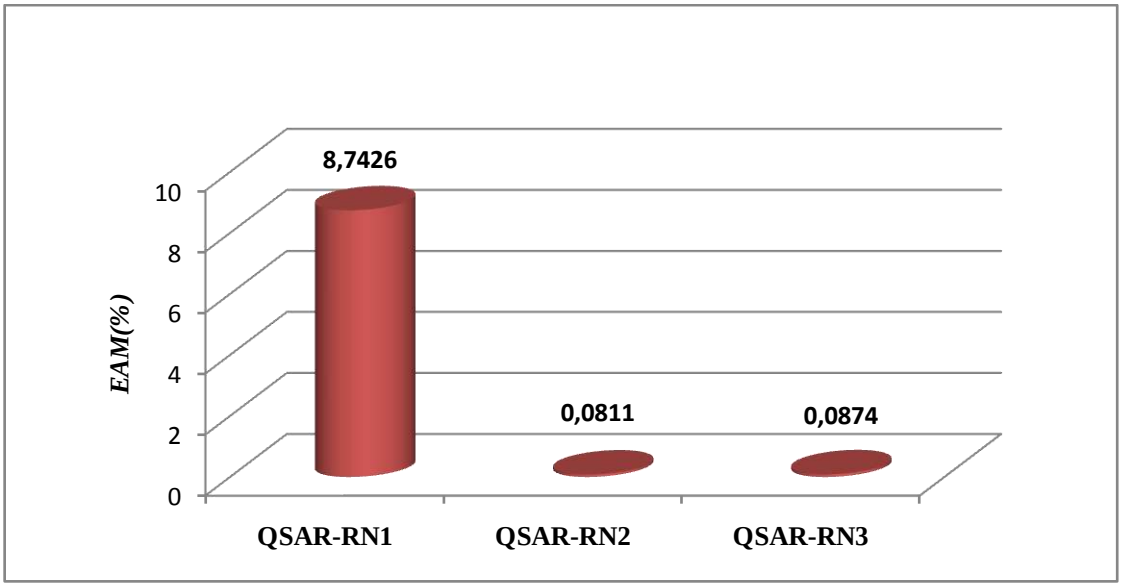
Le tableau III.6 résument les valeurs prisent par les erreurs calculées : la Racine de l'Erreur Quadratique Moyenne (en anglais Root Mean Squared Error (*RMSE*), l'Erreur Absolue Moyenne (en anglais Mean Absolute Error (*MAE*) et l'Erreur Standard de Prédiction (en anglais Standard Error of Predection (*SPE*) qu'on a calculées a partie des équations III.6, III.7 et III.8 respectivement.

Tableau III.6. Les erreurs calculées pour la valeur moyenne de la base d'apprentissage, test et totale pour les trois modèles neuronaux (RN1, RN2 et RN3).

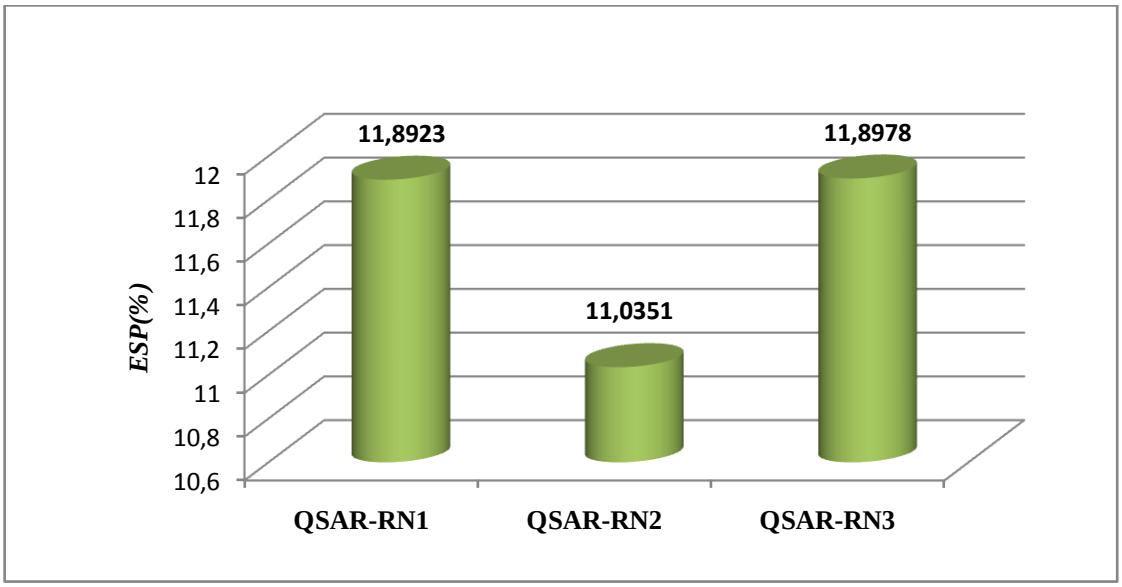
		REQM (%)	*EAM (%)*	*ESP (%)*
QSAR-RN1	**Apprentissage**	10.7045	7.8717	10.6877
	Test	12.8218	10.0264	13.7413
	Totale	11.8688	8.7426	11.8923
QSAR-RN2	**Apprentissage**	0.0939	0.069	9.3738
	Test	0.1669	0.1279	17.3524
	Totale	0.113	0.0811	11.0351
QSAR-RN3	**Apprentissage**	0.1093	0.0809	10.983
	Test	0.1384	0.1134	15.5475
	Totale	0.1157	0.0874	11.8978



(a)



(b)



(c)

Figure III.9. Performances des modèles neuronaux (QSAR-RN1, QSAR-RN2, QSAR-RN3) **(a)** : la Racine de l'Erreur Quadratique Moyenne (REQM%), **(b)** : l'Erreur Absolue Moyenne (EAM%) et**(c)** : l'Erreur Standard de Prédiction (ESP%)

A partir du tableau III.6 et la figure III.9, Les erreurs de modèle QSAR-RN2 sont légèrement inférieures à creux du modèles QSAR-RN1 et QSAR-RN3 .Ces résultats montrent que la normalisation référentielle est la méthode la plus efficace pour la modélisation de la rétention des molécules organiques par les membranes d'OI/NF en utilisant les réseaux de neurones.

III.3.2. Expressions mathématique des modèles neuronaux optimisés (modèle RN)

À partir du QSAR-RN2 optimisée on peut exprimer la rétention de molécules organiques par un modèle mathématique qui y incorpore toutes les entrées E_i comme suit:

Les sorties Z_j de la couche cachée:

$$Z_j = f_H\left[\sum_{i=1}^{14} W_{ji}^I\, E_i + b_j^H\right] = \frac{\exp\left(\sum_{i=1}^{14} W_{ij}^I E_i + b_j^H\right) - \exp\left(-\sum_{i=1}^{14} W_{ij}^I E_i + b_j^H\right)}{\exp\left(\sum_{i=1}^{14} W_{ij}^I E_i + b_j^H\right) + \exp\left(-\sum_{i=1}^{14} W_{ij}^I E_i + b_j^H\right)} \quad \text{(III.8)}$$

j=1, 2,…, 19.

La sortie *Rétention*:

$$Rétention = f_0\left[\sum_{j=1}^{19} W_{1j}^H\, Z_j + b_1^0\right] = \sum_{j=1}^{19} W_{1j}^H\, Z_j + b_1^0 \quad \text{(III.9)}$$

La combinaison des équations (III.8) et (III.9) mène à la formule mathématique suivante qui décrit la rétention des molécules organiques en prenant en compte toutes les entrées E_i .

$$Rétention = \sum_{j=1}^{19} W_{1j}^H \left[\frac{\exp\left(\sum_{i=1}^{14} W_{ij}^I E_i + b_j^H\right) - \exp\left(-\sum_{i=1}^{14} W_{ij}^I E_i + b_j^H\right)}{\exp\left(\sum_{i=1}^{14} W_{ij}^I E_i + b_j^H\right) + \exp\left(-\sum_{i=1}^{14} W_{ij}^I E_i + b_j^H\right)}\right] + b_1^0 \quad \text{(III.10)}$$

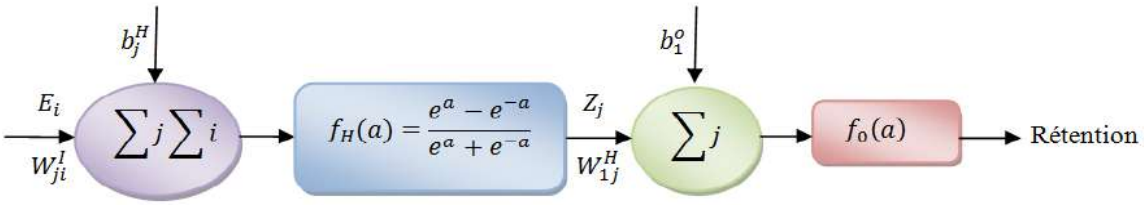


Figure III.10. Représentation schématique du QSAR-RN2 optimisés.

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 ; W_{1j}^H représente le poids synaptique de la connexion du neurone *j* de la couche cachée au neurone de la couche de sortie; b_j^H et b_1^0 représentent respectivement les biais du neurone *j* de la couche cachée et du neurone de la couche de sortie.

L'architecture des réseaux de neurones optimaux obtenus est donnée sur la figure III.11, ci-après.

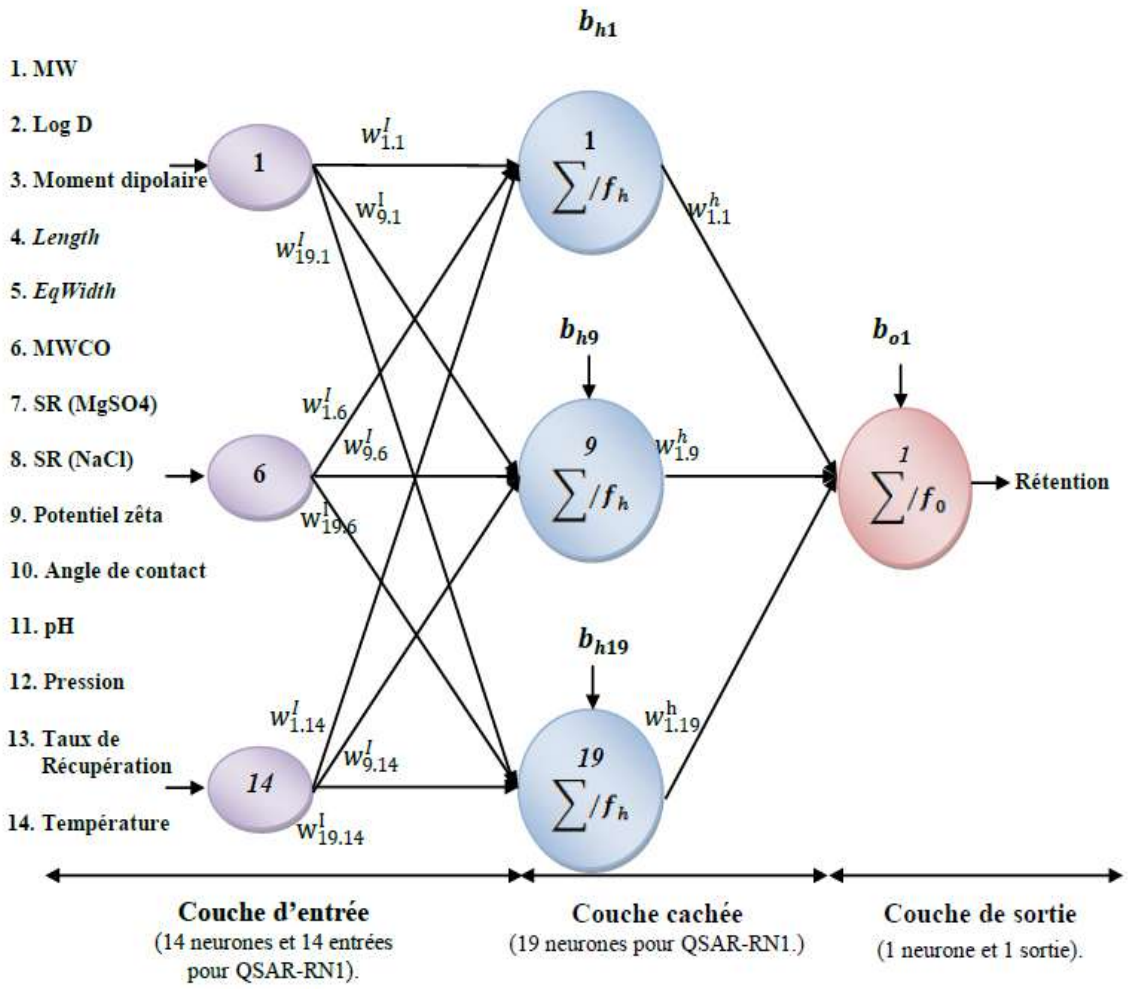


Figure III.11. Architecture finale du réseau de neurone optimal de type feedforward pour la prédiction de la rétention des molécules organiques (neutres et ioniques) par les membranes d'OI/NF.

III.3.2.1.Analyse de sensibilité

Pour voir la contribution ainsi que le profil de variation de chaque variable d'entrée sur la sortie réseau RN2, On a procédé à une analyse de sensibilité par la méthode de "Poids".

- **Méthode de "Poids"**

Cette méthode a été appliquée sur RN*2* en utilisant leur poids. Les contributions des variables d'entrée obtenues par la méthode de "poids" pour ce RN est présentée sur la figure III.12.

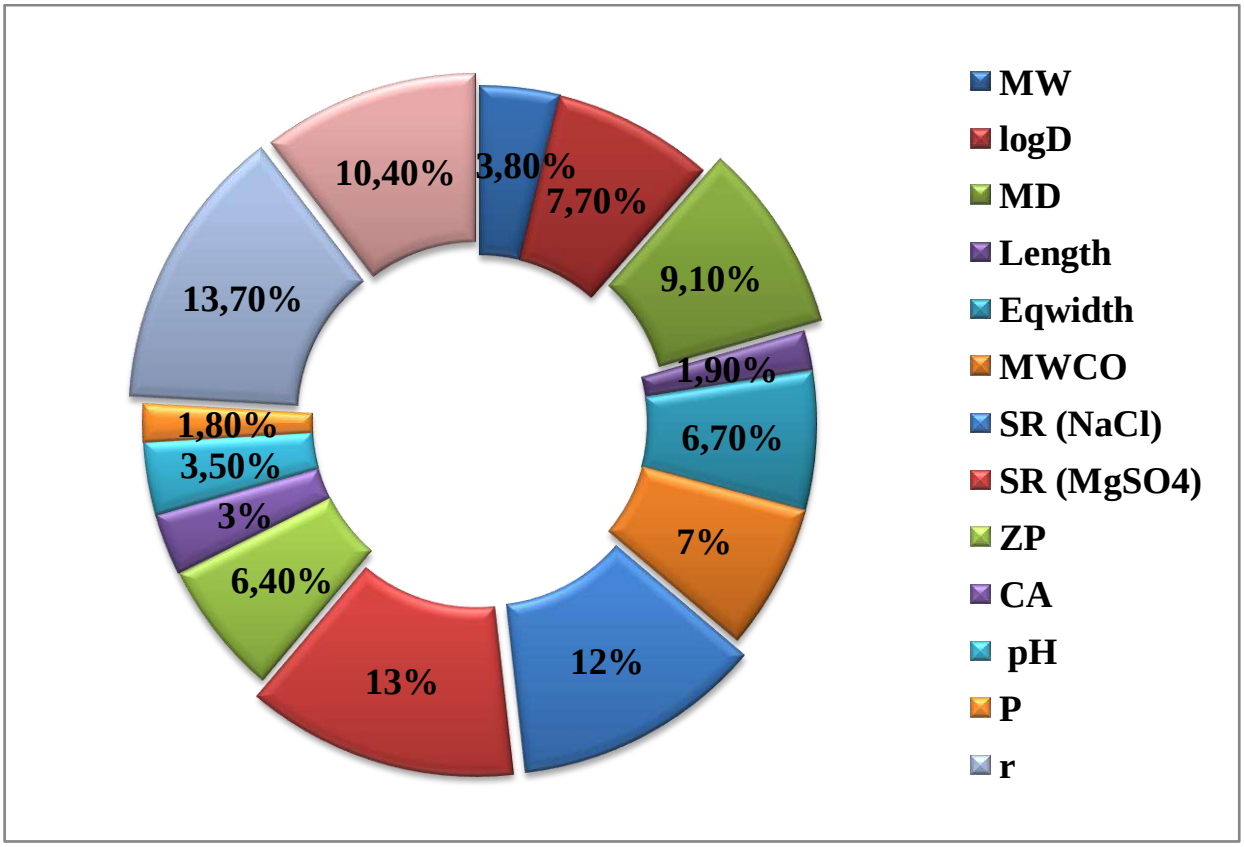


Figure III.12. Présentation des % de l'importance relative (*Ir*) de chaque variable d'entrée sur la rétention de QSAR-RN2.

D'après la figure III.12 on remarque que la rétention des sels pour $MgSO_4$ et NaCl, taux de récupération, température et le moment dipolaire exercent la contribution la plus importante pour le QSAR-RN2 avec une contribution descendante pour les restes entrées. D'où on déduit qu'il est très important de valoriser les caractéristiques des membranes car la sortie inespérée se différent à cause de ces derniers.

III.3.3. Modèle RLM

Le modèle obtenu pour la prédiction de la rétention des molécules organiques, est le modèle linéaire suivant (équation III.11) avec les paramètres statistiques y afférant :

$$Rétention = 0.456025 - 0.00014429(PM) + 0.04412375(logD) + 0.2586445(MD) + 0.959112007(length) + 0.00674625(Eqwidth) - 0.2692468(MWCO) - 0.05723072(SR_{NaCl}) - 0.05529247(SR_{MgSO_4}) - 0.01579126(PZ) + 0.01766275(AC) + 0.04828082(pH) - 0.00618802(P) + 0.7284517(r) - 0.12808143(T) \quad \text{(III.11)}$$

Les signes des coefficients de régression indiquent la direction de l'influence des entrées dans un modèle donné. Selon les coefficients de régression de l'équation (III.11), il est à noter que la longueur moléculaire (*length*) et le poids moléculaire sont ceux qui contribuent le plus à la rétention. Les coefficients de régression des paramètres (*logD, MD, length, Eqwidth, AC, pH, r*) présentent des signes positifs. Ces paramètres ont fourni une influence positive et les valeurs les plus élevées sont favorables à l'amélioration de la rétention des MO. Par ailleurs, les coefficients de régression des paramètres (*PM, MWCO,* SR_{NaCl} , SR_{MgSO_4} , PZ, P, T) présentent des signes négatifs, donnant ainsi un impact négatif sur la rétention.

Par ailleurs, le modèle présente une valeur du coefficient de corrélation *R* de 0,639, ce qui implique une mauvaise concordance entre les valeurs prédite et expérimentales (figure III.13). D'autre part, la Racine de l'Erreur Quadratique Moyenne (*REQM*), l'Erreur Absolue Moyenne (*EAM*) et l'Erreur Standard de Prédiction (*ESP*) pour le modèle ont été indiquées dans le tableau III.7.

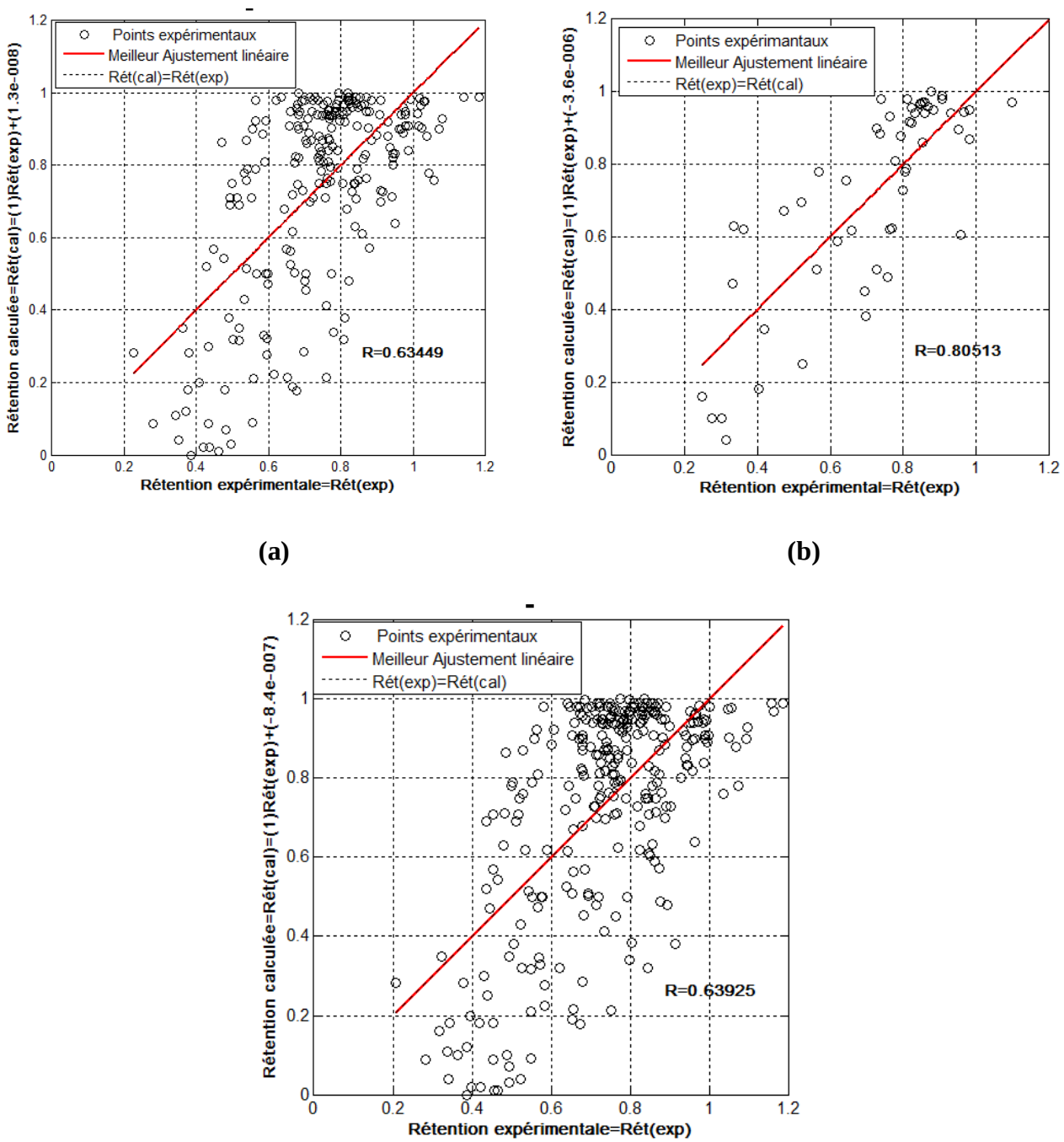


Figure III.13. Comparaison entre les valeurs expérimentales et calculées par les modèles neuronaux optimisés pour l'ensemble de données de QSAR-RLM : **(a)** base d'apprentissage, **(b)** base de test, **(c)** base totale.

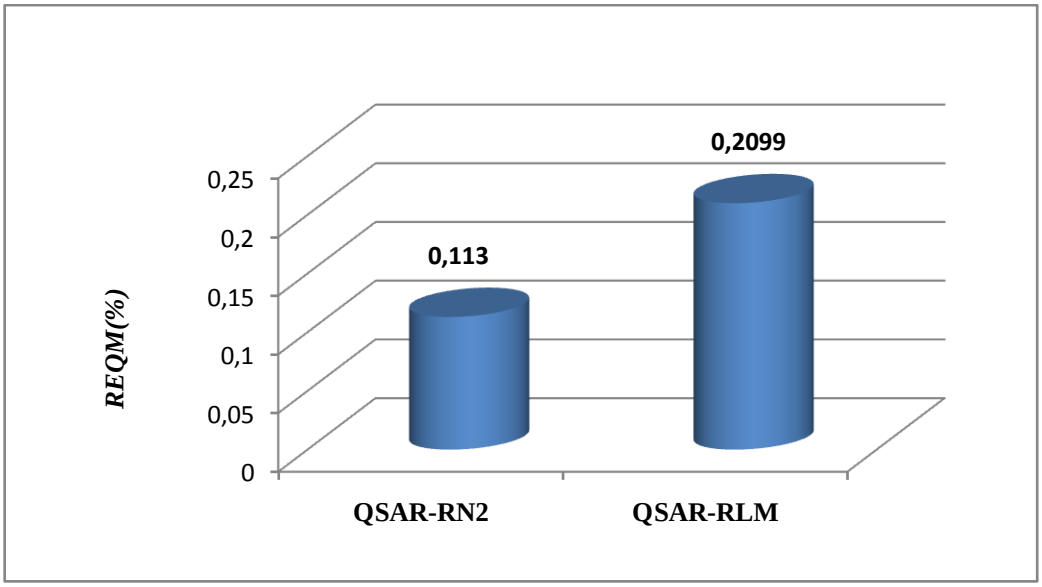
Tableau III.7. Les erreurs calculées pour la valeur moyenne de la base d'apprentissage, test et totale pour le modèle QSAR-RLM.

		REQM (%)	***EPM (%)***	***ESP (%)***
QSAR-MLR	**Apprentissage**	0.2121	0.1762	23.9234
	Test	0.1582	0.1325	18.1576
	Totale	0.2099	0.1741	23.6778

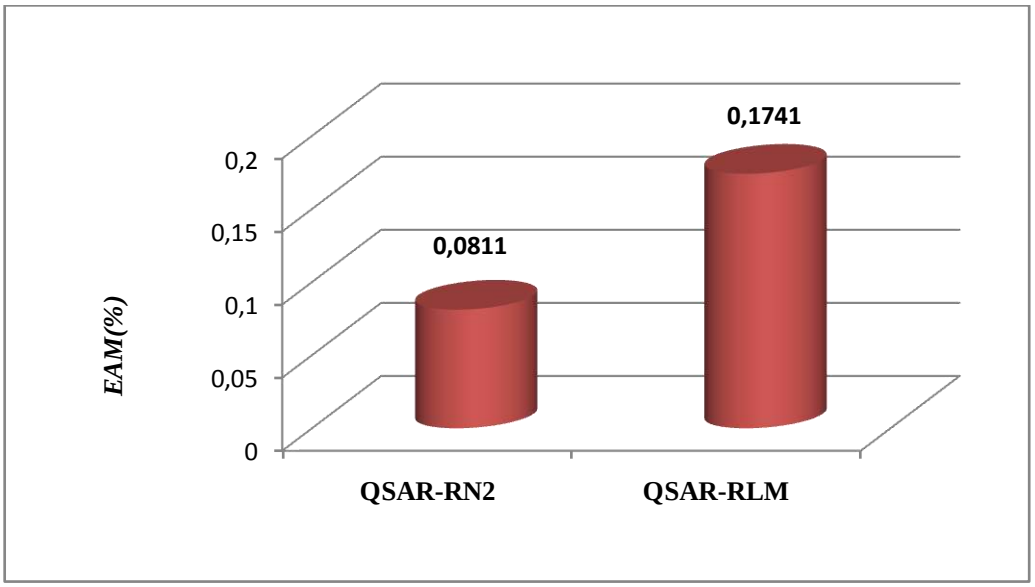
III.3.4. Comparaison entre les modèles RLM et RNA

Pour comparer à la fois les performances et la qualité de la modélisation des deux modèles obtenus dans ce travail (QSAR-RN2 et QSAR-RLM), une évaluation statistique des deux modèles pour l'ensemble de validation est donnée dans les tableaux III.6 et III.7. Les coefficients de corrélation (R) ayant une valeur supérieure à 0,8 indiquent que la valeur de la rétention est acceptable. Il en est de même pour la racine carrée de l'erreur quadratique moyenne (*REQM%*) dont les valeurs sont inférieures à 0,3 qui illustré dans la figure III.14, **(a).** ainsi que pour l'Erreur Absolue Moyenne (*EAM%*) et l'Erreur Standard de Prédiction (*ESP%)* pour le modèle ont été indiquées dans les figures III.14. **(b)** et **(c)**. Cependant, la prédiction déterminée par le modèle QSAR-RN2 est meilleure que celle fournie par le modèle QSAR-RLM.

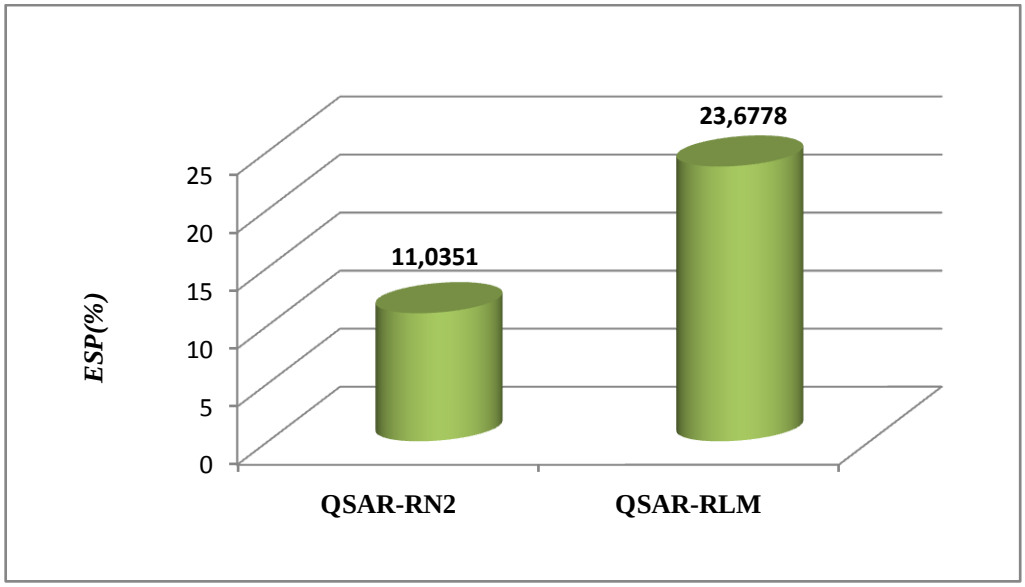
Par rapport aux résidus du modèle QSAR-RLM, ceux donnés par le modèle QSAR-RN2 sont plus près de la ligne médiane représentant l'égalité entre la toxicité observée et celle prédite (figure III.7 et III.13). Aussi, on peut donc conclure que le modèle QSAR-RN2 a un meilleur pouvoir prédictif que le modèle QSAR-RLM. Cela signifie que le modèle obtenu avec le réseau de neurones permet, dans une large mesure, de confirmer qu'une relation non linéaire relie la rétention des MO à leur structure moléculaire.


REQM(%)
0,25
0,2
0,15
0,1
0,05
0
0,113
0,2099
QSAR-RN2
QSAR-RLM

(a)


EAM(%)
0,2
0,15
0,1
0,05
0
0,0811
0,1741
QSAR-RN2
QSAR-RLM

(b)



(c)

Figure III.14. Performances des modèles neuronaux (QSAR-RN2, QSAR-RLM).
(a) : la Racine de l'Erreur Quadratique Moyenne *(REQM%)*, **(b)** : l'Erreur Absolue Moyenne *(EAM%)* et **(c)** : l'Erreur Standard de Prédiction *(ESP%)*

III.3.5. Comparaison entre les algorithmes de retro-propagation

Nous avons vu que le travail d'optimisation de l'architecture du réseau de neurones consiste en résumé à retenir la meilleure fonction d'activation pour un nombre de neurones optimal à l'obtention de la meilleure performance possible compte tenu des données disponibles. Cette optimisation structurelle nous amènera, préalablement à tout calcul, à sélectionner la fonction d'activation parmi les fonctions existantes. Lors de l'apprentissage, un algorithme de rétro-propagation permet de minimiser la fonction d'erreur entre valeurs mesurées et valeurs calculées. Cette minimisation est effectuée en modifiant les poids des connexions de la sortie jusqu'aux entrées. C'est une des règles les plus utilisées pour l'apprentissage supervisé de réseaux de neurones. Nous avons donc utilisé une méthode du second ordre comme algorithme de minimisation itérative des poids du modèle : celle de *nftool* (Levenberg-Marquardt). Cette technique permet en effet de converger vers le minimum de la fonction d'erreur plus rapidement du fait qu'elle considère la dérivée seconde de la fonction d'erreur (Hamdache 2014). La figure III.15 a été indiquée ces caractéristiques. Le temps que l'algorithme BFGS quasi-Newton *(trianbfg)* nécessite plus de calcul à chaque itération que les algorithmes de gradients conjugués, mais possède habituellement une convergence plus rapide.

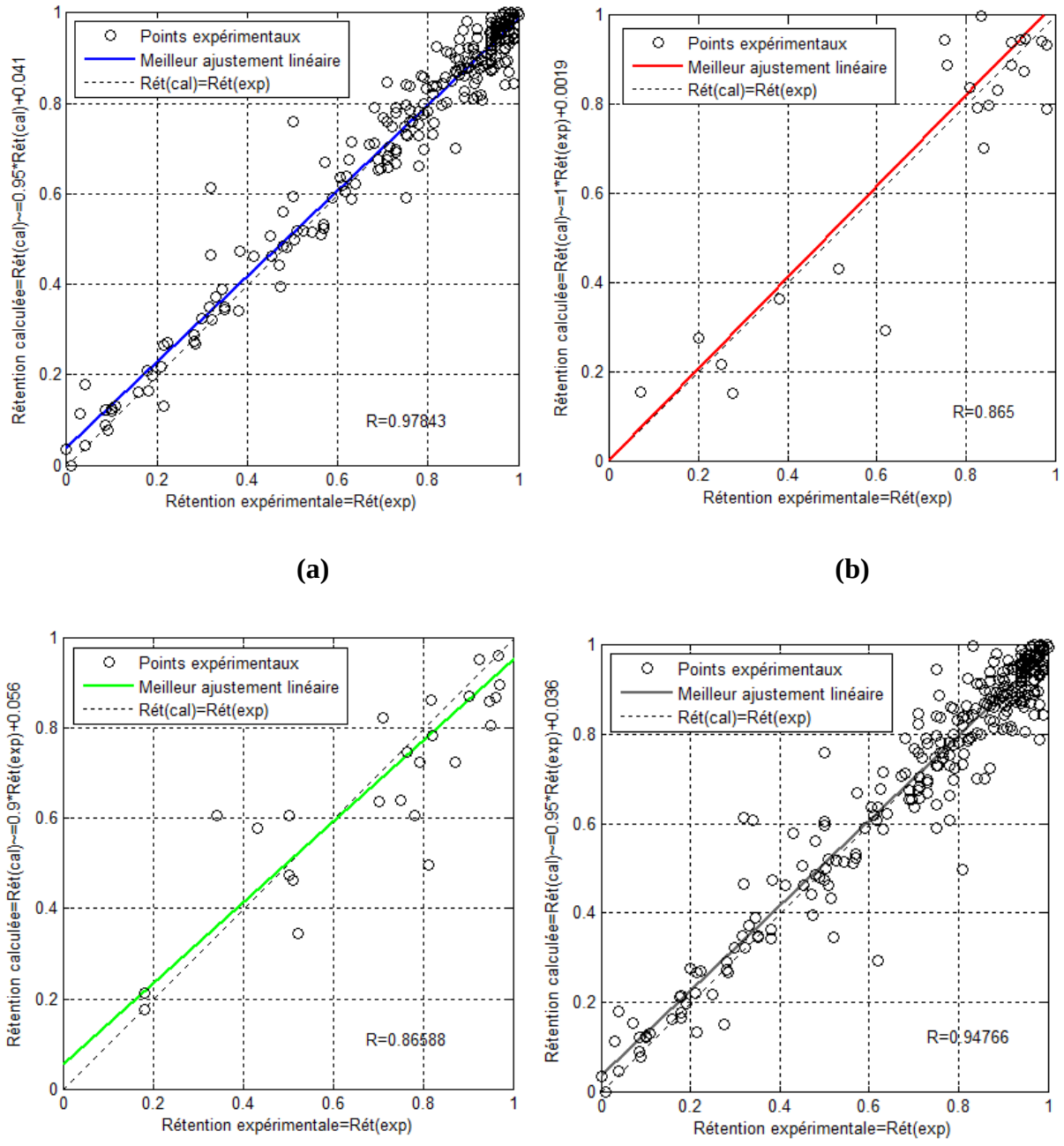


Figure III.15. Comparaison entre les valeurs expérimentales et calculées par les modèles neuronaux optimisés pour l'ensemble de données de QSAR-RN2 : **(a)** base d'apprentissage, **(b)** base validation, **(c)** base test, **(d)** base totale

III.4. Conclusion

Dans ce chapitre, nous avons établi un modèle neuronal (QSAR-RN) et un modèle de régression linéaire (QSAR-RLM) pour la modélisation de la rétention des MO lors l'OI/NF. Ce modèle de prédiction a été obtenu suite à un traitement préliminaire de la base de données initiale, laquelle nous a permis de comparer les deux algorithmes d'apprentissage (Levenberg-Marquardt et quasi-newton). Suite à cela, deux modèles QSAR ont été élaborés. Par comparaison avec les résultats satisfaisants du modèle QSAR-RLM, ceux du modèle neuronal sont plus performants. Les différents paramètres statistiques nous ont permis de constater que le modèle QSAR-RN2 est performant. Par ailleurs, le modèle mathématique issu de l'étude QSAR neuronale a été appliqué à la rétention de MO lors l'OI/NF.

Conclusion générale et Perspectives

L'objectif de cette mémoire était de mettre au point un modèle statistique de prédiction de la rétention des molécules organiques lors l'osmose inverse et nonofiltration. Cette étude s'intègre dans le cadre de la quantification de cette rétention sans recourir à l'expérimentation réel. Pour atteindre cet objectif, les modèles QSAR ont été établis en utilisant deux méthodes d'apprentissage statistique : la régression linéaire multiple (RLM) et les réseaux de neurones artificiels (RN).

Notre première contribution concerne tout d'abord l'élaboration d'une base de données. Dans toute recherche des modèles QSAR, l'établissement d'une base de données fiable est l'une des clés de réussite. On a recueilli la rétention de 291 échantillons. Le choix des variables nécessaires à la description de la rétention puisque cette dernière dépend d'une part à la structure et les propriétés moléculaires et d'autre au part aux propriétés physiques des membranes. Puis on a effectuée deux méthodes de normalisation pour notre base, l'une est référentielle et l'autre réduite par conséquent on a obtenus trois bases de données.

Notre deuxième contribution a porté sur l'utilisation des modèles QSAR pour développer deux modèles pour la prédiction de la rétention des molécules organiques : l'un non linéaire et l'autre linéaire.

Dans un premier temps, on a optée pour la modélisation par utilisation des réseaux de neurones artificiels (QSAR-RN). Pour cela, trois modèles neuronaux de type feedforward sont construits. Les trois modèles neuronaux sont caractérisés par une structure similaire : 14 neurones dans la couche d'entrée (le poids moléculaires (*PM*), le logarithme du coefficient de distribution octanol-eau (*logD*), le moment dipolaire (*MD*), longueur moléculaire (*Length*), la largeur équivalente moléculaire (*Eqwidth*), le poids moléculaire de coupure (*MWCO*), la rétention des sels (*SR*) pour $MgSO_4$ et NaCl, le potentiel zêta (*PZ*), l'angle de contact (*AC*), *pH*, la pression (*P*), le taux de récupération (*r*) et la température (*T*)) , une couche cachée avec un nombre de neurones optimisé au cours du processus d'apprentissage et une couche de sortie avec un seul neurone. Une bonne concordance, entre les valeurs expérimentales et les valeurs prédites, pour chaque modèle neuronal a été obtenue (les racines des erreurs quadratiques moyennes pour la BD simple étaient de: 11.8688 %, 0.113% pour la BD de référence et 0.1157% pour la BD réduite). Par ailleurs, la comparaison entre les trois modèles neuronaux optimisés montré que la normalisation référentielle est la méthode la plus efficace pour la modélisation de la rétention des molécules organiques par les membranes d'OI/NF en utilisant les réseaux de neurones. Car le QSAR-RN2 à une bonne capacité prédictive d'interpolation avec un coefficient de corrélation proche à l'unité. D'autre part, l'analyse de sensibilité appliquée aux réseaux de neurone optimisés (QSAR-RN2) a démontré l'importance

de cinq paramètres (la rétention des sels pour $MgSO_4$ et NaCl, taux de récupération, température et le moment dipolaire) parmi les autres qui choisissent pour modéliser la rétention des molécules organiques.

Une modélisation par régression linéaire multiple (QSAR-MLR) a été menée à l'effet de développer un modèle. Ce modèle RLM obtenu pour la prédiction de la rétention des molécules organiques est un modèle linéaire. Le modèle présente une valeur du coefficient de corrélation R de 0,639, ce qui implique une mauvaise concordance entre les valeurs prédite et expérimentales, nous amène à conclure que le modèle RLM a un mauvais pouvoir prédictif de la rétention des molécules organiques.

A la fin une comparaison a été faite. Premièrement, Pour comparer à la fois les performances et la qualité de la prédiction des deux modèles obtenus dans ce travail (QSAR-RLM et QSAR-RNA), une évaluation statistique des deux modèles a été faite. Cette comparaison nous a permis de constater que la prédiction déterminée par le modèle QSAR-RNA est meilleure que celle fournie par le modèle QSAR-RLM. Une amélioration substantielle des paramètres statistiques pour le modèle QSAR-RNA a été notée. Une seconde comparaison entre les deux algorithmes d'apprentissage Levenberg-Marquardt et BFGS quasi-Newton résulte que l'algorithme de Levenberg-Marquardt est l'algorithme d'apprentissage le plus rapide et qui donne les meilleurs résultats pour les réseaux de taille modérée.

En conclusion, tout en estimant avoir atteint l'objectif tracé pour cette étude, de nombreuses voies de recherche ont émergé. Ces perspectives concernent plus particulièrement la prédiction d'autres phénomènes physique (tel que microfiltration, ultrafiltration,…etc). D'autre part, il faudrait s'intéresser à d'autres méthodes d'apprentissage statistique et d'autre type de modélisation.

Références bibliographiques

-A-

Absar Belkacem (2012)."Modélisation et simulation de systems de dessalement d'eau de mer par osmoses inverse et engeries renouvelables." *Thése de Doctorat*,université Mostaganeme, Algérie.

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.

Allègre C., Moulin P., Maisseu M. et Charbit F. (2004)." Savings and re-use of salts and water present in dye house effluents." *Desalination* 162: 13-22.

Aimar P. (2008)."Filtration membranaire (OI, NF, UF), Mise en oeuvre et performances." *Journal* 2-793.

Ammar M. Y. (2007). "Mise en œuvre de réseaux de neurones pour la modélisation de cinétiques réactionnelles en vue de la transposition batch/continu." *Thèse de Doctorat*, Institut National Polytechnique de Toulouse. Toulouse, France.

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.

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

Aptel P. (2008)." Filtration membranaire (OI, NF, UF), Applications en traitement des eaux." *Journal* 2-794.

Aptel Ph. et Buckley C.A. (1996)."Categories of membrane operations." In: Water Treatment Membrane Process, Chapitre 2, McGraw-Hill, P 24.

Ardhaoui M., Cherichi H., Ognier S. et Ghoul M. (2003). "Caractérisation des propriétés de surface des membranes de nanofiltration: développement d'un outil d'aide à la décision." Récents Progrès en Génie des Procédés 90 (SFGP 2003, St Nazaire): 32-39.

Arzate Alfa (2008)."Procédés de séparation membranaire et leur application dans l'industrie alimentaire."*Revue de littérateure* ACER centre de rechereche, de développement et de transfert technologique acéricole Inc.

Audinos R. (2008)." Membranes semi-perméables, Membranes d'osmose inverse." K 362, Techniques de l'Ingénieur.

Azizi Ahmed (2011)."Modélisation des phénomènes de transports lors de l'osmose inverse: Application aux molécules inorganiques et organiques en amont et en aval d'une unité de production pharmaceutique." *Mémoire de Magister*, université de Médéa, Algérie.

-B-

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: 2795-2809.

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

Bermejo S. et Cabestany J. (1999). "On-line gradient learning algorithms for K-nearest neighbor classifiers." Dans le livre: "Foundations and tools for neural modeling." *Springer Berlin Heidelberg* 1606**:** 546-555.

Bermejo S. et Cabestany J. (2001). "Learning with nearest neighbour classifiers." *Neural Processing Letters* 13(2): 159-181.

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

Bilstad T. (1995)." Nitrogen separation from domestic wastewater by reverse osmosis." *Journal of*

Membrane Science 102: 93-102.

Bishop C. M. (1995). "Neural networks for pattern recognition." Oxford University Press, New York, USA.

Bouamar M. et Ladjal M. (**2007)**. "Système multicapteur utilisant les réseaux de neurones artificiels pour la surveillance des eaux potables." 4th International Conference: *Sciences of Electronic, Technologies of Information and Telecommunications March.* Tunisia : 25-29.

Bouranene Saliha (2008)."Etude de la retention de solutes neuters et ioniques par des membranes de nanofiltration." *Thése de Doctorat*, université de Franche-comté.

-C-

Chikhi Mustapha (2008)."Etude de la complexation des cations metalliques en vue de leur separation par un procéde membranaire." *Thése de Doctorat*,université mentouri Constantine, Algérie.

Childress A. E. et Desmukh S. S. (1998)." Effects of humic substances and anionic surfactants on the surface charge and performance of reverse osmosis membranes." *Desalination* 118: 167-174.

-D-

Dames M. (2005). "Gestion de procèdes discontinus : Méthodologie de modélisation et d'optimisation d'opérations de dispersion liquide-liquide en cuve agitée." *Thèse de Doctorat*, Institut National Polytechnique de Toulouse. Toulouse, France.

Danis P. (2003)." Dessalement de l'eau de mer." Techniques de l'Ingénieur J 2700.

Debnath A. K (2001). "Quantitative Structure-Activity Relationship (QSAR) Paradigm-Hansch Era to New Millenium." *Mini Reviews in Medicinal Chemistry* I: 187-195.

Demuth H. et Beale M. (2000). "Neural network toolbox for use with Matlab." *User's Guide version* 4. USA, The MathWorks Inc.

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

Dickson J. M., Matsuura T., Blais P. et Sourirajan S., (1975)." Reverse osmosis separations of some organic and inorganic solutes in aqueous solutions using aromatic polyamide membranes." *Journal of Applied Polymer Science* 19: 801-819.

Dreyfus G Martinez J. M., Samuelides M., Gordon M. B., Badran F., Thiria S. et Hérault L. (2002). " Réseaux de neurones : Méthodologie et applications." *Éditions Eyrolles*, Paris.

Dreyfus G. (2008). "Réseaux de neurones: Méthodologie et applications." Eyrolles, Paris, France.

-E-

Ecosse D., (2001). "Techniques alternatives en vue de subvenir à la penurie d'eau dans le monde. Mem. D.E.S.S. (Qualite et Gestion de l'Eau)." *Fac. Sciences, Amiens* : 62.

-F-

Fell. (1995)."Membrane separation technologie;chapitre 4 Reverse osmosis."

Fontanals N'uria (2015)." Rejection of organic micropollutants by clean and fouled nanofiltration membranes."*Journal of chemistry* (9):1-9.

-G-

Galambos I., Mplina J. M., Jaray P., Vatai G.et Bekassy-Molnar E. (2004)."High organic content industrial wastewter treatment by membrane filtration."*Desalination* 162:117-120.

Garson G. D. (1991). "Interpreting neural-network connection weights." *Artificial Intelligence Expert* 6(4): 46-51.

Gevrey M., Dimopoulos I. et Lek S. (2003). "Review and comparison of methods to study the contribution of variables in artificial neural network models." *Ecological Modelling* 160(3): 249-264.

Girard B. et Fukumoto L. R. (2000). "Membrane processing of fruit juices and beverages." *A review*. Critical Reviews in Food Science and Nutrition 40: 109-175.

Goh A. T. C. (1995). "Back-propagation neural networks for modeling complex systems." *Artificial Intelligence in Engineering* 9(3): 143-151.

Gonzalez-Diaz.H, Olazabal.E, Santana.L, Uriarte.E, Gonzalez-Diaz.Y et Castanedo.N (2007). "QSAR study of anticoccidial activity for diverse chemical compounds: Prediction and experimental assay of trans-2-(2-nitrovinyl) furan." *Bioorg. Med. Chem*. 15: 962-968.

Gumi T., Valiente M., Khulbe K. C., Palet C. et Matsuura T. (2003)." Characterization of activated composite membranes by solute transport, contact angle measurement, AFM and ESR." *Journal of Membrane Science* 212 (1-2): 123-134.

-H-

Hamadache. M (2014). "Modélisation en vue de la prédiction ou la corrélation de l'active toxicologique a partir de la structure moléculaire." *Thèse de Doctorat*, université de Médéa ,Algérie.

Hassoun M. H. (1995). "Fundamentals of artificial neural networks." Mit Press, Massachusetts, USA.

Haykin S. S. (1999). "Neural networks: A comprehensive foundation." Prentice Hall, New Jersey, USA.

Hiemenz P.C. (1986)."Principales of colloid and surface chemistry."Marcel Dekker Inc.

Hong Huixiao, Qian Xie, Weigong Ge, Feng Qian, Hong Fang, Leming Shi, Zhenqiang Su, Roger Perkins et Weida Tong (2008). "Mold2, Molecular Descriptors from 2D Structures for Chemoinformatics and Toxicoinformatics." *Journal of Chemical Information and Modeling* (48): 1337-1344.

Hornik K. (1991). "Approximation capabilities of multilayer feedforward networks." *Neural Networks* 4 (2): 251-257.

Hornik, K., M. Stinchcombe et H. White (1989). "Multilayer feedforward networks are universal approximatbrs". *Neural Networks*, 2, 359-366.

Hu J. Y., Ong S. L., Shan J. H., Kang J. B. et Ng W. J. (2003). "Treatability of organic fractions derived from secondary effluent by reverse osmosis." *Water Research* 37: 4801-4809.

Hunter R.J. (1996)." Introduction to modern colloid science." *Oxford science publications*.

http://www.kochmembrane.com/sep_uf.htm.(consulté le 13/01/2016).

http://www.kochmembrane.com/sep uf.htm.(consulté le 13/01/2016).

-J-

Jacob Matthieu (2011)."Réutilisation des eaux usées épurées par association des procédés biologiques et membranaires."*Thése de Doctorat,* Université de Toulouse.

-K-

Khaouane L. (2013). "Étude et modélisation de la biosynthèse Des antibiotiques a partir de différentes Souches productrices -cas de Pleuromutiline." *Thèse de Doctorat,* Université de Médéa. Médéa, Algérie.

Kiso Y., Nishimura Y., Kitao T. et Nishimura K. (2000)." Rejection properties of non-phenylic presticides with nanofiltration membranes." *Journal of Membrane Science* 171: 229-237.

Kiso Y., Kon T., Kitao T. et Nishimura K. (2001a)." Rejection properties of alkyl phthalates with

nanofiltration membranes." *Journal of Membrane Science* 182 (1-2): 205-214.

Kiso Y., Sugiura Y., Kitao T. et Nishimura K. (2001b). "Effects of hydrophobicity and molecular size on rejection of aromatic pesticides with nanofiltration membranes." *Journal of Membrane Science* 192: 1-10.

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.

Kumar S., Zafar M., Prajapati J. K., Kumar S. et Kannepalli S. (2011). "Modeling studies on simultaneous adsorption of phenol and resorcinol onto granular activated carbon from simulated aqueous solution." *Journal of Hazardous Materials* **185**(1): 287-294.

-L-

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

Libotean D., Giralt J., Rallo R., Cohen Y., Giralta F., Ridgway H. F., Rodriguez G. et Phipps D., (2008). "Organic compounds passage through RO." *Journal of Membrane Science* 313 (1-2): 23–43.

Liew M. K. H., Tanaka S. et Morita M. (1995). "Separation and purification of lactic acid: fundamental studies on the reverse osmosis down-stream process." *Desalination* 101: 269-277.

Lin S.H. et Yang C.R. (2004)."Chimical and physical pretreatement of chemical mechanical polishing wastewater from semiconductor fabrication." *Journal of Hazardous Materials* B108:103-109.

Lu Z. M., Lei J. Y., Xu H. Y., Shi J. S. et Xu Z. H. (2011). "Optimization of fermentation medium for triterpenoid production from *Antrodia camphorata* ATCC 200183 using artificial intelligence-based techniques." *Applied Microbiology and Biotechnology* **92**(2): 371-379.

Lyklema J. (1995)."Fundamentals of interface and colloid science." Vol II, solid-liquid interfaces,Academic Press.

-M-

Maldonado .A (2006). "Diversité Moléculaire : Application au Criblage Virtuel, Corrélation avec des Propriétés Physico-chimiques". *Thèse de Doctorat*, université Paris 7. Denis Diderot, France.

Maurel A. (1998). "Osmose inverse in: Les séparations par membrane dans les procédés de l'industrie alimentaire." Daufin G., René F., Aimar P. Paris, La voisier - Tec & Doc: 68-85.

Metaiche M. (2006). "Optimisation des systèmes de dessalement par osmose inverse : conception, paramètres de fonctionnement et simulation numérique". *Thèse de doctorat*, École nationale polytechnique. Alger. Algérie.

Metaiche Mehdi (2014)."Technologie membranaire." *Cours*,université de Bouira, Algérie.

-N-

Nascimento C. A. O., Giudici R. et Guardani R. (2000). "Neural network based approach for optimization of industrial chemical processes." *Computers & Chemical Engineering* 24 (9-10): 2303-2314.

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: 45-52.

-O-

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

-P-

Padilla A.P. et Tavani E.L. (1999)."Treatement of an industrial effluent by reverse osmosis." *Desalination* 129(1-3),219-226.

Petrinic I., Korenak J., Povodnik D. et Hélix-Nielsen C. (2015)."A feasibility study of ultrafiltration/reverse osmosis (UF/RO)-based wastewater treatment and reuse in the metal finishing industry." *Journal of Cleaner Production* **101**: 292-300.

-R-

Rivals I. (1996) " Les réseaux de neurones formels pour le pilotage de robots mobiles". *Revue de l'association amicale des ingénieurs SUPÉLEC* No178, septembre-octobre.

Roorda J. H., (2004). "Filtration characteristics in dead end ultrafiltration of wwtp-effluent". *Thèse de doctorat*, Université Delft. Delft. Netherlands.

-S-

Sadmani A. H. M. A., Andrews R. C. et Bagley M. D. (2014a). "Impact of natural water colloids and cations on the rejection of pharmaceutically active and endocrine disrupting compounds by nanofiltaration ." *Journal of Membrane Science* 450 : 272-281.

Sadmani A. H. M. A., Andrews R.C. et Bagley M. D. (2014b)."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 Camille (2008)."Etudes des mecanismes de transfert de molécules organiques en osmose inverse." *Thèse de Doctorat*, Université de Montepellier.

Schäfer A. I., Pihlajamäki A., Fane A. G., Waite T. D. et Nyström M. (2004). "Natural organic matter removal by nanofiltration: effects of solution chemistry on retention of low molar mass acids

versus bulk organic matter." *Journal of Membrane Science* 242: 73-85.

Scotti M.T., Fernandes M.B., Ferreira M.J.P. et Emerenciano V.P. (2007). "Quantitative structure-activity relationship of sesquiterpene lactones with cytotoxic activity". *Bioorg. Med. Chem* 15: 2927-2934.

Selassie C. D. (2003). "History of quantitative Structure-Activity relationships", in *Book* "History of quantitative Structure-Activity relationships", D. J. Abraham, Editor, John Wiley & Sons, Inc: California.

Shahmansouri A. et Christopher B. (2013). "Application of quantitative structure property relationships (QSPRs) to predict the rejection of organic solutes by nanofiltration". *Separation and Purification Technology* 118: 27–638.

Shim Y., Lee H. J., Lee S., Moon S. H. et Cho J. (2002)." Effects of natural organic matter and ion species on membrane surface charge." *Environnemental Science and Technology* 36: 3864- 387.

Si-Moussa C. (2009). "Modélisation des équilibres de phases a pressions élevées par les réseaux de neurones artificiels." *Thèse de Doctorat*, Ecole Nationale Supérieure Polytechnique Alger, Algérie.

Sridadhar et Parasad T.G. (2002)."A combination of mechanistic and emprirical models to predict growth and yield of sunflower as influenced by irrigation and moisture stress." Helia.25:39-50.

Stone M. (1977). "An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion". *Journal of the Royal Statistical Society B* 39(1): 44-47.

-T-

Talbi M. L. (2011). "Analyse et traitement du signal électrocardiographique (ECG)." *Thèse de Doctorat*, Université Mentouri de Constantine. Constantine, Algérie.

Triki Z. (2014). "Études, Analyses et Optimisation de la Consommation Énergétique des Unités de Dessalement pour les Sites Isolés". *Thèse de doctorat*, Université Constantine 1. Constantine. Algérie.

-U-

Urbani D., Roussel-Ragot P., Personnaz L. et Dreyfus G. (1994). "The selection of neural models of nonlinear dynamical systems by statistical tests." *Neural Networks for Signal Processing [1994] IV. Proceedings of the 1994 IEEE Workshop*. 6-8 septembre 1994.

-V-

Van der Bruggen B., Schaep J., Wilms D. et Vandecasteele C. (1999)." Influence of molecular size, polarity and charge on the retention of organic molecules by nanofiltration." *Journal of*

Membrane Science 156: 29-41.

Van der Bruggen B., Braeken L. et Vandecasteele C. (2002b)." Evaluation of parameters describing flux decline in nanofiltration of aqueous solutions containing organic compounds." *Desalination* 147 (1-3): 281-288.

Verliefde A.B.D., (2008). "Rejection of organic micropollutants by high pressure membranes (NF/RO) ".*Water Management Academic Press.*

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

-Y-

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., Verliefde A., Kime T.-U., Sadmani A., Kennedy M. et Amy G. (2009a) "Artificial neural network models based on QSAR for predicting rejection of neutral organic compounds by polyamide nanofiltration and reverse osmosis membranes" *Journal of Membrane Science* 313: 23–43.

Yangali-Quintanilla V., Sadmani A., McConville M., Kennedy M. et Amy G. (2009b)." Rejection of pharmaceutically active compounds and endocrine disrupting compounds by clean and fouled nanofiltration membranes."*Water research* 43: 373-384.

Yangali-Quintanilla V., Anwar S, McConville M., Kennedy M. et Amy G. (2010). "A QSAR model for predicting rejection of emerging contaminants (pharmaceuticals, endocrine disruptors) by nanofiltration membranes" *water research 44*: 373–384.

Yangali Quintanilla V. (2010a). "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., 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.

Yangjeh, A. H., et Jenagharad, M. D. (2009). "Application of a genetic algorithm and an artificial neural network for global prediction of the toxicity of phenols to Tetrahymena pyriformis". *Monatsh Chem.* 140: 1279.

Young G. L, Yun S. L, Jong J. J., Sangho L., Dae R. Y., In S. K. et Joon H. K. (2009). "Artificial neural network model for optimizing operation of seawater reverse osmosis desalination plant". *Desalination* 247: 180– 89.

-Z-

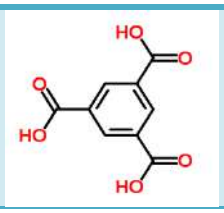
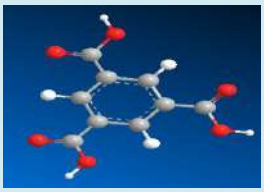
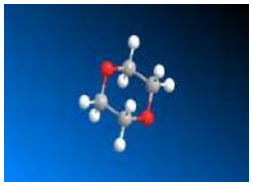
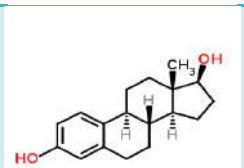
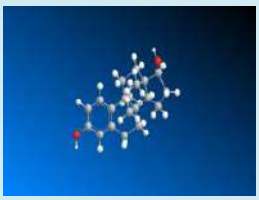
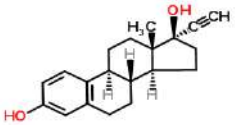
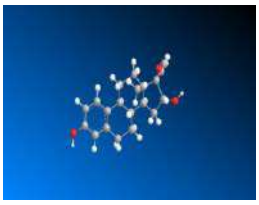
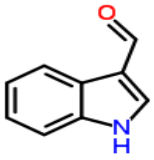
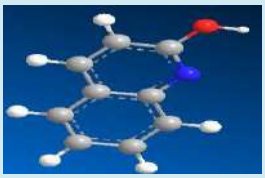
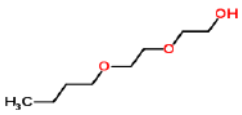
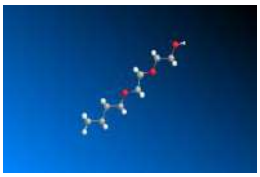
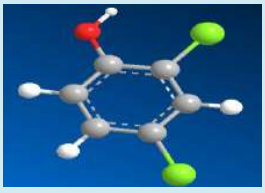
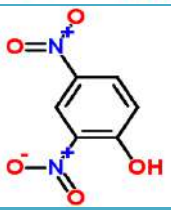
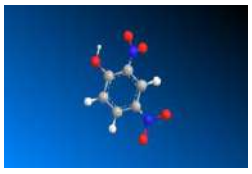
Zafar M., Kumar S., Kumar S. et Dhiman A. K. (2012a). "Artificial intelligence based modeling and optimization of poly(3-hydroxybutyrate-co-3-hydroxyvalerate) production process by using *Azohydromonas lata* MTCC 2311 from cane molasses supplemented with volatile fatty acids: A genetic algorithm paradigm." *Bioresource Technology* **104**(0): 631-641.

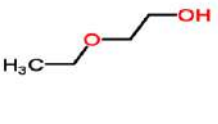
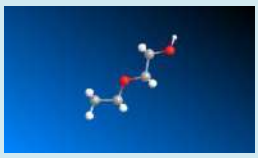
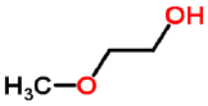
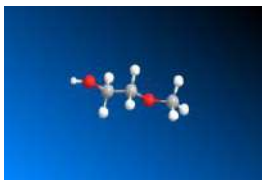
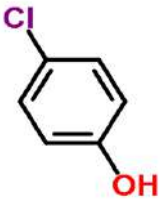
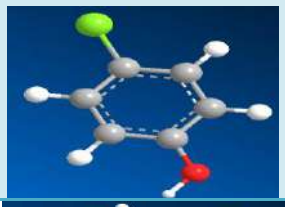
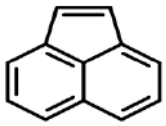
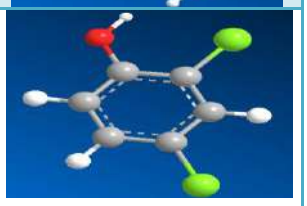
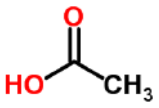
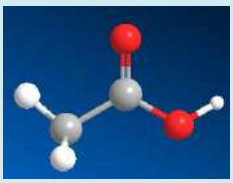
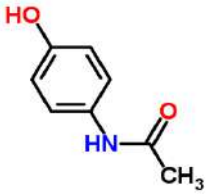
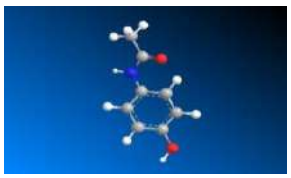
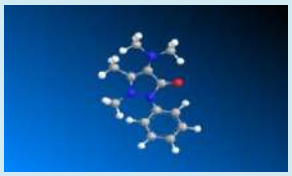
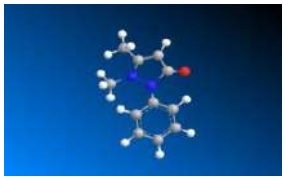
Zafar M., Kumar S., Kumar S. et Dhiman A. K. (2012b). "Optimization of polyhydroxybutyrate (PHB) production by *Azohydromonas lata* MTCC 2311 by using genetic algorithm based on artificial neural network and response surface methodology." *Biocatalysis and Agricultural Biotechnology* **1**(1): 70-79.

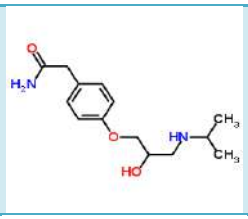
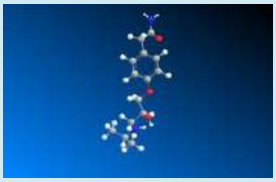
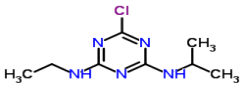
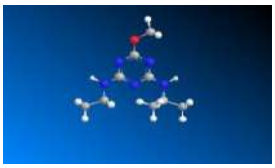
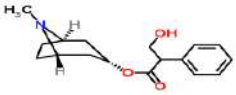
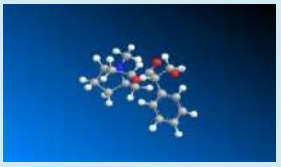
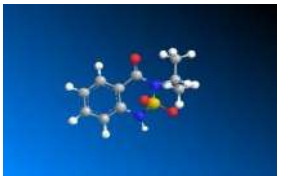
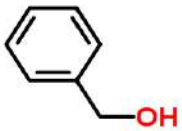
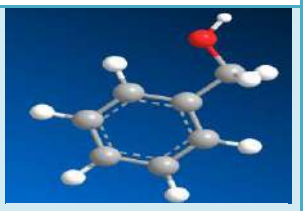
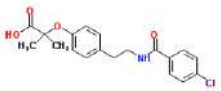
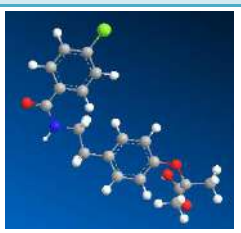
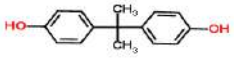
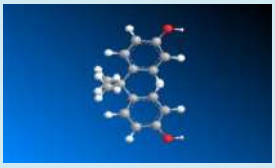
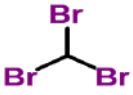
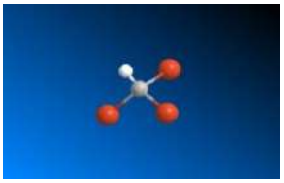
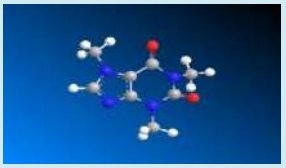
Zhao Y. et Taylor J. S., (2005). "Incorporation of osmotic pressure in an integrated incremental model for predicting RO and NF permeate concentration". *Desalination* 174: 145-159.

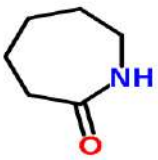
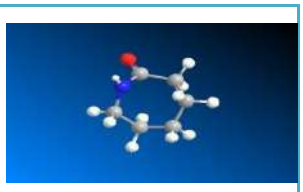
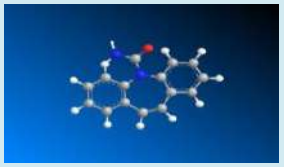
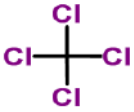
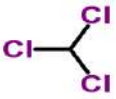
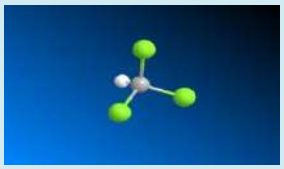
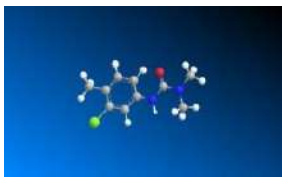
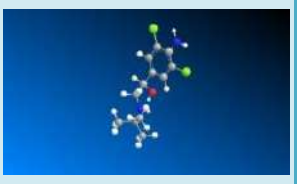
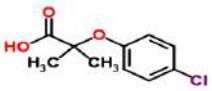
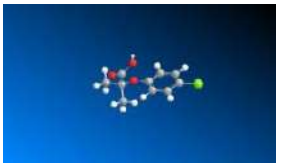
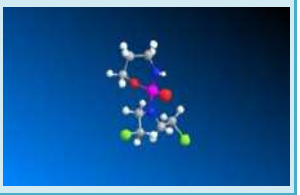
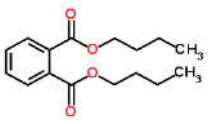
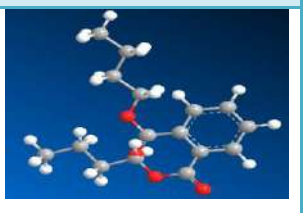
Annexe

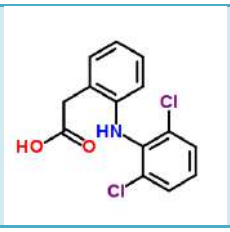
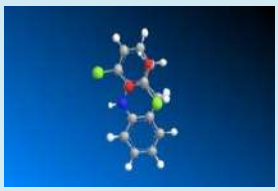
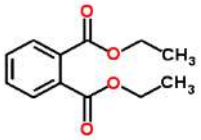
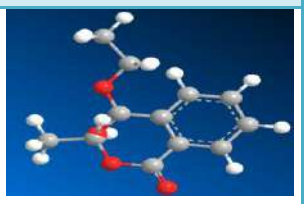
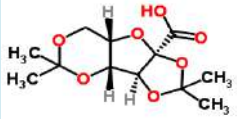
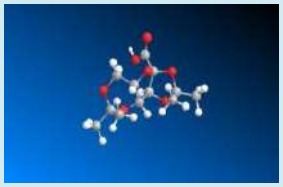
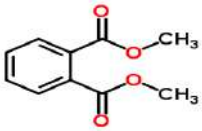
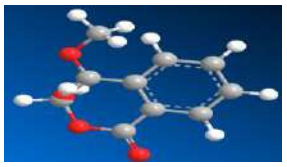
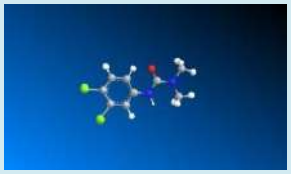
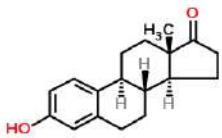
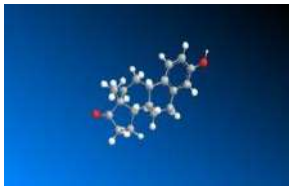
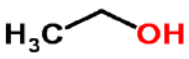
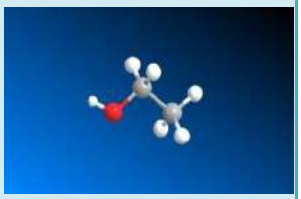
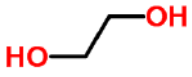
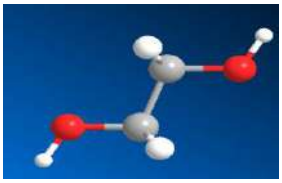
Tableau A. 1. Les molécules organiques et leurs formules et structures.

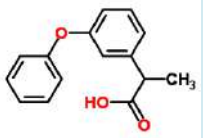
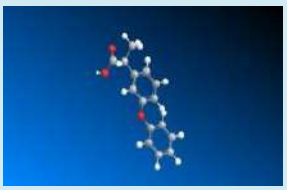
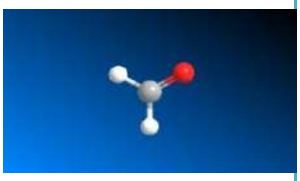
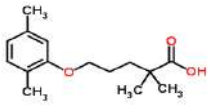
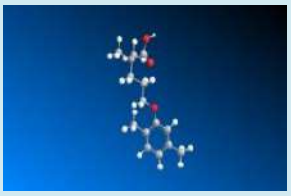
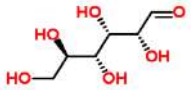
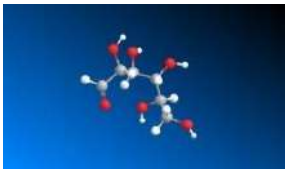
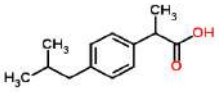
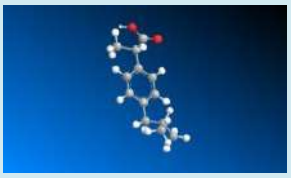
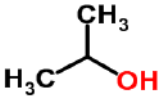
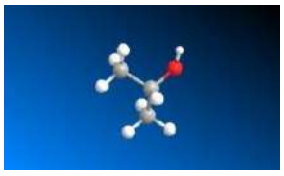
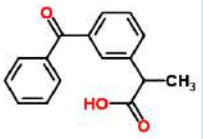
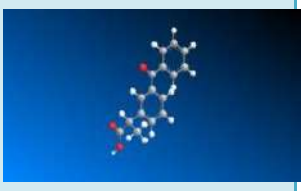
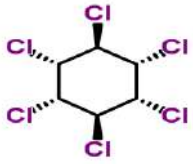
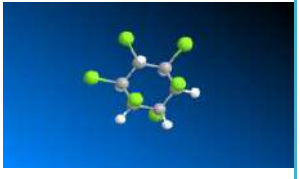
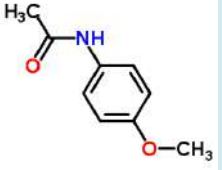
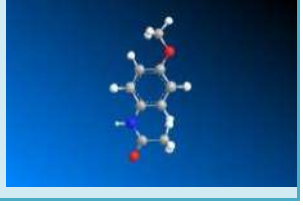
N°	NAME	MW (g/mol)	chemical formula	chemical structure 2D	Chemical structure 3D
1	**1,3,5-Benzenetricarboxylic acid**	210,14	$C_9H_6O_6$		
2	**1,4-Dioxane**	88,105	$C_4H_8O_2$		
3	**17β- Estradiol**	272,382	$C_{18}H_{24}O_2$		
4	**17α- Ethynylestradiol**	296,403	$C_{20}H_{24}O_2$		
5	**2-(1H)-Quinoline**	145,16	C_9H_7NO		
6	**2-(2-Butoxyethoxy)**	162,22	$C_8H_{18}O_3$		
7	**2,4-Dichlorophenol**	163	$C_6H_4Cl_2O$		
8	**2,4-dinitrophenol**	184,106	$C_6H_4N_2O_5$		

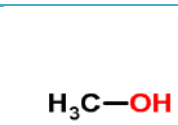
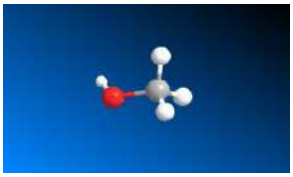
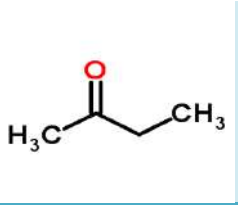
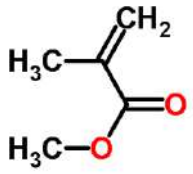
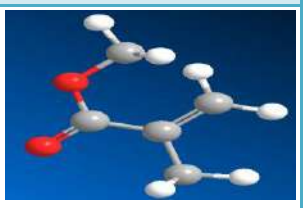
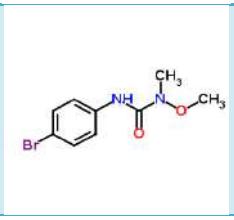
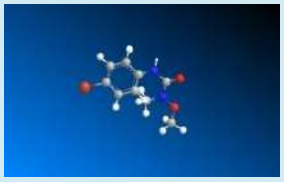
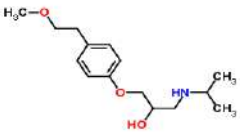
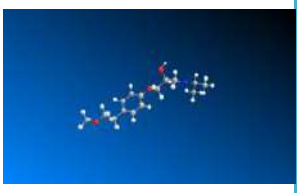
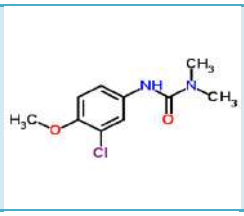
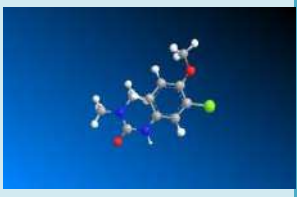
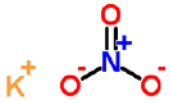
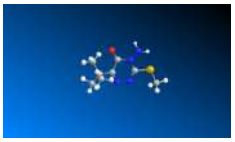
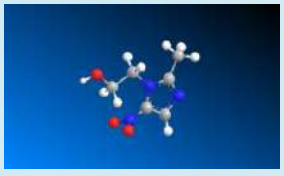
9	**2-ethoxyethanol**	90,12	$C_4H_{10}O_2$		
10	**2-Methoxyethanol**	76,094	$C_3H_8O_2$		
11	**4-Chlorophenol**	128,556	C_6H_5ClO		
12	**Acenaphtylene**	154,208	$C_{12}H_8$		
13	**Acetic acid**	60,052	$C_2H_4O_2$		
14	**Acetaminophen**	151,163	$C_8H_9NO_2$		
15	**Aminopyrine**	231,294	$C_{13}H_{17}N_3O$		
16	**Antipyrine**	188,226	$C_{11}H_{12}N_2O$		

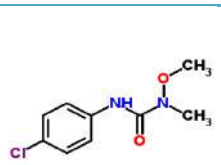
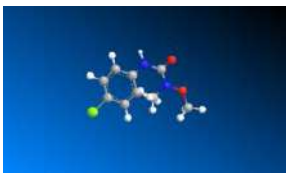
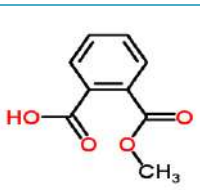
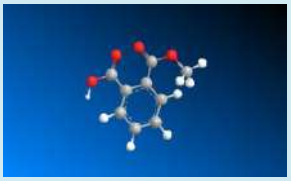
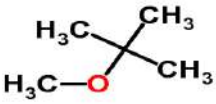
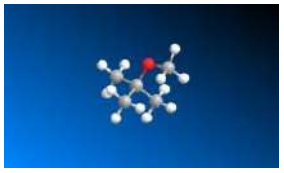
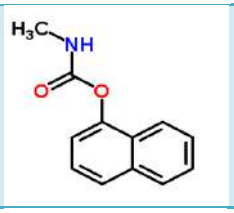
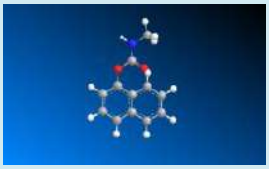
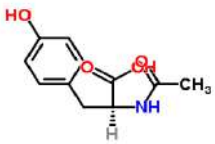
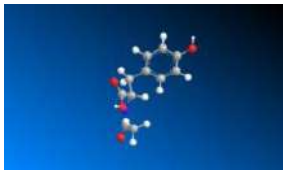
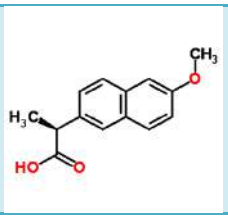
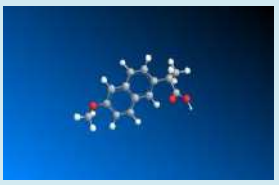
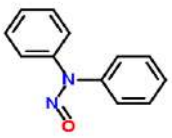
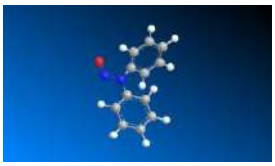
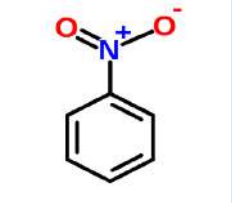
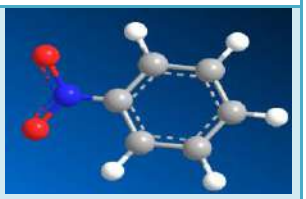
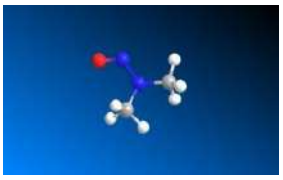
17	**Atenolol**	266,336	$C_{14}H_{22}N_2O_3$		
18	**Atrazine**	215,683	$C_8H_{14}ClN_5$		
19	**Atropine**	289,369	$C_{17}H_{23}NO_3$		
20	**Bentazon**	240,279	$C_{10}H_{12}N_2O_3S$		
21	**Benzyl alcohol**	108,138	C_7H_8O		
22	**Bezafibrate**	361,819	$C_{19}H_{20}ClNO_4$		
23	**Bisphenol A**	228,286	$C_{15}H_{16}O_2$		
24	**Bromoform**	252,731	$CHBr_3$		
25	**Caffeine**	194,191	$C_8H_{10}N_4O_2$		

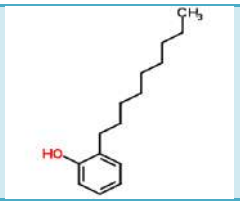
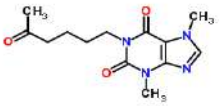
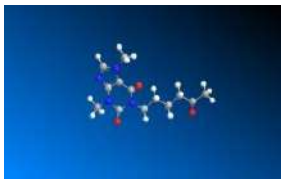
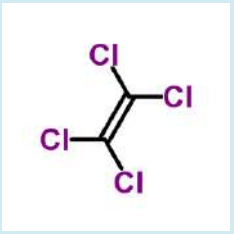
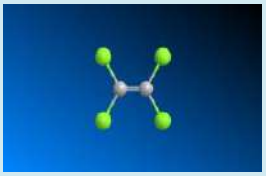
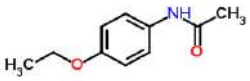
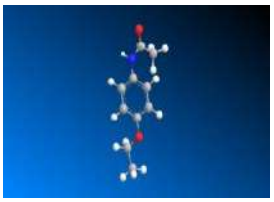
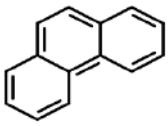
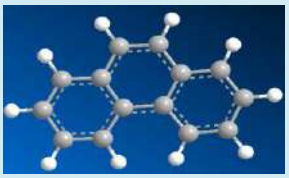
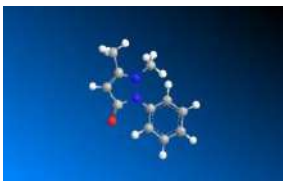
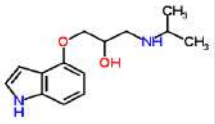
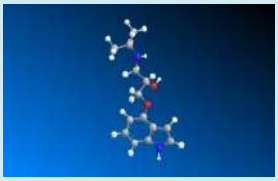
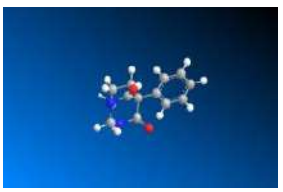
26	**Caprolactam**	113,158	$C_6H_{11}NO$		
27	**Carbamezepine**	236,269	$C_{15}H_{12}N_2O$		
28	**Carbontetrachloride**	153,823	CCl_4		
29	**Chloroform**	119,378	$CHCl_3$		
30	**Chlorotoluron**	212,676	$C_{10}H_{13}ClN_2O$		
31	**Clenbuterol**	277,19	$C_{12}H_{18}Cl_2N_2O$		
32	**Clofibric acid**	214,645	$C_{10}H_{11}ClO_3$		
33	**Cyclophosphamide**	261,08	$C_7H_{15}Cl_2N_2O_2P$		
34	**Dibutyl phtalate**	278,344	$C_{16}H_{22}O_4$		

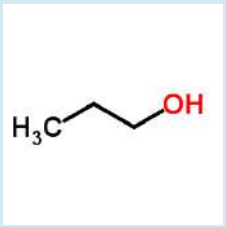
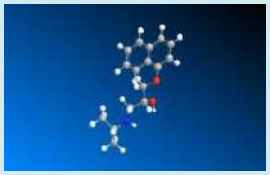
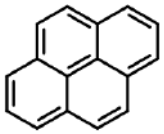
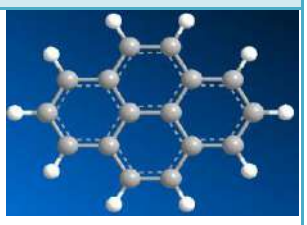
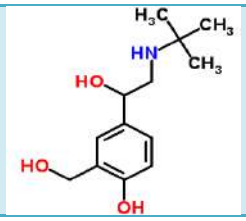
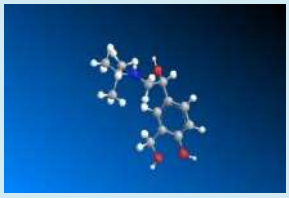
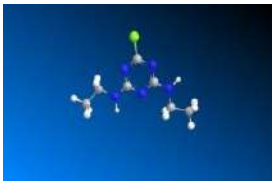
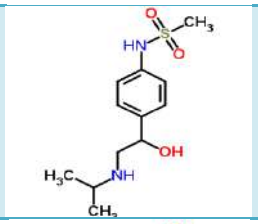
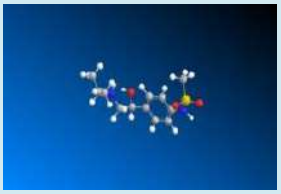
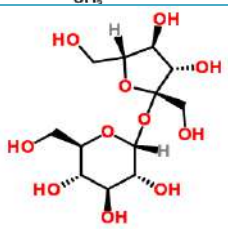
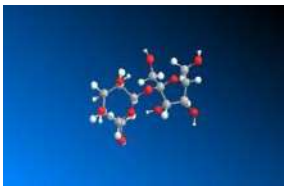
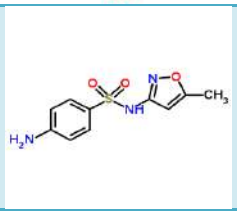
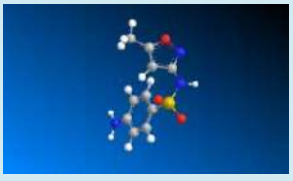
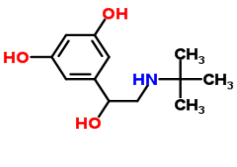
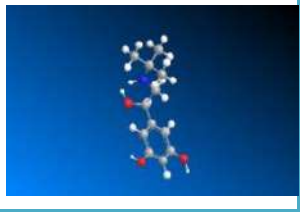
35	**Diclofenac**	296,149	$C_{14}H_{11}Cl_2NO_2$		
36	**Diethyl phtalate**	222,237	$C_{12}H_{14}O_4$		
37	**Dikegulac**	274,267	$C_{12}H_{18}O_7$		
38	**Dimethyl phtalate**	194,184	$C_{10}H_{10}O_4$		
39	**Diuron**	233,094	$C_9H_{10}Cl_2N_2O$		
40	**Estrone**	270.366	$C_{18}H_{22}$		
41	**Ethanol**	46,068	C_2H_6O		
42	**Ethylene glycol**	62,068	$C_2H_6O_2$		

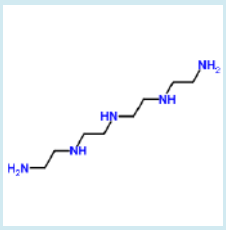
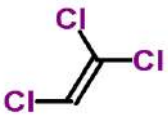
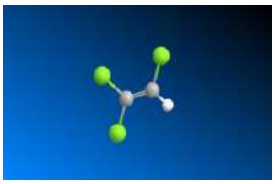
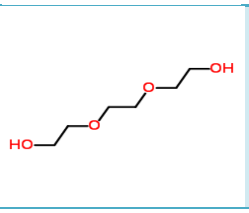
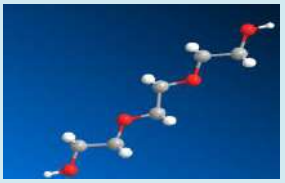
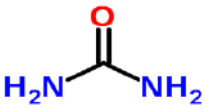
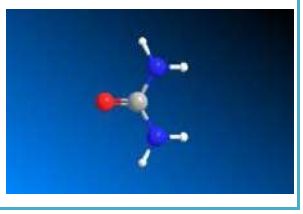
43	**fenoprofen**	242,27	$C_{15}H_{14}O_3$		
44	**Formaldehyde**	30,026	CH_2O		
45	**Gemfibrozil**	250,333	$C_{15}H_{22}O_3$		
46	**Glucose**	180,156	$C_6H_{12}O_6$		
47	**Ibuprofen**	206,281	$C_{13}H_{18}O_2$		
48	**Isopropanol**	60,095	C_3H_8O		
49	**Ketoprofen**	254,281	$C_{16}H_{14}O_3$		
50	**Lindane**	290,83	$C_6H_6Cl_6$		
51	**Methacetin**	165,189	$C_9H_{11}NO_2$		

52	**Methanol**	32,04	CH_4O		
53	**Methylethylketon**	72,106	C_4H_8O		
54	**Methylmetacrylate**	100,116	$C_5H_8O_2$		
55	**Metobromuron**	259,1	$C_9H_{11}BrN_2O_2$		
56	**Metoprolol**	267,364	$C_{15}H_{25}NO_3$		
57	**Metoxuron**	228,675	$C_{10}H_{13}ClN_2O_2$		
58	**Metribuzin**	214,288	$C_8H_{14}N_4OS$		
59	**Metronidazole**	171,159	$C_6H_9N_3O_3$		

60	**Monolinuron**	214,649	$C_9H_{11}ClN_2O_2$		
61	**Monomethylphtalate**	180,157	$C_9H_8O_4$		
62	**Methyl tert-butyl ether**	88,184	$C_5H_{12}O$		
63	**NAC standard(1-naphthyl-n-methylcarbamate)**	201,221	$C_{12}H_{11}NO_2$		
64	**N-acetyl-L-tyrosine**	223,225	$C_{11}H_{13}NO_4$		
65	**Naproxen**	230,259	$C_{14}H_{14}O_3$		
66	**NDPA (diphenylnitrosamine)**	198,221	$C_{12}H_{10}N_2O$		
67	**Nitrobenzene**	123,109	$C_6H_5NO_2$		
68	**N-Nitrosodimethylamine**	74,082	$C_2H_6N_2O$		

69	**Nonylphenol**	220,35	$C_{15}H_{24}O$		
70	**Pentoxifylline**	278,307	$C_{13}H_{18}N_4O_3$		
71	**Perchloroethene**	165,833	C_2Cl_4		
72	**Phenacetine**	179,216	$C_{10}H_{13}NO_2$		
73	**Phenatherene**	178,229	$C_{14}H_{10}$		
74	**Phenazone**	188,226	$C_{11}H_{12}N_2O$		
75	**Pindolol**	248,321	$C_{14}H_{20}N_2O_2$		
76	**Primidone**	218,252	$C_{12}H_{14}N_2O_2$		

77	**propanol**	60,095	$C_{16}H_{21}NO_2$		
78	**pyrene**	202,251	$C_{16}H_{10}$		
79	**Salbutamol**	239,311	$C_{13}H_{21}NO_3$		
80	**Simazin**	201,657	$C_7H_{12}ClN_5$		
81	**Sotalol**	272,364	$C_{12}H_{20}N_2O_3S$		
82	**Sucrose**	342,297	$C_{12}H_{22}O_{11}$		
83	**Sulfamethoxazole**	253,278	$C_{10}H_{11}N_3O_3S$		
84	**Terbutaline**	225,284	$C_{12}H_{19}NO_3$		

85	**Tetraethylenepentamine**	189,302	$C_8H_{23}N_5$		
86	**Trichloroethene**	131,388	C_2HCl_3		
87	**Triethylene glycol**	150,173	$C_6H_{14}O_4$		
88	**Urea**	60,055	CH_4N_2O		

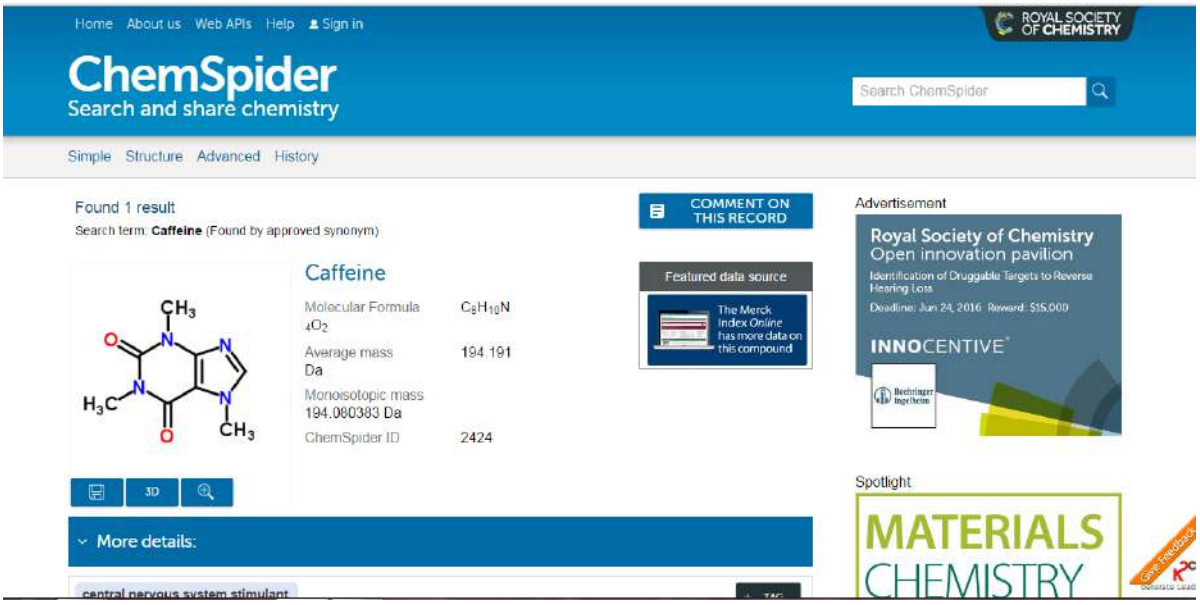


Figure B.1. Capture d'écran de logiciel Chemspider.

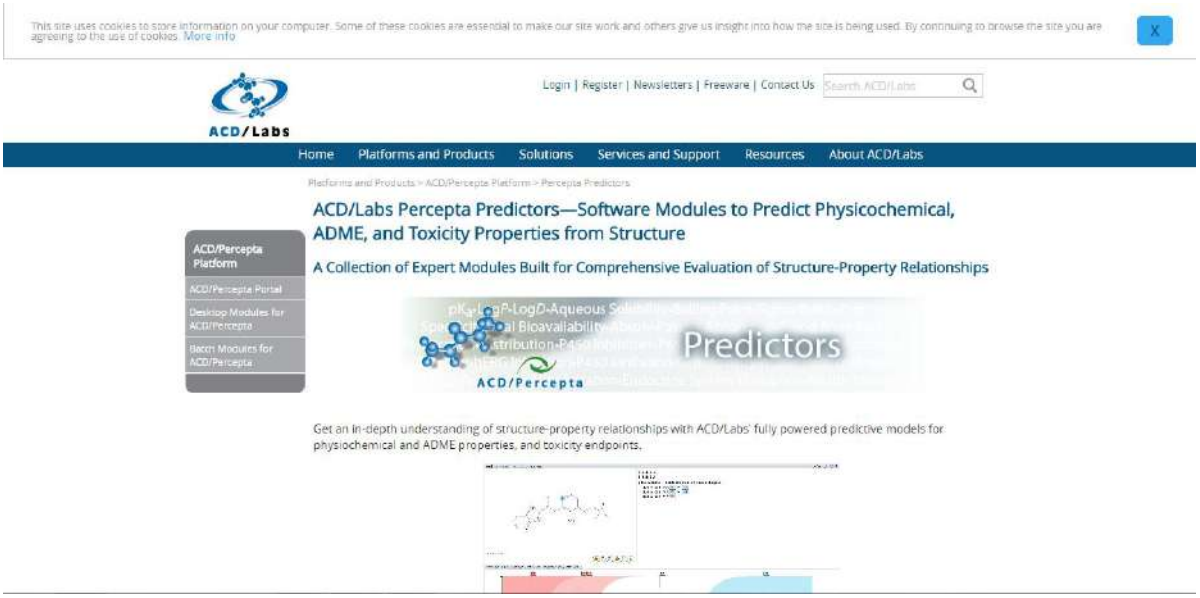


Figure B.2. Capture d'écran de logiciel ACD/Labs.

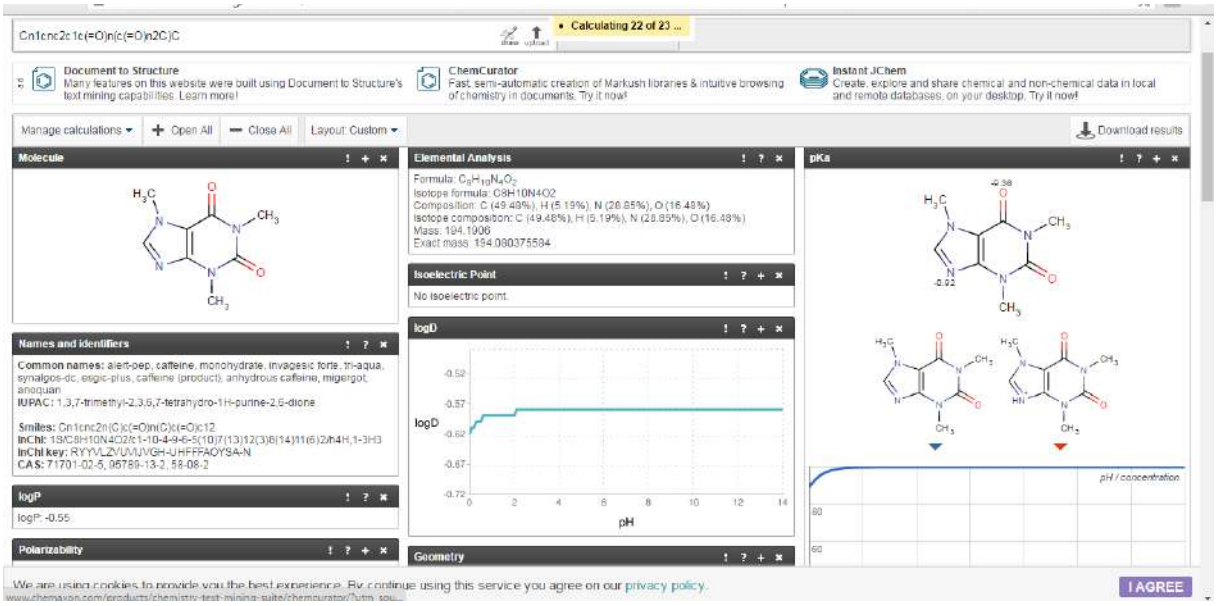
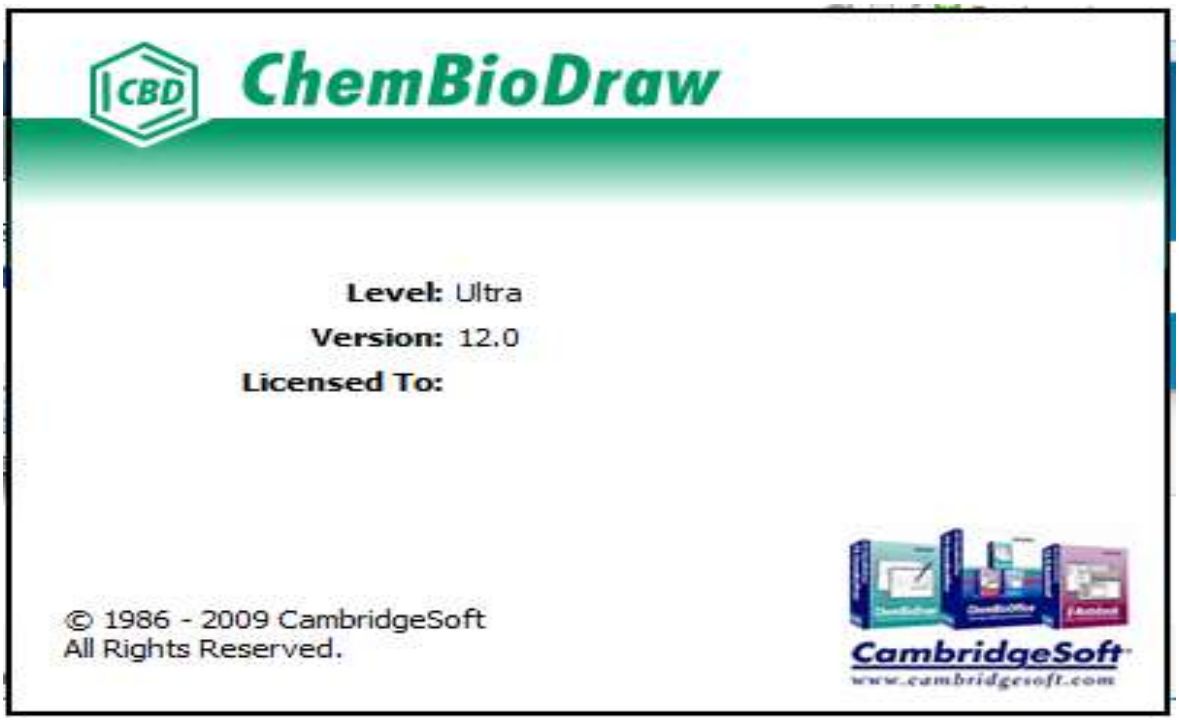
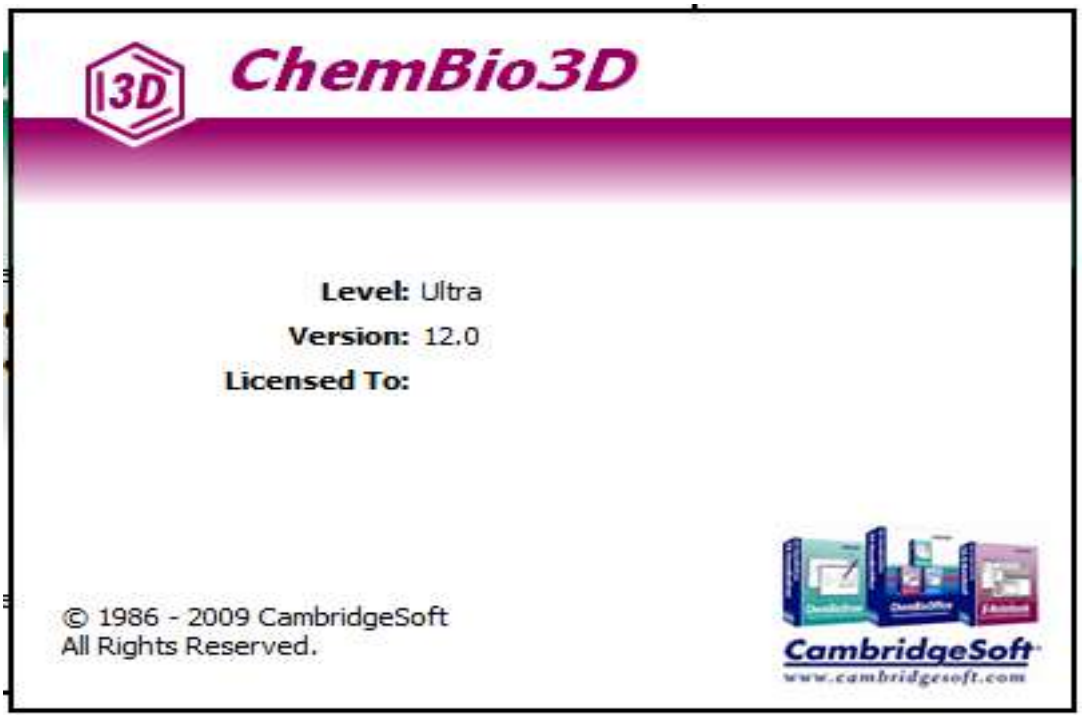


Figure B.3. Capture d'écran de logiciel ChemAxon.



(a)



(b)

Figure B.4. Capture d'écran des logiciels : (a) ChemBioDraw et (b) ChemBio3D.

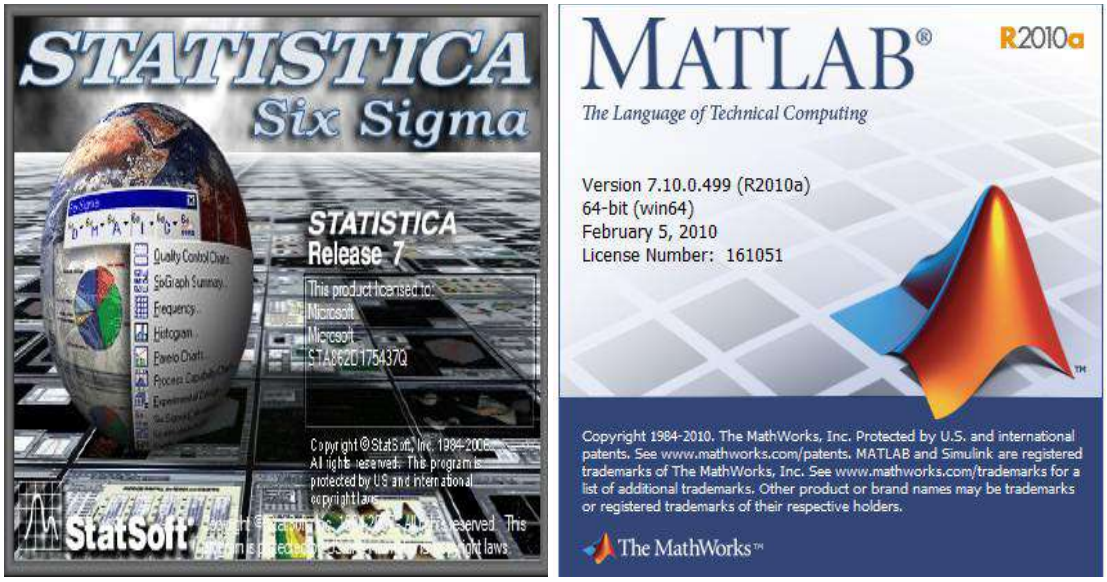


Figure B.5. Capture d'écran des logiciels STATISTICA et MATLAB.

Tableau C.1. Matrice de corrélations.

	MW(g/mol)	log Kow	log D	Moment dipolaire(de)	length(nm)	width(nm)	depth(nm)	Eqwidth(nm)	Zeta potentiel(mv)	Contact angle	Ph	Pression (K pa)	Recovery	T(°C)	MWCO(Da)	NaCl	SR	rejection
***MW* (g/mol)**	1	0,40855854	0,10755001	0,22765999	0,67540426	0,86509103	0,89869893	0,90342144	-0,08372257	-0,09687899	0,05242094	-0,31038863	0,2163989	-0,29681672	0,30393798	-0,34030922	-0,28478228	0,64270938
log Kow	0,40855854	1	0,73239716	-0,24524133	0,28778966	0,30406984	0,39853955	0,35541934	-0,1896853	-0,05337054	0,06473935	-0,06832851	-0,05093485	-0,10768027	0,07378993	0,06032364	-0,14762301	0,28647815
log D	0,10755001	0,73239716	1	-0,2357397	0,01517356	0,05134254	0,09307093	0,09000652	-0,24872022	-0,03812181	0,15386459	-0,00068845	-0,33469957	0,06431689	-0,05417913	0,29584681	-0,15313291	0,00049239
Moment dipolaire (de)	0,22765999	-0,2452413	-0,2357397	1	0,08697147	0,33110586	0,17023745	0,25762488	-0,10558232	0,02671611	0,06475981	-0,01934073	-0,05144887	-0,16618461	0,17529166	-0,13385571	-0,22217128	0,17899837
***length*(nm)**	0,67540426	0,28778966	0,01517356	0,08697147	1	0,62954053	0,80749678	0,75659009	-0,11490785	-0,07908054	0,04320865	-0,25905335	0,24316544	-0,36991241	0,36551655	-0,29653816	-0,27626535	0,59920398
***width*(nm)**	0,86509103	0,30406984	0,05134254	0,33110586	0,62954053	1	0,85677789	0,92460169	-0,15370894	-0,08761711	0,03059478	-0,28667094	0,16515668	-0,34870831	0,33009498	-0,29094946	-0,3335269	0,63746216
***depth*(nm)**	0,89869893	0,39853955	0,09307093	0,17023745	0,80749678	0,85677789	1	0,96608577	-0,15306366	-0,12549287	0,0468547	-0,27697627	0,21323295	-0,37164964	0,3765344	-0,34619472	-0,34931599	0,66719323
Eqwidth(nm)	0,90342144	0,35541934	0,09000652	0,25762488	0,75659009	0,92460169	0,96608577	1	-0,18098975	-0,11441601	0,04749724	-0,29806332	0,17321852	-0,37929565	0,37342756	-0,34255961	-0,38431644	0,66999457
Zeta potentiel (mv)	-0,08372257	-0,1896853	-0,24872022	-0,10558232	-0,11490785	-0,1537089	-0,1530637	-0,18098975	1	-0,48749604	-0,08139536	0,04214803	0,28034515	0,16204341	0,13359654	-0,26361207	-0,00757899	-0,14109297
Contact angle	-0,09687899	-0,0533705	-0,03812181	0,02671611	-0,07908054	-0,0876171	-0,1254929	-0,11441601	-0,48749604	1	-0,08758364	-0,0796437	-0,29560559	0,15802658	-0,58093682	0,22708467	0,42858641	0,05420376
pH	0,05242094	0,06473935	0,15386459	0,06475981	0,04320865	0,03059478	0,0468547	0,04749724	-0,08139536	-0,08758364	1	0,06224905	-0,07926952	0,00841704	0,0530921	0,04681275	-0,07528928	0,03672831
Pression (K pa)	-0,31038863	-0,0683285	-0,00068845	-0,01934073	-0,25905335	-0,2866709	-0,2769763	-0,29806332	0,04214803	-0,0796437	0,06224905	1	-0,32699116	0,43006553	-0,10970887	0,1344614	0,00118655	-0,19960369
Recovery	0,2163989	-0,0509349	-0,33469957	-0,05144887	0,24316544	0,16515668	0,21323295	0,17321852	0,28034515	-0,29560559	-0,07926952	-0,32699116	1	-0,53527814	0,42144504	-0,27335547	0,25383293	0,15061036
***T* (°C)**	-0,29681672	-0,1076803	0,06431689	-0,16618461	-0,36991241	-0,3487083	-0,3716496	-0,37929565	0,16204341	0,15802658	-0,00841704	0,43006553	-0,53527814	1	-0,50522608	0,07083182	0,08196664	-0,25940424
***MWCO*(Da)**	0,30393798	0,07378993	-0,05417913	0,17529166	0,36551655	0,33009498	0,3765344	0,37342756	0,13359654	-0,58093682	0,0530921	-0,10970887	0,42144504	-0,50522608	1	-0,37068159	-0,63017968	0,1147761
NaCl	-0,34030922	0,06032364	0,29584681	-0,13385571	-0,29653816	-0,2909495	-0,3461947	-0,34255961	-0,26361207	0,22708467	0,04681275	0,1344614	-0,27335547	0,07083182	-0,37068159	1	0,51403529	-0,16925978
SR	-0,28478228	-0,147623	-0,15313291	-0,22217128	-0,27626535	-0,3335269	-0,349316	-0,38431644	-0,00757899	0,42858641	-0,07528928	0,00118655	0,25383293	0,08196664	-0,63017968	0,51403529	1	-0,08511836
rejection	0,64270938	0,28647815	0,00049239	0,17899837	0,59920398	0,63746216	0,66719323	0,66999457	-0,14109297	0,05420376	0,03672831	-0,19960369	0,15061036	-0,25940424	0,1147761	-0,16925978	-0,08511836	1