République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université YAHIA FARES DE MEDEA

Faculté des Sciences et de la Technologie
Département de Génie des Procédés

Mémoire de Fin d'Etudes pour l'Obtention du Diplôme de Master en Génie des Procédés
Option : **Pollution Chimique et Gestion Environnementale**

Thème :

Prédiction de la toxicité des insecticides par les réseaux de neurones artificiels au moyen des descripteurs sélectionnés par la méthode "stepwise MLR"

Proposé par :
Mr. HAMADACHE Mabrouk

Dirigé par :
M^elle^ KHAOUANE Latifa

Etudié par :
- *M^elle^ EULDJI Fouzia*
- *M^elle^ BELAISSA Besma*

Promotion : 2012-2013

Remerciements

Nous remercions tout

D'abord, ALLAH le tout-puissant, de nous avoir donné le courage et la volonté pour terminer ce travail.

Nous tenons à exprimer nos plus vifs remerciements au professeur Mr. Salah HANINI et notre prompteur Mr. Mabrouk HAMADACHE, enseignants à l'UYFM, pour leur aide et leur accompagnement. Nous leur devons les remerciements les plus sincères.

Nous exprimons toute notre gratitude à notre co-promotrice Melle «KHAOUANE Latifa »pour l'effort fourni, les conseils Prodigués, sa patience et sa persévérance dans le suivi de ce travail.

Nous remercions Mr. SI MOUSSA Cherif et Mr. Abdallah Abdallah El hadj qui n'ont pas cessé de nous encourager et de nous conseiller tout au long de la préparation de ce mémoire.

Nous adressons également nos vifs remerciements au Mr. Mohamed NEDJHIOUI chef de département du Génie de procédés et environnement.

Nous remercions également tous les enseignants de l'UYFM qui nous ont guidés tout au long de notre vie universitaire.

Dédicace

Je dédie ce modeste travail à :

Mes chers parents qui m'ont soutenu et encouragés durant la période de mes études et à qui Je dois tout ce que je suis aujourd'hui.

Mes très chères sœurs : Bahia, Sabira, Cherifa
A mes frères : Mohamed, Abdenour, Amine.
Mes adorables neveux et nièces

A ma Binôme Belaissa Besma avec qui j'ai passé Les meilleurs moments.

A mes copines : Selma, Souhila, Hafida, Amina, Merieme, fella, Amel, Hamida, Naima, Sammy qui m'ont aidé tout au long de ce travail.

A tous mes collègues de travail de laboratoire des analyses médicales de la polyclinique de Draa-smar pour leur patience et leur soutien.

A toute la promotion Master
Pollution chimique et gestion environnementale
LMD
«2012/2013»

Fouzia

Dédicace

Je dédie ce mémoire à ma très chère et douce mère B. Fatima Zahra et mon cher père Ahmed qu'ALLAH vous protège et ne me prive jamais de vous.

Á ma seconde mère L. Radia et mon second père Djamel qui m'ont aidé par leur amour et leur encouragement à réaliser mon travail

Á mes chers frères :M'hamed et Wassim

Á mes sœurs : Latifa et leurs enfants Djamel et Rym sans oublier son mari Nacer Ichrak ;Lilia et Amira .

Á ma chère amie et binôme Euldji Fouzia pour sa patience et son sérieux.

Á toutes mes amies surtout :C. Selma ; M. Souhila ; D. Amel ;K. Hafidha ;A .Meriem ;F.Amina ;B. Fulla ; B .Hamida .

Á la promotion du Master 2012-2013(LMD) du département du Génie des Procédés Pharmaceutiques et Environnement.

Besma

Lecture	Signification	Unité
B	Biais d'un neurone artificiel	(-)
CL50	Concentration Létale 50	[mmole/l]
DJA	Dose Journalière Admissible	[mg/kg/jour]
DL50	Dose Létale 50	[mg/kg]
EAM	Erreur Absolue Moyen	(-)
ELUMO	Energie de l'Orbitale Moléculaire inoccupée plus bas	[Joule]
F	Fisher statistique.	(-)
IGC50	la concentration pour inhiber la croissance de 50%	(mmol/L)
K	taux de réaction constante	(-)
μ	un petit nombre positif qui représente le pas de déplacement en direction du minimum le plus proche	(-)
P	le potentiel post-synaptique du neurone	(-)
pKa	constante d'équilibre	(-)
Q	corrélation croisée validé coefficient (procédure de leave-one-out)	(-)
R	Coefficient de régression	(-)
REQM	Racine d'erreur quadratique Moyenne	[mg/kg]
SavN	Super delocalisability moyenne nucléophile	(-)
Sp	somme des constantes hydrophobes fragmentaire	(-)
W	Poids synaptique des connexions inter neuronales	(-)
X	l'état du neurone de la couche précédente	(-)

Abréviation	*Signification*
2D	Deux dimensions
3D	Trois dimensions
ACD	Comité d'aide au développement
BFGS	Broyden–Fletcher–Goldfarb–Shanno algorithm
DOG	Difference of Gaussians
EAM	Erreur Absolue Moyen
ELUMO	énergie de l'orbitale moléculaire inoccupée plus bas
EPA	Environemental protection agency
FAO	Organisation des Nations unies pour l'alimentation et l'agriculture
IA	Intelligence Artificielle
IUPAC	International Union of pure and Applied Chemistry
MLP	Multi-Layer Perceptron
NOAEL	La dose sans effet toxique observable
OMS	L'Organisation mondiale de la santé
PDB	Protein Data-Bank
PPDB	Pesticide Properties DataBase
ProPred	Prediction of HLA-DR binding
QSAR	Quantitative Structure-Activité Relationship
RBF	Radial Basis Function
REQM	Racine d’erreur quadratique Moyenne
RNA	Réseau de Neurone Artificiel
SDF	Sans domicile fixe
SMILES	Simplified Molecular Input Line Entry Specification
XOR	La fonction OU exclusif

LISTE DES TABLEAUX

Page

Chapitre I

Tableau I.1 Historique de l'évolution des trois plus grandes classes des pesticides des années 1900 à nos jours 5

Tableau I.2 Classification des pesticides selon la dose létale 9

Tableau I.3 Classement toxicologique d'un produit 11

Chapitre II

Tableau II.1 Notation SMILES pour quelques molécules 31

Chapitre IV

Tableau IV.1 Liste des insecticides utilisés dans cette étude 63

Tableau IV.2 Structure des RNSO utilisés dans la modélisation de la toxicité 78

Tableau IV.3 Récapitulatif des erreurs (REQM, ERAM) et des coefficients de régression (R) obtenues. 79

Tableau IV.4 valeurs du coefficient de régression R pour les fongicides, herbicides et Insecticides 81

LISTE DES FIGURES

page

Chapitre I

Figure I.1	Le marché mondial des pesticides par région du monde	10
Figure I.2	marché mondial des pesticides par catégorie de produit	10
Figure I.3	Mécanismes de transport des pesticides dans l'environnement	16
Figure I.4	Processus impliqué dans le devenir des pesticides dans les sols conditionnant leur disponibilité et, par conséquent, leur efficacité phytosanitaire ou la manifestation de leur caractère polluant	18

Chapitre II

Figure II -1	Etude de QSAR et son application	20

Chapitre III

Figure III.1	Schématisation d'un neurone biologique	36
Figure III.2	Neurone formel	37
Figure III.3	Neurone formel modèle étendu	38
Figure III.4	Fonction binaire à seuil	39
Figure III.5	Fonction identité	40
Figure III.6	Fonction linéaire à seuil	40
Figure III.7	Fonction sigmoïde logarithmique	41
Figure III.8	Fonction sigmoïde tangentielle	41
Figure III.9	Fonction Gaussienne	42
Figure III.10	comparaison entre un neurone biologique et un neurone formel	43

Figure III.11	Réseau multicouche non bouclé classique	44
Figure III.12	Principe d'un perceptron	45
Figure III.13	Réseau Radial Basis Function (RBF)	46
Figure III.14	Un réseau de neurone bouclé à deux entrées contenant une boucle entre les neurones (1) et (2).	47
Figure III.15	Modèle de Hopfield	48
Figure III.16	Réseau de type machine de Boltzmann	49
Figure III.17	Architecture d'un modèle de Kohonen	50
Figure III.18	Illustration de l'apprentissage supervisé	51
Figure III.19	Illustration de l'apprentissage non supervisé	51
Figure III.20	Séparation des exemples d'apprentissage par la droite de décision	53
Figure III.21	Sens de propagation du calcul des sorties s_i	55
Figure III.22	Sens de rétro propagation des gradients δ_i.	56

Chapitre IV

Figure IV.1	Méthodologie de calcul de la toxicité par les réseaux de neurones artificiels.	62
Figure IV.2	Procédure d'élaboration de la base de données	68
Figure IV.3	Collection des notations SMILES par ppdb	69
Figure IV.4	Elaboration du fichier texte des molécules sous forme smilles	69
Figure IV.5	Les étapes pour le calcul des descripteurs par E Dragon1.0.	70
Figure IV.6	Pourcentage des descripteurs avant prétraitement.	74
Figure IV.7	Pourcentage des descripteurs après prétraitement	75
Figure IV.8	Pourcentage des descripteurs après la méthode stepwise	76
Figure IV.9	Schéma simplifié de l'architecture du RNSO	78
Figure IV.10	Validation des valeurs calculées par RNSO avec les valeurs expérimentale : (a) test, (b) apprentissage	80

SOMMAIRE

Page

Introduction générale 1

Chapitre I : LES PESTICIDES

I.1. Introduction 3

I.2. Étymologie 3

I.3. Définition 3

I.4. historique 4

I.5. Composition des pesticides 6

I.6. Classification des pesticides 6

I.6.1 Selon leur mode d'action 6

I.6.2 Selon leur structure chimique 8

I.6.3 Selon leur usage 8

I.6.4 Classification des pesticides selon l'organisation mondiale de la santé 8

I.7. Le marché mondial 9

I.8. Toxicité Des Pesticides 10

I.9. Impacts chroniques sur la sante humaine et ampleur du problème 11

I-9-1 L'ampleur du problème 12

I-9-2 Les pesticides à toxicité aigüe 12

I-9-3 Impacts chroniques sur la santé 13

I.10. Effets sur l'environnement 16

I.11. Conclusion 18

Chapitre II : RELATION QUANTITATIVE STRUCTURE-ACTIVITE

II.1. Introduction 19

II.2. Définition 19

II.3. Historique 20

II.4. Application du QSAR 21

II-4-1 En chimie 21

II-4-2 En biologie 22
II-4-3 Analyse de données 23
II-5 La QSAR en toxicologie de l'environnement 24
II-6 La QSAR pour la prédiction de la toxicité 24
II-7 Les systèmes experts pour la prédiction de la toxicité 27
II-8 Calcul des propriétés physicochimiques 28
II-9 Smiles (Simplified Molecular Input Line Entry Specification) 29
II-9-1 Notation SMILES 29
II-9-2 Principes de Smiles et théorie des graphes 29
II-10. Conclusion 32

Chapitre III : LES RESEAUX DE NEURONES ARTIFICIELS

III.1. Introduction 33
III.2. Historique 34
III.3. Neurone biologique 35
III.3.1. Dendrites (pôles récepteurs) 36
III.3.2. Soma 36
III.3.3. Axone (pôle émetteur) 36
III.3.4. Synapses 36
III.4. Neurone formel 37
III.4.1. Entrées 38
III.4.2. Poids synaptique 38
III.4.3. Seuil 38
III.4.4. Fonction d'activation 39
III.4.4.1. Fonction binaire à seuil 39
III.4.4.2. Fonction identité 40
III.4.4.3. Fonction linéaire à seuil 40
III.4.4.4. Fonction non linéaire 41
III.4.4.5. Fonction Gaussienne 42
III.4.5. Sortie du neurone "S" 42
III.5. Étude comparative 43
III.6. Architecture et connexionnisme des réseaux de neurones 43
III.6.1. Réseau de neurones non bouclés 43

III.6.1. 1. Perceptron 45
III.6.1.2. Perceptron multicouche (MLP) 45
III.6.1.3. Réseaux RBF (Radial Basis Function) 46
III.6.2. Réseaux de neurones bouclés 47
III.6.2.1. Modèle de Hopfield 48
III.6.2.2. Modèle de Boltzmann 49
III.6.2.3. Modèle de Kohonen 49
III.7. Apprentissage 50
III.7.1. Apprentissage supervisé 50
III.7.2. Apprentissage non supervisé 51
III.8. Règles d'apprentissage 52
III.8.1. Règle d'apprentissage de Hebb 52
III.8.2. Règle d'apprentissage du perceptron de F. Rosenblatt 53
III.8.3. Autres règle d'apprentissage 54
III.9. Théorie de la généralisation 54
III.10. Critère d'arrêt pour l'apprentissage 56
III.11. Procédure de modélisation neuronale 57
III.11.1. Sélection des données 58
III.11.2. Collecte des données 59
III.11.3. Séparation des bases de données 59
III.11.4. Normalisation des données 59
III.11.5. Choix du type de réseau de neurones 59
III.11.6. Choix du nombre de neurones dans la couche cachée 60
III.11.7. Choix de la fonction de transfert 60
III.11.8. Estimation des paramètres ou apprentissage 60
III.11.9. Validation et généralisation 60
III.12. Conclusion 61

Chapitre IV : MODELISATION NEURONALE DE LA TOXICITE DES INSECTICIDES

IV-1 Introduction 62
IV-2 Méthodologie de calcul 62
IV-2-1 Elaboration de la base de données 63
IV.2.2. Prétraitement de la base de données 75
IV.2.3. Normalisation de la base de données partielle 76

IV.2.4. Mise en forme des données 77
IV.2.5. Conception et optimisation 77
IV.3. Structure de réseau de neurone optimale obtenue (RNSO) 77
IV.3.1. L'algorithme d'apprentissage et la fonction d'activation 78
IV.4. Validation des valeurs calculées par RNA avec les valeurs expérimentales 78
IV.5. Comparaison avec d'autres travaux 81
IV.6. Conclusion 82

Conclusion générale 83

ملخص: التنبؤ العصبي من سمية مبيدات الحشرات

الهدف من هذا العمل هو إقامة علاقات بين البنية والنشاط من اجل تنبؤ سمية 141 مبيد حشرات مع حساب 08 معيار (التي تم الحصول عليها بواسطة الأسلوب التدرجي) تحت شكل ممثلات كيميائية فيزيائية لكل جزيء.

استعملت إستراتيجية المحاولة والخطأ كمعيار توقف من اجل النمذجة العصبونية مع قدرة التعميم، لغرض تحديد المعايير الأكثر تأثيراعلى سمية مبيدات الحشرات. الشبكة العصبونية المثالية المتحصل عليها مكونة من طبقة إدخال، طبقة خفية وطبقة خروج. المعلمات الذي لها أكبر تأثير على سمية مبيدات الحشرات هي 3D-مورس،GETAWAY و المجموعات الوظيفية .

الكلمات المفتاحية: تنبؤ، سمية، ممثلات، علاقات البنية النشاط والكمية، شبكة عصبونية

Résumé : Prédiction de la Toxicité des Insecticides par les Réseaux de Neurones Artificiels

Le but de ce travail est d'établir des relations structure-activité (SAR) en mesure de prédire la toxicité dans un ensemble de données de 141composés de insecticides, et 8 paramètres calculées (obtenues par la méthode stepwise) sous forme de descripteurs QSAR de type physico-chimiques pour chaque molécule.

La stratégie d'essai et erreur a été utilisée pour la modélisation neuronale avec la capacité de généralisation comme critère d'arrêt à fin de déterminer les paramètres les plus pertinents sur la toxicité des insecticides.

Le réseau de neurones optimal obtenu est un perceptron multicouche : une couche d'entrée, une couche cachée et une couche de sortie. Les paramètres les plus influant sur la toxicité des insecticides sont 3D-MoRSE, GETAWAY, les groupes fonctionnels.

Mots clés : Prédiction, Toxicité, Descripteur, QSAR, Réseau de Neurones.

Abstract: Neural prediction of the toxicity of insecticides

The aim of this work is to establish structure-activity relationships (SAR) to predict toxicity in a dataset consisting of 141 insecticides and 08 parameters calculated (obtained by the stepwise method) as QSAR descriptors of physicochemical type for each molecule.

The trial and error strategy was used for neuronal modeling with the generalization ability as stopping criterion for the purpose of determining the most relevant parameters on the toxicity of insecticides.

The optimal neural network obtained is a multilayer perception: an input layer, a hidden layer and an output layer. The parameters have the greatest influence on the toxicity of insecticides 3D-Morse, 3D-MoRSE, GETAWAY and functional groups .

Keywords: Prediction, Toxicity, descriptor, QSAR, Neural Networks.

INTRODUCTION GENERALE

Les pesticides (insecticides, raticides, fongicides, et herbicides) sont des composés chimiques dotés de propriétés toxicologiques, utilisés essentiellement par les agriculteurs pour lutter contre les animaux (insectes, rongeurs) ou les plantes (champignons, mauvaises herbes) considérés nuisibles aux plantations. Par ailleurs et dans une moindre proportion, ils sont utilisés pour un usage domestique.

Malheureusement, malgré le fait que les pesticides soient bénéfiques, leur emploi massif a pour conséquence la contamination généralisée de notre environnement. Les eaux de surfaces et les eaux souterraines sont très fréquemment polluées. Cette contamination généralisée de l'environnement et de la nourriture par les pesticides rend inévitable la contamination de l'être humain. Leurs effets secondaires nocifs ont été rapidement mis en évidence et leur toxicité, liée à leur structure moléculaire, ne se limite pas en effet aux seules espèces que l'on souhaite éliminer.

La tendance actuelle est le recours à l'interdiction de l'utilisation des pesticides reconnus toxiques. Cependant, pour arriver à statuer sur la toxicité ou non d'un pesticide, de longues et harassantes analyses sont effectuées sur des animaux pendant une très longue durée. D'un autre côté, le coût de ces analyses est très élevé.

Pour surmonter ces problèmes, des tentatives ont été faites pour proposer des approches de modélisation permettant la simulation du devenir des insecticides et leurs toxicités vis-à-vis de l'environnement par élaboration de relations entre les structures des molécules et leurs activités (toxicité).

C'est dans ce contexte que s'inscrit ce travail. Il consiste d'une part à l'élaboration d'une base de données basée sur la caractérisation des molécules par un ensemble de descripteurs (nombre réels mesurés ou calculés à partir des structures moléculaires) et d'autre part à la modélisation de la prédiction de l'effet toxicologique à partir de ces descripteurs en utilisant les réseaux de neurones artificiels de type perceptron multicouche.

Cette technique, doit passer non seulement par le calcul des différents descripteurs, mais aussi par la sélection des descripteurs les plus pertinents par rapport à la grandeur à modéliser.

Ce manuscrit s'articule autour de quatre chapitres :

- Le chapitre I présente des généralités sur les pesticides (définition, propriétés,...).
- Le chapitre II introduit la notion de QSAR.
- Dans le troisième chapitre on donne un aperçu général sur la modélisation par les réseaux de neurones artificiels.
- Le quatrième chapitre présente la méthodologie ainsi que les résultats obtenu lors de la modélisation de la toxicité des insecticides.

Les perspectives inhérentes aux résultats de ce travail seront évoquées dans la conclusion générale.

CHAPITRE I
GENERALITES SUR LES PESTICIDES

I.1 Introduction

Les pesticides sont des produits chimiques qui servent à lutter contre les ravageurs tels qu'insectes, mauvaises herbes, maladies des plantes, nématodes et rongeurs. L'augmentation de leur emploi depuis 1945 a fortement contribué à accroitre la production agricole, à protéger le bétail contre des maladies telles que la trypanosomiase et l'homme contre le paludisme et la filariose, à réduire les pertes de céréales en cours de stockage, et d'une façon générale à améliorer la condition humaine. La consommation mondiale annuelle de pesticides dépasse probablement 2 millions de tonnes. La plus grande partie est fabriquée dans les pays industrialisés, et la valeur totale de la production annuelle est supérieure à 10 milliards de dollars des Etats-Unis. La consommation mondiale de pesticides augmente et continuera à augmenter jusqu'à la fin du siècle en même temps que se poursuivra la course entre production alimentaire et population. L'emploi des pesticides soulève toutefois des difficultés.(D.A. Lindquist)

I.2 Étymologie

Le mot *pesticide* vient de l'anglais, sur le modèle de nombreux mots se terminant par le suffixe -*cide* (latin -*cida*, du verbe latin *caedo, caedes, caedere, caedi, caedum* : « tuer »), et sur la base du mot anglais *pest* (animal, insecte ou plante nuisible), lequel provient (comme le français *peste*) du latin *pestis* qui désigne notamment un animal nuisible (www.wikipedia.org/wiki/Pesticide).

I.3 Définition

Selon le Code de conduite de la FAO sur la distribution et l'utilisation des pesticides, « un pesticides est une substance ou association de substances qui est destinée à repousser, détruire ou combattre les ravageurs, y compris les vecteurs de maladies humaines et animales, et les espèces indésirables de plantes ou d'animaux ».

Les ravageurs (déprédateurs ou nuisibles) causent des dommages durant la production, la transformation, le stockage, le transport et la commercialisation des produits. Les produits à protéger peuvent être des plantes ou parties de plantes, des denrées alimentaires, des produits agricoles, du bois et des produits ligneux et des aliments pour animaux **(**www. environnement. Dangers-risques-pesticides.pdf).

Selon les nuisibles ciblés, on parlera de :

- Insecticide (utilisé contre les insectes) ;
- Fongicide (utilisé contre les champignons) ;
- Bactéricide (utilisé contre les bactéries) ;
- Raticide (utilisé contre les rats et souris) ;
- Herbicide (utilisé contre les mauvaises herbes) ;
- Avicide (utilisé contre les oiseaux granivores);
- Nématicide (utilisé contre les nématodes) ;
- Molluscide (utilisé contre les mollusques), etc.

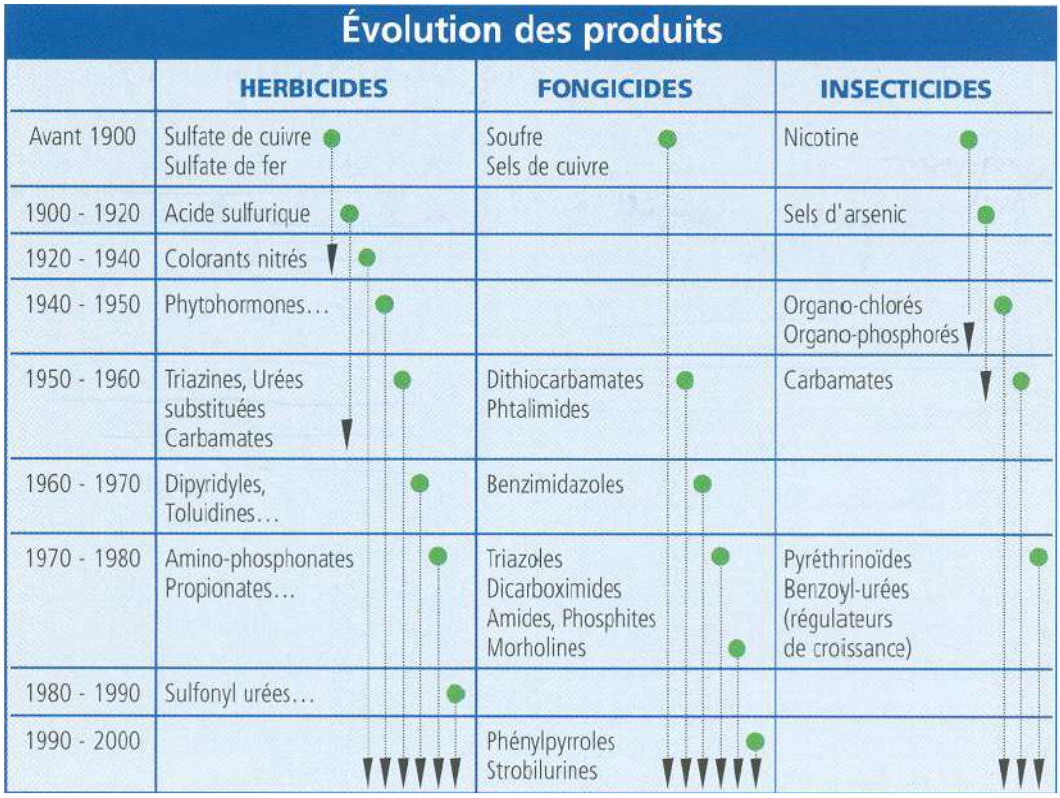
I.4 Historique

La première utilisation des pesticides en agriculture date de l'antiquité. Leur développement a ensuite suivi celui de la chimie minérale. Les composés alors employés étaient des dérivés de composés minéraux ou des plantes comme par exemple. Ceux à base d'arsenic, de cuivre, de zinc, de manganèse ou de sulfate du développement de la chimie organique. Les composés synthétiques majoritaires ont d'ailleurs été à l'origine de l'expansion rapide des produits phytosanitaires à partir des années 1940 (Tableau I.1).

Une définition détaillée de ces composés est disponible dans la directive européenne 91/414/CE du 15 juillet 1991 qui vise à assurer la sureté et qualité des produits de consommation (Directive européenne du conseil du 15 juillet 1991(Journal officiel))**.** De manière plus concise, les pesticides ou produits phytosanitaires, sont définis comme des substances dont les propriétés chimiques contribuent à la protection des plantes cultivées et des produits récoltés. Ce sont des formulations contenant une ou plusieurs substances chimiques minérales ou organiques, synthétiques ou naturelles. La plupart de celle utilisée aujourd'hui est de nature organique dont un petit nombre est extrait ou dérivé des plantes. Les formulations sont en général composées d'une ou de plusieurs substances actives et d'un ou de plusieurs adjuvant .la substances active exerce une

action générale ou spécifique sur les organismes nuisibles ou sur les végétaux ; c'est elle qui confère au produit l'effet désiré. L'adjuvant quant à lui est une substance dépourvue d'activité biologique jugée suffisante dans la pratique mais capable de modifier les propriétés physiques, chimique ou biologique des produits phytosanitaires. Il renforce l'efficacité, la sécurité du produit et sa facilité d'utilisation (Association De Coordination Technique Agricole. (2006) - "Index phytosanitaire ACTA)**.**

Tableau I.1. Historique de l'évolution des trois plus grandes classes des pesticides des années 1900 à nos jours**,(** www. senat, fr/rap/102-215-2/102-215-21pdf).



Les produits phytosanitaires permettent la régulation de la croissance des plantes et la conservation des récoltes. Il améliore ainsi à la fois la quantité et la qualité des denrées alimentaires. Dans le passé, en l'absence de traitement adapté, des récoltes et des cultures entières ont été anéanties entrainant des crises alimentaires (famine). Ainsi, ils aident au contrôle d'un grand nombre de maladies et de parasites de plantes (Garrido-Frenche A et al. 2004)**.** l'usage non agricole des pesticides concerne l'entretien des espaces publics ,des infrastructures de transport et des jardins des particuliers .En France , en 2004 , sur les 76000 tonnes commercialisées, 90 à 94% ont été destinées à l'agriculture.

Le reste se partage équitablement entre les usages amateurs et les usages collectifs **(**www. observatoire pesticides. gouv. Fr). Dans de nombreux domaines, le terme «résidus de pesticides » est largement employé ; il correspond aussi bien au composé qu'à ses produits de dégradation qui peuvent être présents à la fois dans l'environnement mais aussi dans des produits de consommation.

I.5 Composition des pesticides

Un pesticide est constitué de nombreuses molécules comprenant (www.notre-planete.info/ecologie/alimentation/pesticides) :

- Une (ou plusieurs) substance active responsable de l'effet du pesticide.
- Un solvant. Ce sont souvent des huiles végétales dans le cas des liquides, de l'argile ou du talc dans le cas des solides.
- Des additifs n'ayant pas d'activité biologique, mais pouvant modifier les propriétés du pesticide et faciliter son utilisation. Il y a plus de 100000 pesticides commercialisés, fabriqués à partir de 900 substances actives différentes. Chaque année, entre 15 et 20 nouvelles substances actives sont créées.

I.6 Classification des pesticides

Les pesticides disponibles aujourd'hui sur le marché sont caractérisés par une telle variété de structures chimiques, de groupes fonctionnels et d'activité que leur classification est complexe .d'une manière générale, ils peuvent être classés en fonction de la nature de l'espèce à combattre et en fonction de la nature chimique de la principale substance active qui le compose et aussi en fonction de leur usage .donc les pesticides sont classés :

I.6.1 Selon leur mode d'action

Il existe principalement trois grandes familles chimiques qui sont les herbicides, les fongicides et les insecticides.

- **Les herbicides** représentent les pesticides les plus utilisés dans le monde toutes cultures confondues. Ils sont destinés à éliminer les végétaux rentrant en concurrence avec les

plantes à protéger en ralentissant leur croissance. Les herbicides possèdent différents mode d'action sur les plantes **:**

- les perturbateurs de la régulation d'une hormone « l'auxine » (principale hormone agissant sur l'augmentation de la taille des cellules).
- les perturbateurs de la photosynthèse.
- les inhibiteurs de la division cellulaire.
- les inhibiteurs de la synthèse des lipides.
- les inhibiteurs de la synthèse de cellulose.
- les inhibiteurs de la synthèse d'acide aminés.

- **Les fongicides** permettent quant à eux de combattre la prolifération des maladies des plantes provoquées par des champignons ou encore des bactéries. Les fongicides peuvent agir différemment sur les plantes :
  - les inhibiteurs respiratoires.
  - les inhibiteurs de la division cellulaire.
  - les perturbateurs de la biosynthèse des acides aminés ou des protéines.
  - les perturbateurs du métabolisme des glucides.

- **Les insecticides** sont utilisés pour la protection des plantes contre les insectes, ils interviennent en éliminant ou empêchant leur reproduction. Différent types existent comme :
  - les neurotoxiques.
  - les régulateurs de croissance.
  - ceux agissant sur la respiration cellulaire.

Outre, les trois grandes familles de pesticides mentionnés ci-dessus, différentes familles peuvent être citées comme par exemple :

- les acaricides contre les acariens.
- les nemanticides contre les vers du groupe des nématodes.
- les rodonticides contre les rongeurs.
- les taupicides contre les taupes.
- les molluscicides contre les limaces et escargots.
- les corvicides et corvifuges contre les corbeaux et les autres oiseaux ravageurs de culture.

I.6.2 Selon leur structure chimique

Les principaux groupes chimiques sont :

- les organochlorés.
- les organophosphorés.
- les carbamates.
- les pyrethrynoides.
- les triazines.
- les urées substituées.

I.6.3 Selon leur usage

A. Les pesticides à usage agricole ou produits phytopharmaceutiques qui sont des substances chimiques minérales ou organiques, de synthèse ou naturelles. Elles sont utilisées pour la protection des végétaux contre les maladies et contre les organismes nuisibles aux cultures.
B. Les pesticides à usage non agricole ou biocides qui sont similaires aux premiers, utilisés par exemple en hygiène publique (lutte anti-vectorielle) et dans d'autres applications comme la conservation du bois, la désinfection, ou certains usages domestiques.

Environ 300 à 350 matières actives sont autorisées au Maroc et entrent dans la composition de près de 1000 spécialités commerciales, dont plus de 80% homologuées pour utilisation agricole **(**The WHO Recommended Classification of Pesticides by Hazard. 2004).

I.6.4 Classification des pesticides selon l'organisation mondiale de la santé

L'Organisation mondiale de la santé (OMS) classe les pesticides par dangerosité en se basant sur leur dose létale médiane orale ou cutanée...(www.fao.orgagAGP...printfriendly.html).
Chaque pesticide est alors placé dans une des quatre classes:

- Ia, extrêmement dangereux;
- Ib, très dangereux;
- II, modérément dangereux;
- III, légèrement dangereux.

Tableau I.2. Classification des pesticides selon la dose létale,
(www.wikipedia.orgwikiToxicity_Class.08)

La classe de toxicité	La dose létale
extrêmement dangereux	< 5 mg/kg
très dangereux	5 - 50 mg/kg
modérément dangereux	50 - 500 mg/kg
légèrement dangereux	> 500 mg/kg

I.7 Le marché mondial

Le marché mondial (environ 40 milliards de dollars) est globalement stable depuis quelques années (2000). Il faut noter que certains évènements climatiques récents (chaleur et sècheresse en Europe, pluie en Océanie) influencent fortement ces chiffres.

En Europe et en Amérique du Nord, les herbicides représentent 70 à 80% des produits utilisés (notamment à cause de la forte augmentation des cultures de maïs) tandis que sous les tropiques, 50% des produits appliqués sont des insecticides. La diversification des cultures, avec l'amélioration du niveau de vie dans certains pays, modifie également cet équilibre, ainsi la Chine a converti l'équivalent de la surface de l'Angleterre de rizières en cultures maraichères, entrainant une diversification des produits mis en œuvre.

Les usages non-agricoles représentent environs 12% du marché global (dont plus d'un tiers des usages pour les Etats-Unis). En France, les usages non agricoles représentent entre 6 et 10% de la consommation totale.

Les pesticides et les engrais ont permis, depuis les années 60, de multiplier la productivité par 3, les pertes occasionnées aux cultures représentent pourtant encore près de 30% en Europe pour le maïs contre 50% en Afrique ; moins de 30% en Asie pour le riz contre plus de 50% en Afrique. Malgré les efforts considérables réalisés, il reste encore dans le Monde près d'un milliard de personnes mal nourries, et il faudra d'ici à 2050 nourrir une population qui aura progressée de près de 50% (Fergani et Mami. 2011).

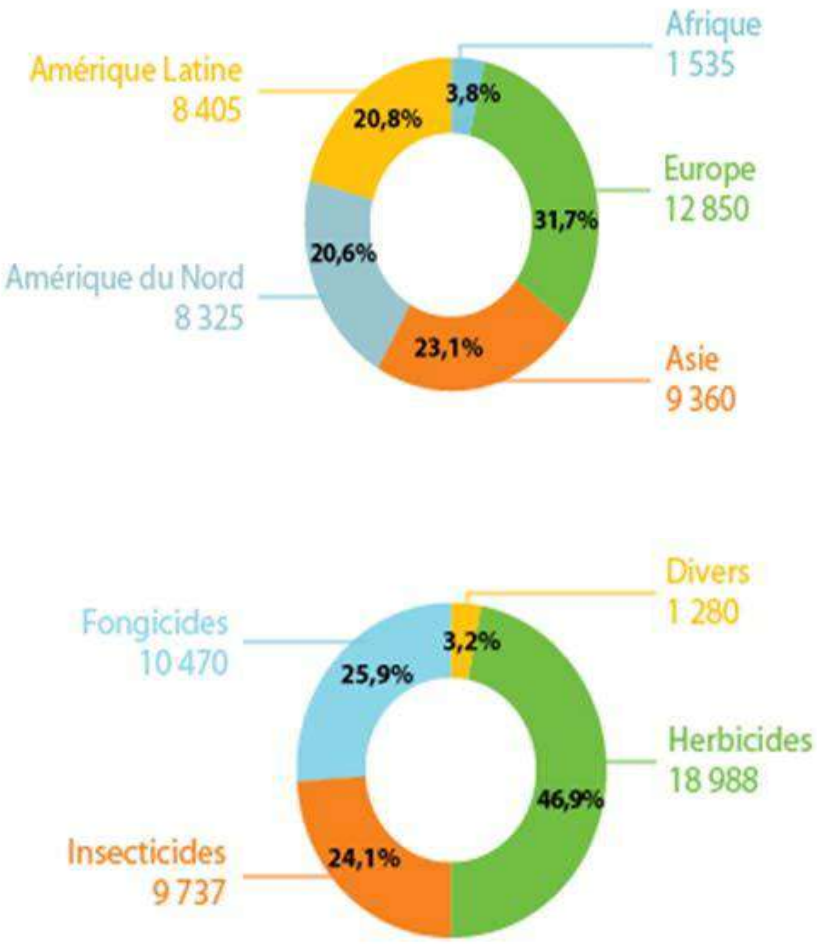


Figure I.2 : Le marché mondial des pesticides par catégorie de produit (Fergani et Mami. 2011)

I.8 Toxicité des pesticides

La toxicité aigüe d'une substance chimique est estimée par des séries de tests réalisés sur des animaux de laboratoire. La valeur la plus utilisée est la dose létale 50 (DL50) qui correspond à la quantité de substance active qui, administrée en une fois, par ingestion, inhalation ou voie cutanée, entraine la mort de 50% des animaux testés. Plus la DL 50 est basse, plus le produit est dangereux.

La toxicité chronique d'une substance chimique administrée en doses répétées est également estimée par des tests normalisés sur des animaux de laboratoire. L'effet carcinogène et l'effet sur la

reproduction sont ainsi estimés. Par ces études, une dose sans effet (NOAEL) peut être déterminée. Cela correspond à la dose maximale n'entrainant pas d'effet statistiquement significatif sur le groupe testé. Tous ces tests permettent la détermination d'une Dose Journalière Acceptable (DJA ou ADI en anglais) qui indique la quantité de substance qu'un humain peut ingérer chaque jour, durant toute sa vie, sans danger pour sa santé (www.era-orleans.org/AFIP/FR/pesticides.)**.** Sur le tableau I.3 est donné le classement toxicologique des pesticides.

Tableau I.3 : Classement toxicologique d'un produit **(**Arrêté du 28 Mars 1989 relatif aux conditions de classement, d'emballages des préparations pesticides).

Valeur de référence	**Ingestion DL50 (rat, orale) (mg/kg)**		**Pénétration cutanée DL50 (rat, cutanée) (mg/kg)**		**Inhalation CL50 (mg/l)**
Formulation	liquide	Solide	liquide	solide	gazeux
Très toxique (T+)	<25	<5	<50	<10	<0.5
Toxique (T)	25 et <200	5 et <50	50 et <400	10 et <100	0.5 et <2
Nocif (Xn)	200 et < 2000	50 et <500	400 et 4000	100 et <1000	2 et <20

I.9 Impacts chroniques sur la sante humaine et ampleur du problème

Les pesticides dangereux ne sont pas seulement un problème du monde en développement. Les pays fortement industrialisés continuent d'utiliser de grandes quantités de pesticides dangereux. Ces derniers continuent encore de causer des agressions sur la santé et l'environnement. Pratiquement tous les pays ont besoin de réformes supplémentaires pour réduire et éliminer les préjudices causés par l'exposition aux pesticides. Néanmoins, l'exposition à des pesticides dangereux est un problème particulièrement grave dans beaucoup de pays en développement. Dans de nombreux pays, l'exposition aux pesticides toxiques est un grave problème de santé avec des proportions épidémiques (Jack Weinberg)**.**

I.9.1 L'ampleur du problème

Plus d'un tiers de tous les travailleurs du monde entier sont dans l'agriculture, mais le pourcentage varie considérablement selon les régions. Dans des pays fortement industrialisés, l'agriculture occupe généralement moins de 5% de la main-d'œuvre. En Amérique latine, au Moyen-Orient et dans les pays de l'ex-Union soviétique, près d'un cinquième (20%) des travailleurs sont employés dans l'agriculture. En Afrique du Nord et l'Asie de l'Est, le nombre s'élève à plus d'un tiers. En Asie du Sud et du Sud-est, près de la moitié de la main-d'œuvre est dans l'agriculture. Et en Afrique sub-saharienne, les deux tiers de tous les travailleurs sont utilisés dans les activités agricoles **(** Employment bysector,ILO2007)**.**

Un récent rapport de la Banque mondiale estime que 355.000 personnes dans le monde meurent chaque année d'empoisonnement involontaire dû à des pesticides **(**www.siteresources.worldbank.org/INTWDR2008/Resources/WDR_00_book.pdf). Une étude plus ancienne faisant autorité (J. Jeyaratnam. 1990)**.** Estime qu'il y a peut-être un million de cas d'intoxications involontaires graves causés par les pesticides chaque année, et deux millions de cas de personnes supplémentaires hospitalisées pour cause de tentatives de suicide avec des pesticides. L'auteur note que cela ne reflète forcément qu'une infime partie du véritable problème et estime qu'il pourrait y avoir jusqu'à 25 millions de travailleurs agricoles dans les pays en développement souffrant d'une forme d'intoxication au travail chaque année à cause de pesticides, bien que la plupart des incidents ne soient pas enregistrés et la plupart des patients ne cherchent pas des soins médicaux. Une des conclusions de cet auteur est que les intoxications par les pesticides peuvent constituer dans certains pays en développement un grave problème de santé publique comme le sont les maladies contagieuses (www. fr.wiktionary.org/wiki/ibid**).**

I.9.2 Les pesticides à toxicité aigüe

Le pesticide aldicarbe qui a été pris dans l'exemple ci-dessus est l'un des nombreux pesticides qui sont classés comme très dangereux. Cette classification remonte à une décision de 1973 de l'Organisation mondiale de la Santé de mettre au point un système de classification des pesticides en fonction des risques qu'ils posent. Dans cette décision, le terme «pesticides dangereux» est défini comme un pesticide qui fait peser un grave risque sur la santé, c'est-à-dire un pesticide qui peut causer des préjudices graves à la suite d'une exposition pendant une courte période.

L'OMS tient une liste à jour des pesticides très dangereux qui sont très toxiques dans sa publication: (*The Recommended Classification of Pesticides Hazard*. 2004) L'OMS utilise comme principale méthode pour identifier les pesticides qui posent un risque grave pour la santé un test dans lequel les rats sont nourris avec le pesticide ou exposés à travers la peau. Ce test produit une valeur appelée DL50 (dose létale 50%), qui est une estimation statistique du nombre de milligrammes de la substance toxique par kilogramme de poids corporel qui va tuer 50% d'une population importante de cobayes. Dans ce système, l'OMS considère un pesticide extrêmement dangereux (classe Ia) ou très dangereux (classe Ib) si elle a une valeur DL 50 faible **(**www.wikipedia.org/wiki/Ibid)**.**

I.9.3 Impacts chroniques sur la santé

Au cours des dernières années, le terme pesticide extrêmement dangereux a été élargi et est maintenant utilisé pour décrire non seulement les pesticides très toxiques, mais aussi les pesticides qui causent de graves effets chroniques sur la santé. Il est généralement beaucoup plus difficile de démontrer les effets chroniques sur la santé que de démontrer des effets toxiques aigus, mais d'importantes recherches sont en cours. Quand il y a une preuve sérieuse qui établit des liens entre un pesticide et un effet chronique grave sur la santé, ce pesticide est ainsi considéré comme un pesticide extrêmement dangereux. Les effets chroniques sur la santé liés aux pesticides concernent les cancers et tumeurs, les troubles du système nerveux, des problèmes de reproduction, les effets sur le système immunitaire, la perturbation du système endocrinien **(** Sowing Poison et al. 2009)

i. Le cancer

Il est connu que beaucoup matières actives des pesticides sont à l'origine ou sont soupçonnés d'être à l'origine du cancer. Les pesticides sont associés, soit par des éléments de preuve en laboratoire ou des études épidémiologiques, à une longue liste de cancers, y compris le myélome multiple, les sarcomes des tissus mous, le sarcome d'Ewing, le lymphome, le lymphome non hodgkinien, la leucémie, le mélanome, le neuroblastome ou de la tumeur de Wilm, les tumeurs à cellules germinales, le rétinopathie (tumeur de l'œil) et le cancer de l'oesophage, de l'estomac, de la prostate, du testicule, du sein, de l'ovaire, du col de l'utérus, de la vessie thyroïde, des poumons, du cerveau, des reins, du pancréas, du foie, du colon et du rectum**.**
(www.unep.org/ceh/main01.html; www.linkedin.com/in/stephanielynnwatts)

ii. Troubles du système nerveux

L'exposition aux pesticides a été associée à l'altération du développement du système nerveux qui peut entraîner une baisse de l'intelligence et des troubles du comportement **(** UNEP et al. 2002)**.** Il existe des preuves reliant divers pesticides aux effets sur le système nerveux central, le système nerveux périphérique et le développement prénatal du cerveau, notamment:

- Un développement plus lent des aptitudes et une augmentation de l'agressivité chez les enfants.
- Des effets dépressifs qui peuvent entrainer des suicides. Une neuropathie retardée, impliquant une dégénérescence des nerfs périphériques des membres avec des maux et des douleurs musculaires et des symptômes grippaux.
- Le changement de personnalité, les troubles de concentration et de mémoire, les troubles du langage, le sens de l'odorat très développé, la détérioration de l'écriture, une altération de la tolérance à l'exercice et les déficits neuromusculaires.
- La maladie de Parkinson et le parkinsonisme - des troubles avec des symptômes tels que la maladie de Parkinson qui peuvent toutefois être réversibles (www.linkedin.com/in/stephanielynnwatts).

iii. Les problèmes de la reproduction

Certains pesticides peuvent être liés à un certain nombre de problèmes de reproduction, y compris des anomalies congénitales, la stérilité, les grossesses prolongées, l'avortement spontané, les mort-nés, les naissances prématurées, le retard de croissance intra-utérin, mortalité périnatale, l'endométriose, et la baisse de spermatozoïdes. Des études épidémiologiques ont établi un lien entre l'exposition des parents à certains pesticides et des anomalies du tube neural, des malformations congénitales cardiaques, la fente labiale et la fente palatine, les malformations congénitales musculo-squelettiques, anomalies du tractus urinaire et anomalies des organes génitaux masculins (www.fr.wiktionary.org/wiki/ibid).

iv. Les effets sur le système immunitaire

L'exposition aux pesticides peut compromettre le système immunitaire ce qui augmente le risque de maladies infectieuses et de cancer. Cela devient particulièrement préoccupant quand le

système immunitaire est déjà compromis par d'autres facteurs, telle que la malnutrition**.** Certains pesticides comme le thymus et la rate altèrent le développement des organes du système immunitaire et portent atteinte à la capacité des globules blancs et des lymphocytes à tuer les bactéries, les virus et les cellules cancéreuses (www.linkedin.com/in/stephanielynnwatts).

v. Perturbation du système endocrinien

Certains pesticides, avec de très petites doses, peuvent imiter ou bloquer les hormones ou peuvent déclencher une activité hormonale inappropriée. Cela peut en rajouter aux effets nocifs sur la santé telle que la stérilité, la baisse du nombre de spermatozoïdes et le cancer du sein**.** D'autres effets indésirables éventuels sur la santé incluent l'avortement spontané, le sex-ratio biaisé dans la descendance des communautés exposées, les anomalies du tractus reproducteur mâle et femelle, y compris des malformations génitales et d'autres anomalies congénitales, la puberté précoce, le syndrome des ovaires poly kystiques, une atteinte de la fonction immunitaire ainsi qu'une variété de cancers. Une étude a identifié 127 pesticides soupçonnés d'avoir des effets perturbateurs sur le système endocrinien (UNEP et al. 2002)**.**

vi. Le suicide

L'auto-intoxication aux pesticides dangereux est probablement la méthode de suicide la plus fréquemment utilisée dans le monde et constitue une pratique largement confinée dans les pays en développement **(** J. M. Bertolote et al. 2006)**.** Les recherches montrent que beaucoup d'individus qui tentent de se suicider ne veulent pas mourir.

Dans les pays très industrialisés, une forme courante de tentative de suicide réside dans la surdose de somnifères ou de drogues similaires. Le taux de mortalité résultant de telles tentatives peut baisser jusqu'à 0,5%, et les études menées dans les pays industrialisés constatent que seulement 2% se suicident au cours des 12 mois qui suivent la première tentative. D'autre part, une tentative de suicide par ingestion de pesticides agricoles peut entrainer des taux de mortalité au-delà de l'ordre de 10% à 20% **(** Gunnell and Eddelston. 2003)**.**

Selon les statistiques de l'OMS, 18 millions de personnes tentent de se suicider chaque année et cela se traduit par plus de 800.000 morts **(**preventing_suicide.pdf)**.** Selon un dépliant de l'OMS sur la prévention du suicide: *«On trouve le taux le plus élevé de comportement suicidaire chez les femmes dans les zones rurales des pays asiatiques, où le facteur de risque principal est la disponibilité*

*des pesticides, qui transforme les tentatives de suicide impulsives par empoisonnement en actes meurtriers en moins de 3 heures»***.**

Ce dépliant liste les interventions qui ont démontré leur efficacité dans la prévention des comportements suicidaires. Le premier point de la liste est : «Contrôle de l'accessibilité des substances toxiques (en particulier les pesticides dans les zones rurales des pays asiatiques**)** (www.fr.wiktionary.org/wiki/ibid)**.**

I.10 Effets sur l'environnement

Les pesticides employés pour lutter contre les organismes nuisibles peuvent se retrouver dans l'environnement. Ils risquent alors d'engendrer une contamination ponctuelle ou diffuse. Cette pollution diffuse semble être la source dominante d'apport de pesticides vers les eaux de surface et souterraines .Celle-ci peut être associée à plusieurs mécanismes de transport qui permettent aux pesticides de quitter la parcelle agricole et de se retrouver dans l'environnement. La figure I.4 illustre ces mécanismes.

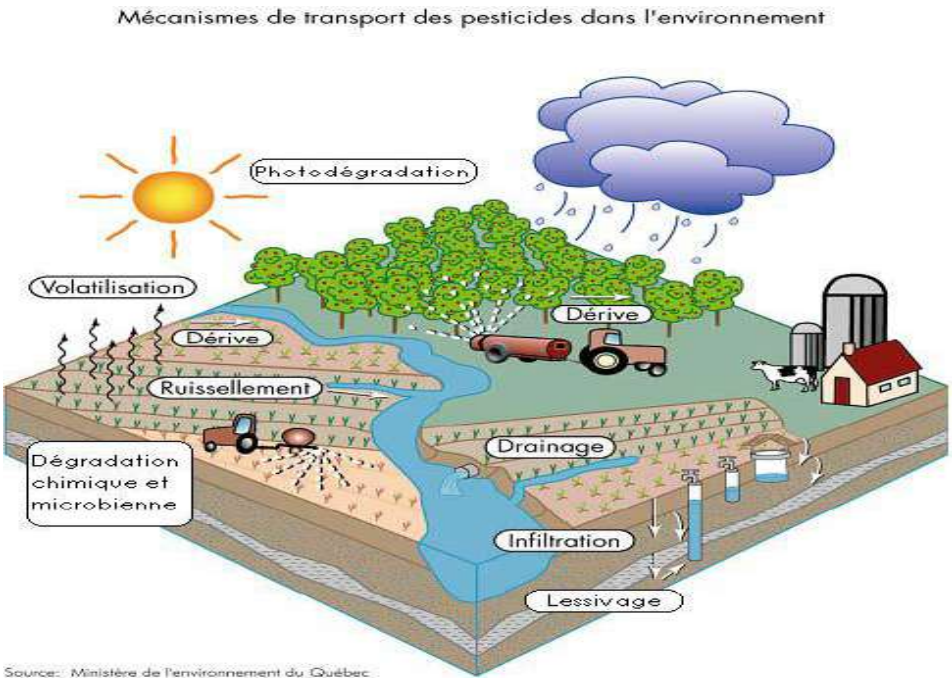


Figure I-3 : Mécanismes de transport des pesticides dans l'environnement (Les pesticides en milieu agricole : état de la situation environnementale et initiatives prometteuses. pdf)

Les principaux mécanismes de transport des pesticides vers les eaux de surface et souterraines sont le ruissèlement (pesticide sous forme dissoute ou lié à des particules de sols) et le lessivage. Les caractéristiques physicochimiques propres à chacun des pesticides - telle la persistance, l'adsorption, la pression de vapeur et la solubilité du pesticide - influencent sa tendance à quitter la parcelle. Les caractéristiques du sol et de la nappe d'eau souterraine, les précipitations, le travail du sol, les méthodes d'application, les pratiques culturales et le choix du pesticide jouent également un rôle important sur les pertes de pesticides par ruissèlement et lessivage. Le drainage du surplus d'eau qui s'infiltre dans les sols agricoles par les canaux de drainage souterrain entraine aussi les pesticides vers les eaux de surface.

Les pesticides peuvent aussi être transportés par l'air. Lors de l'application, la dérive des gouttelettes de pesticides fait en sorte que des produits peuvent se retrouver à l'extérieur de la parcelle. Le phénomène de la dérive est influencé, entre autres, par la taille des gouttelettes pulvérisées, la vitesse du vent, le type d'équipement utilisé et la hauteur de la pulvérisation.

Les pesticides peuvent également se retrouver dans l'air sous forme de vapeurs, un phénomène appelé la volatilisation. Après l'application de certains pesticides, ceux-ci sont libérés sous forme de gaz ou adhèrent à des particules comme les poussières. Les résidus de pesticides sous forme gazeuse dans l'atmosphère ont souvent une demi-vie de courte durée, car ils sont susceptibles à la photo dégradation et sont transportés sur de courtes distances. Par contre, les pesticides moins sujets à la photo dégradation et qui ont une demi-vie beaucoup plus longue. Par la suite, ces résidus de pesticides sous forme gazeuse peuvent être soustraits de l'atmosphère par condensation et retourner dans l'eau ou sur le sol par l'effet des précipitations ou par le dépôt de fines particules solides.

La volatilisation du pesticide à partir des surfaces traitées est principalement fonction de la chaleur, de la pression de vapeur associée au pesticide et de l'humidité relative. La présence de certains pesticides (atrazine, métolachlore, EPTC) dans l'eau de pluie. Une fois dispersés dans l'environnement, certains pesticides peuvent être dégradés par la lumière ou les microorganismes, tandis que d'autres persistent et peuvent s'accumuler ou se transformer en d'autres contaminants parfois plus à risque que le produit d'origine (Les pesticides en milieu agricole : état de la situation environnementale et initiatives prometteuses .pdf).

Sur la figure I.4 ci-contre, est illustré le processus impliqué dans le devenir des pesticides dans les sols conditionnant leur disponibilité et, par conséquent, leur efficacité phytosanitaire ou la manifestation de leur caractère polluant.

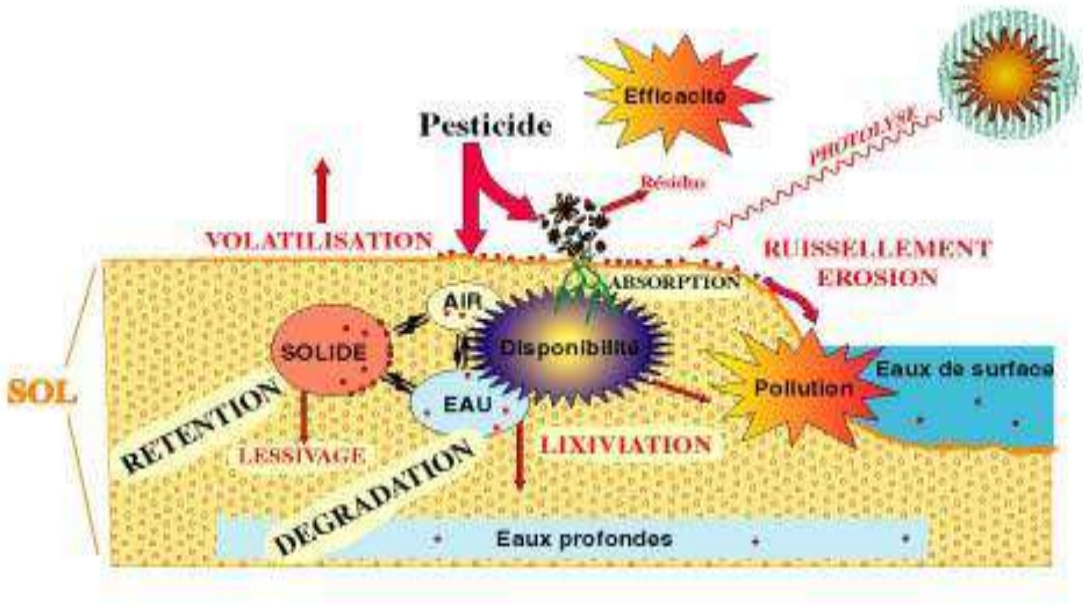


Figure I-4 : Processus impliqué dans le devenir des pesticides dans les sols conditionnant leur disponibilité et, par conséquent, leur efficacité phytosanitaire ou la manifestation de leur caractère polluant.

Des études émanant de divers groupes de recherche ont été menées en France depuis la fin des années 80 et se poursuivent à l'heure actuelle. Les premières données permettent de constater la présence de pesticides dans toutes les phases atmosphériques, en concentrations variables dans le temps (caractère parfois saisonnier, lien avec les périodes d'application) et l'espace (proximité des sources) ; même des composés peu volatils ou interdits (lindane par exemple) sont parfois détectés.

I.11 Conclusion

Tout au long de ce chapitre, nous avons tenté de donner un maximum d'informations sur les pesticides commençant par mentionné la composition de ces dernier et leurs classification sur 4 différents échelles (leurs mode d'action, leurs structure chimique, leurs usage et finalement selon le classement de l'organisation mondiale de la santé).

Nous avons montré dans ce chapitre les différents impactas des pesticides (exercé sur la sante humaine) et leurs effets sur l'environnement : L'exposition de l'humanité et de son environnement aux pesticides est une source de graves problèmes de santé dans toutes les régions du monde, mais surtout dans les pays en développement. Aussi, tous les efforts doivent être conjugués afin de minimiser ou d'éliminer carrément l'utilisation de pesticides reconnus dangereux pour l'homme, la faune et la flore. D'un autre côté, les efforts des scientifiques axés sur la recherche de méthodes de prédiction de la toxicité des nouveaux pesticides doivent être soutenus et encouragés.

CHAPITRE II

RELATION QUANTITATIVE STRUCTURE - ACTIVITE

II.1 Introduction

Découvrir de nouveaux médicaments de la manière la plus efficace et la moins couteuse possible constitue un enjeu majeur pour les années à venir. Il est admis que, en moyenne, pour une molécule qui arrive sur le marché en tant que médicament innovant, 10 000 molécules sont synthétisées et testées. De plus, le développement d'un médicament demande généralement entre 10 et 15 ans de recherches. Il s'agit en effet de trouver une molécule qui doit à la fois présenter des propriétés thérapeutiques particulières, et posséder le minimum d'effets secondaires indésirables. Le prix de revient d'un médicament est essentiellement dû à ces synthèses longues, couteuses et finalement inutiles. Pour cette raison, l'industrie pharmaceutique s'oriente vers de nouvelles méthodes de recherche, qui consistent à prédire les propriétés et activités de molécules avant même que celles-ci ne soient synthétisées. Deux disciplines de la « chimie computationnelle » se sont développées en réponse à ce besoin : les relations structure-activité ou QSAR. Elle consiste essentiellement en la recherche de similitudes entre molécules dans de grandes bases de données de molécules existantes dont les propriétés sont connues. La découverte d'une telle relation permet de prédire les propriétés physiques et chimiques et l'activité biologique de composés, de développer de nouvelles théories ou de comprendre les phénomènes observés. Elle permet également de guider les synthèses de nouvelles molécules, sans avoir à les réaliser, ou à analyser des familles entières de composés.**(** www.neurones.espci.fr/Theses_PS/These%20AG-convertie-PASTEL.pdf)

II.2 Définition

Une relation quantitative structure à activité (en anglais : *Quantitative structure-activity relationship* ou QSAR) est le procédé par lequel une structure chimique est corrélée avec un effet bien déterminé comme l'activité biologique ou la réactivité chimique. Ainsi par exemple l'activité biologique peut être exprimée de manière quantitative, comme pour la concentration de substance nécessaire pour obtenir une certaine réponse biologique. De plus lorsque les propriétés ou structures

physicochimiques sont exprimées par des chiffres, on peut proposer une relation mathématique, ou *relation quantitative structure à activité*, entre les deux. L'expression mathématique obtenue peut alors être utilisée comme moyen prédictif de la réponse biologique pour des structures similaires.

La QSAR la plus commune est de la forme : activité = f(propriétés physicochimiques et/ou structurales). (www.wikipidia.org)

Les informations extraites à partir des résultats d'étude de QSAR peuvent être utilisées pour obtenir une meilleure connaissance des structures moléculaires et probablement le mode d'action au niveau moléculaire. Ces informations peuvent alors être utilisées pour prévoir les propriétés physicochimiques et les activités biologiques de nouveaux composés ainsi que pour concevoir de nouvelles structures (**Figure II -1**)

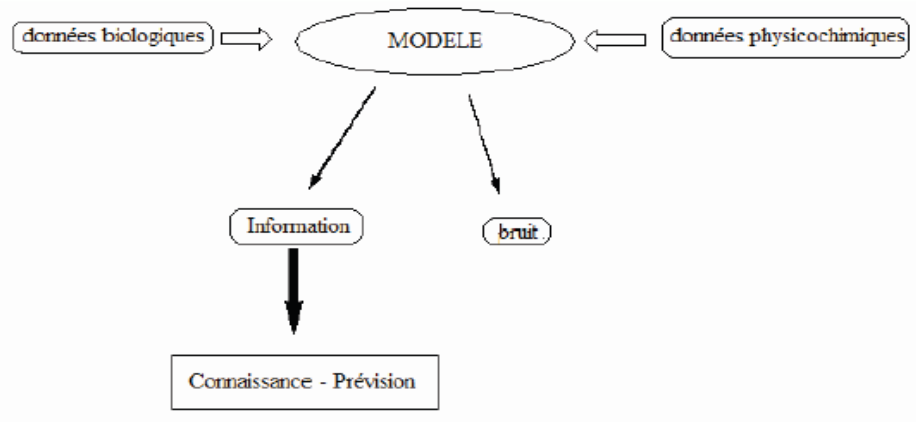
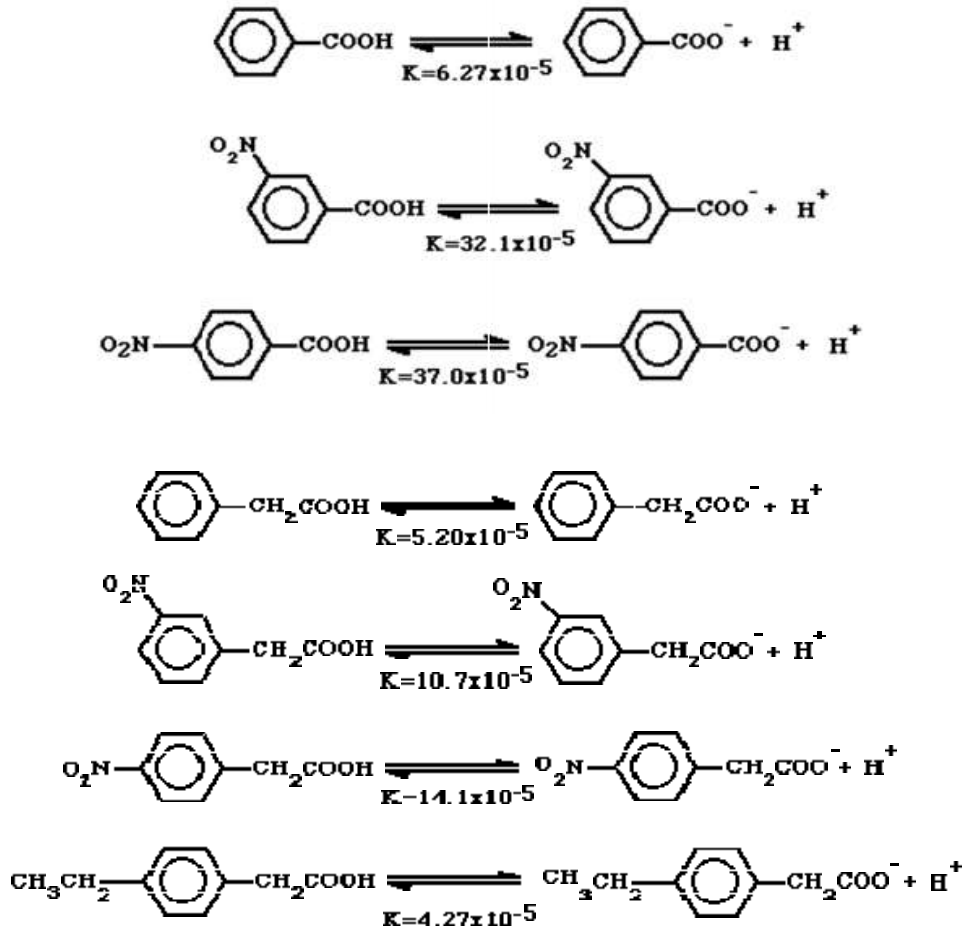


Figure II.1 : Etude de QSAR et son application **(** H. Waterbeemd, S. Rose. 2003)

II.3 Historique

- **19ème siècle** : Début de QSAR.
- **1863** : Des chercheurs appartenant à l'AFA0 Cros à l'Université de Strasbourg ont montrés que la diminution de la solubilité des alcools dans l'eau entrainait une augmentation de la toxicité de ces alcools pour les mammifères,
- **1890** : Hans Horst Meyer de l'Université de Marburg et Charles Ernest Overton de l'Université de Zurich, travaillant indépendamment, ont noté que la toxicité des composés organiques dépend de leur caractère lipophile.

- **1894-1987** : Louis Hammett à fait une corrélation entre les propriétés électroniques des acides et des bases organiques avec leurs constantes d'équilibre et leurs réactivité. En envisageant la dissociation de l'acide benzoïque (réaction 1), Hammett a observé que l'ajout de substituant sur le noyau aromatique de l'acide benzoïque a un effet ordonné et quantitatif sur la constante de dissociation (réactions 2-3). Hammett a également observé que les substituants ont un effet similaire sur la dissociation des acides organiques et d'autres bases (cas de la dissociation des acides phénylacétique schématisées par les réactions 4-7) **(** Ami et Melzi. 2011)**.**



II.4 Applications de la QSAR

II.4.1 En chimie

Une des premières applications de la QSAR concerne la prédiction des points d'ébullition (Rouvray,D. H. et al. 1991)**.**

Il est bien connu par exemple que pour une famille de composés chimiques, particulièrement en chimie organique, il existe une corrélation forte entre la structure et les propriétés observées. On peut citer comme exemple simple, la relation entre le nombre de carbones dans les alcanes et leur point d'ébullition. Il existe une tendance nette à l'augmentation de la température d'ébullition avec le nombre d'atomes de carbone, ce qui sert de moyen prédictif pour les points d'ébullition des alcanes les plus lourds. Les méthodes de Hammett, de Taft et prédiction de pKa sont des applications particulièrement intéressantes (Fraczkiewicz R. 2007).

II.4.2 En biologie

L'activité biologique des molécules est mesurée habituellement au moyen d'essais afin d'établir le niveau d'inhibition d'une transduction de signal ou d'une voie métabolique particulière. Les produits chimiques peuvent être biologiquement actifs par leur toxicité. La recherche de médicament implique parfois l'utilisation de la QSAR afin d'identifier les structures chimiques pouvant présenter de bons effets inhibiteurs sur des cibles spécifiques et possèdent une faible toxicité (activité non-spécifique). Un intérêt particulier est la prédiction du coefficient de partition log P, qui est une mesure importante pour l'identification de la «similarité médicamenteuse » selon la règle des cinq de Lipinski.

Alors que de nombreuses analyses QSAR traitent des interactions d'une famille de molécules avec une enzyme ou un site récepteur, la QSAR peut aussi être utilisée pour étudier les interactions entre les domaines structuraux des protéines. Les interactions entre protéines peuvent être quantitativement analysées pour des variations structurales résultantes d'une mutagenèse **(** Freyhult EK et al. April 2003) La réduction du risque de paradoxe RAS fait partie de l'apprentissage automatique, et particulièrement la prise en compte du fait que seul un nombre fini de données est disponible (voir aussi estimateur de minimum de variance non biaisé). En général, tous les problèmes QSAR peuvent être partagés entre une partie codage **(**Hendrik T et al. 2002)**.** Et une partie apprentissage **(** Strok et al. 2001).

II.4.3 Analyse de données

La première étape dans la modélisation par **QSAR** est la représentation des structures moléculaires par des nombres nommés des descripteurs moléculaires. Un nombre relativement important de caractéristiques ou de descripteurs moléculaires est calculé pour le codage, qui peut cependant manquer de pertinence sur l'interprétation structurale. Il existe donc un problème de choix de variable, qui peut se résoudre par des méthodes d'apprentissage, appliquées comme étape de post-traitement ou de prétraitement. Parmi ces méthodes, les machines à vecteurs de support, les décisions, ou les réseaux de neurones peuvent être utilisées pour induire un modèle d'apprentissage prédictif.

Les descripteurs basés sur les propriétés pharmacophores mettent en relation les propriétés physicochimiques des atomes avec la topologie de la molécule. Ces propriétés pharmacophoriques sont essentiellement les atomes donneurs et accepteurs de liaisons hydrogène, les charges, les centres hydrophobes et les cycles aromatiques. La distance topologique entre deux éléments du pharmacophore donne une empreinte pharmacophorique qui pourra être utilisée pour la recherche de molécules de topologie similaire (Ami et Melzi. 2011)

D'autres descripteurs de type QSAR (Quantitative Structure-Activity Relationship) sont des modèles prédictifs qui ont pour objectif d'estimer la capacité d'une molécule à devenir un médicament suffisamment stable dans l'organisme pour être actif et concernent par exemple la lipophilie d'une molécule **(** Corinne H. 2008).

Différents formats de fichiers ont été décrits pour représenter la structure des molécules. Les plus connus sont SMILES, SDF, mol2 (Tripos) et PDB (Protein Data-Bank). Le format SMILES (Simplified Molecular Input Line Entry System) permet de représenter les molécules sur une seule ligne de caractères. Les tirets simples ou doubles représentent des liaisons de même nature, et un système de parenthèses permet de repérer les embranchements au niveau de la molécule. Les chiffres permettent d'identifier les connections de cycle. Le problème de ce code est sa dégénérescence. En effet, à une même molécule peuvent correspondre plusieurs codes SMILES. Le code SMILES canonique a permis de résoudre ce problème. L'avantage principal de ce format est qu'il est léger, permettant le stockage des

bases des molécules de grande taille dans un espace mémoire restreint. Cependant, d'autres informations sont souvent nécessaires, notamment la structure 3D de la molécule. Des formats comme PDB, SDF ou mol2 contiennent les coordonnées cartésiennes de chaque atome, et éventuellement leur charge partielle (mol2), ainsi que toutes les liaisons formant la molécule. Open Babel permet la conversion d'un format 2D vers un autre format 3D. La conversion d'un format 2D vers un format 3D nécessite de générer de l'information et est rendue possible grâce à des outils comme Omega, Corina, Catalyst, Concord ou Frog. Selon la molécule, différents états d'ionisation et stéréo-isomères peuvent être générés et il est important de le vérifier avant (Ami Et Melzi. 2011)

II.5 La QSAR en toxicologie de l'environnement

La QSAR est un outil pour la prédiction de l'activité biologique, et se prête ainsi facilement à la prédiction de la toxicité pour l'environnement. Au cours des 20 dernières années, la QSAR dans le domaine environnemental n'a cessé de croitre en importance, et Nendza **(** Nendza, M. 1998) a admirablement résumé ses réalisations. Il a maintenant atteint le stade où certains organismes de règlementation, tels que l'Agence américaine de protection de l'environnement, utilise régulièrement des QSAR pour prédire la toxicité à des fins règlementaires. Il est prévu que cette utilisation va augmenter fortement à l'avenir, à mesure que des assurances sont demandées sur la sécurité des produits chimiques d'une part et plus la pression publique est exercée contre l'utilisation des animaux dans les essais de toxicité d'autre part. Il convient de noter, toutefois, que les données expérimentales sur la toxicité sont nécessaires en premier lieu afin de développer une QSAR et il y'a encore une pénurie de données de bonne qualité dans de nombreux domaines.

II.6 La QSAR pour la prédiction de la toxicité

La plupart des données sur la toxicité de l'environnement ont été obtenues en utilisant des animaux aquatiques comme les poissons de diverses espèces, les daphnies, TetrahymenapyriformisVibriofischeri, et les algues. Cronin et Dearden **(** Cronin, M.T.D et al. 1995) ont passé en revue la littérature concernant la modélisation QSAR de la toxicité aquatique.

Huit modes d'action ont été identifiées chez les poissons, à savoir la narcose non polaire, narcose polaire, le découplage de la phosphorylation oxydative, l'irritation de la membrane respiratoire, l'inhibition de l'acétylcholinestérase, la saisie du système nerveux central, l'inhibition de la photosynthèse, et l'alkylation**(** Bradbury et al. 1994). Toutefois, ceux-ci sont généralement plus regroupés comme suit: la narcose non polaire, narcose polaire, la réactivité sélective, et les mécanismes d'action spécifiques. Il est important, afin d'obtenir une prédiction RQSA correcte de la toxicité, que le mode d'action d'un produit chimique soit correctement identifié.

La corrélation QSAR première narcose non polaire a été développée par Könemann **(** Könemann, H; Toxicol. 1981) qui a corrélé la toxicité aiguë de divers produits chimiques industriels.

$$Log1/LC_{50}=0.87logp-1 \qquad (II.1)$$

Avec n=50, r^2=0.98, s=0.23 et où CL50 = concentration (mmol/L) pour tuer 50% des poissons dans un délai déterminé, n = nombre de produits chimiques, r = coefficient de corrélation, et s = erreur-type de l'estimation.

Veith et Broderius ont signalé que certains produits chimiques réactifs, tel que le phénol (2) et l'aniline (3) produits dérivés, qui produisent des effets toxiques compatibles avec narcose ont néanmoins été prévus plus toxiques que pour la narcose non polaire. Ce mode d'action est maintenant appelé narcose polaire, et peut résulter de la présence d'un groupe de liaisons hydrogène fortes dans une molécule.

En règle générale, la toxicité des produits chimiques réactifs ne peut être modélisée que par l'inclusion d'un ou plusieurs descripteurs qui reflètent la réactivité. Généralement, une telle réactivité électrophile est dans la nature de la molécule, puisque les groupes nucléophiles tels que NH, OH et SH abondent dans les macromolécules biologiques. Cronin et Schultz **(**Cronin, M.T.D, Schultz, T.W. 2001) ont trouvé ce qui suit : la QSAR pour la toxicité des composés aromatiques par rapport à un cilié aquatiques, Tetrahymena pyrifomis :

$$Log1/IGC_{50}=0.603logp-0.330^{E}_{LUMO}-1 \qquad (II.2)$$

Avec n=239, r^2=0.8000, q^2=0.796, s=0.335, f=476etoù IGC50 = la concentration (mmol/L) pour inhiber la croissance de 50%, ELUMO = énergie de l'orbitale moléculaire inoccupée plus bas, Q = corrélation croisée validé coefficient (procédure de leave-one-out), et F = Fisher statistique.

Les statistiques fournies par l'équation II.3 sont les statistiques préférées pour la déclaration des QSAR. La corrélation de coefficient d'une validation croisée est une mesure de la productivité des QSAR (par opposition à sa capacité de seulement corrélative); F donne une mesure de probabilité que la corrélation n'a pas apparue par hasard.

Comme la plupart des QSAR sont développées à des fins prédictives, il est important que leur capacité prédictive soit évaluée. Bien que la validation croisée est une méthode d'évaluation de la capacité prédictive, il s'agit d'une méthode interne, et peut être critiquée comme étant simplement de donner une indication de la cohérence interne de l'ensemble des données utilisées pour élaborer les QSAR (l'ensemble d'apprentissage). Une meilleure méthode consiste à utiliser le modèle QSAR pour prédire la toxicité des produits chimiques non utilisés dans l'ensemble d'apprentissage (validation externe).

D'autres mesures de réactivité peuvent être utilisée dans QSAR impliquant des produits chimiques réactifs. Un exemple intéressant est donné par Mekenyan et Veith **(**Mekenyan, O.G.Veith, G.D, 1993) qui ont fait une corrélation entre la toxicité des produits chimiques avec trois modes d'action différents pour P. promelas par l'utilisation du descripteur orbitale moléculaire connue sous le nom de superdelocalisability:

$$\text{Log } 1/LC_{50}=0.56 \log p + 13.7\, s^{N}_{av} - 1.49 \qquad \textit{(II.3)}$$

Avec n=114,r^2=0.81,s non donné et où SavN = superdelocalisability moyenne nucléophile.

Les produits chimiques spécifiquement actifs sont caractérisés par les insecticides organophosphorés comme le malathion (4).Hermens et al**(** Hermens. J et al. 1987)**.** ont découvert qu'ils pouvaient modéliser la toxicité de ces composés à P.reticulata avec un QSAR à deux mandats

$$\text{Log } 1/\, LC_{50}=0.23\sum 0.80 + \square k + 2.77 \qquad \textit{(II.4)}$$

Avec n=9,r^2=0.92,s=0.19 et où Sp = somme des constantes hydrophobes fragmentaire, et k = taux de réaction constante déterminée expérimentalement avec le 4-nitrobenzylpyridine.

II.7 Les systèmes experts pour la prédiction de la toxicité

Un système expert a été défini **(** Dearden, J.C. 1997)**.** Comme « tout système formel, pas nécessairement sur ordinateur, qui permet à un utilisateur d'obtenir des prévisions rationnelles sur les propriétés des produits chimiques ». Greene**(** Greene, N. 2002)**.** A récemment passé en revue les systèmes experts pour la prédiction de la toxicité.

Un certain nombre de ces systèmes sont disponibles dans le commerce, alors que la plupart d'entre eux s'occupent principalement des risques pour la santé de l'homme tels que la cancérogénicité, la tératogénicité et la sensibilisation de la peau. Certains comprennent des modules pour la prédiction éco-toxicologique.

Parmi ces systèmes d'expert, on cite quelques uns :

- **TOPKA** :

  Ce système utilise une approche QSAR, qui repose en grande partie sur des descripteurs topologiques, pour prédire des valeurs CL50 pour Daphnia magna et P. promelas.

- **MULTICASE**

  Ce système identifie des fragments structuraux qui sont responsables d'une toxicité donnée, y compris des valeurs CL50 chez les poissons.

- **HAZARDEXPERT**

Ce système intègre les connaissances d'experts de l'homme pour identifier les fragments structurels liés à la toxicité. Il fournit des prédictions à travers une série de niveaux trophiques avec les différents régimes de dosage

- **ASTER**

Ce système a été développé par l'Agence américaine de protection de l'environnement. Les points de terminaison écotoxicologiques sont largement couverts, des

valeurs de la CL50 nommément pour P. promelas, méné de-mouton (Cyprinodon) et D. magna. Il s'agit essentiellement d'une base de données, mais utilise QSAR pour prédire la toxicité lorsque des données expérimentales ne sont pas disponibles. Il utilise des modèles QSAR pour les différents types de produits chimiques (stupéfiants non-polaires, polaires et de stupéfiants ainsi de suite). En raison des restrictions de sécurité, ASTER n'est actuellement pas accessible au public.

- **ECOSAR**

Ce système est librement téléchargeable sur site. Il utilise QSAR hydrophobicité et prédit la toxicité pour les poissons, les daphnies et les algues vertes (Fergani et Mami. 2011).

II.8 Calcul des propriétés physicochimiques

Un certain nombre de propriétés physicochimiques sont d'une importance environnementale, que les QSAR données ci-dessus illustrent.

- **Coefficient de partage (P)**

C'est sans doute le plus important d'entre eux, et il ya de nombreux logiciels disponibles pour le calcul de log P. Parmi les meilleurs, on cite les suivants: Analyse interactive qui permet gratuitement de calculer en ligne ; Biobyte, SPARC**,** qui permet gratuitement de calculer en ligne, et le logiciel SRC KOWWIN, qui est téléchargeable gratuitement à partir de l'EPA**.**

- **Solubilité aqueuse**

C'est une autre propriété importante de l'environnement. L'analyse interactive et de sites web SPARC permet le libre calcul en ligne de solubilité, et WSKOWWIN de la SRC est téléchargeable gratuitement à partir de l'EPA site web.

- **Point d'ébullition**

Il est disponible à partir de SPARC (librement en ligne), MPBPVP de la SRC**,** ACD et ProPred (Fergani et Mami. 2010).

- **pression de vapeur**

Il est disponible à partir de SPARC (librement en ligne) et ACD.

- **point de fusion**

Il est disponible auprès de MPBPVP et ProPred.

- **Le partage air-eau**

Ce coefficient (constante de Henry) a des implications importantes pour la distribution de produits chimiques dans l'environnement. De nombreuses tentatives ont été faites pour le prédire; Dearden et Schüürmann **(** Dearden, J.C. Schüürmann, G. 2001)**.** Ont récemment revu la prévision RQSA de la constante d'Henry. Le logiciel SRC HENRYWIN est téléchargeable gratuitement à partir de l'EPA site web.

II.9 Notation Smiles (Simplified Molecular Input Line Entry Specification)

II.9.1 Notation SMILES

The Simplified Molecular Input Line Entry Specification ou SMILES est un langage symbolique de description de la structure des molécules chimiques sous forme de courtes chaines de caractères ASCII. Les chaines SMILES peuvent être importées dans la plupart des éditeurs de structure moléculaire pour être reconverties en représentation 2D ou en modèle 3D, et par conséquent, permettent de construire des bases de données informatiques **.**

La définition du langage SMILES a été initialement introduite par David Weininger à la fin des années 1980. Plusieurs extensions et modifications ont été ajoutées depuis, notamment par la société DaylightChemical Information Systems, dont Weininger est le fondateur. Il existe d'autres langages à ligne de texte pour représenter les molécules, dont notamment InChl, introduit par l'IUPAC. SMILES est cependant plus répandu, car plus facile à lire par l'utilisateur humain et surtout parce qu'il est implémenté dans de nombreux logiciels de chimie moléculaire (Ami et Melzi. 2011).

II.9.2 Principes de Smiles et théorie des graphes

SMILES est bâti sur la représentation d'une molécule chimique sous forme d'un graphe, au sens mathématique du terme. Les atomes sont les sommets du graphe, et les

liaisons chimiques en sont les arêtes. Chaque sommet du graphe est étiqueté avec le symbole de l'atome correspondant (C, N, O…)**(** Weininger D. 1988)**.**

La chaine de caractère SMILES correspondant à une molécule donnée est obtenue en parcourant progressivement le graphe suivant un algorithme de parcours en profondeur et en imprimant successivement le symbole de chaque sommet traversé**.** Le graphe de la molécule est préalablement simplifié en supprimant tous les atomes d'hydrogène (notation implicite) et en ouvrant les cycles pour transformer le graphe en arbre couvrant (spanningtree). À chaque fois qu'un cycle est ouvert, un suffixe numérique est ajouté pour indiquer la connexion des sommets correspondant à la liaison chimique supprimée. Des parenthèses sont utilisées pour indiquer les points de branchement sur l'arbre**.**

De par la nature de ce langage, il existe, pour une même molécule organique constituée de plusieurs atomes autres que l'hydrogène, plusieurs écritures possibles, selon l'atome de départ et l'ordre de parcours de la structure moléculaire. Ainsi, CCO, OCC, C(C) O et C(O) C sont toutes les quatre, des représentations valides de la molécule d'éthanol CH_3–CH_2–OH**.**

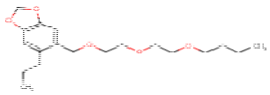
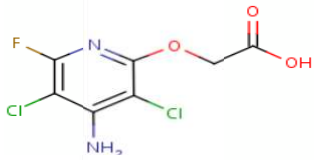
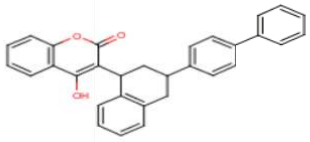
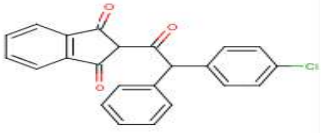
Les principes théoriques du langage SMILES sont présentés de manière exhaustive sur le manuel du SMILES théorique proposé par DaylightChemical Information Systems (en anglais). Le tableau II.1 illustre la notation SMILES pour quelques molécules.

Il existe des algorithmes informatiques pour transformer automatiquement le schéma d'une molécule donnée en chaîne SMILES. Comme il n'existe en général pas de description unique d'une molécule, il y a également plusieurs SMILES pouvant décrire la même structure chimique. Tout dépend de l'atome dont on part pour décrire la structure, de l'ordre dans lequel on décrit les branches du graphe moléculaire et de la manière dont on ouvre les cycles (Weininger D, 1988)**.**

Il existe plusieurs logiciels utilisés pour traduire la structure en formule SMILES, citons :

- **ChemOffice :** est une suite de logiciels et d'utilitaires offrant le nec plus ultra en matière de recherche. Elle se décline dans les versions Pro ou Ultra selon les besoins du scientifique ou du chercheur dans le domaine de la chimie **.**

Tableau II.1 : Notation SMILES pour quelques molécules

Nom	Formule brute	Notation de smille	Structure chimique
dinoseb	C10H12N2O5	OC1=C([N+]([O-])=O)C=C([N+]([O-])=O)C=C1C(CC)C	
Piperonyl butoxide	C19H30O5	CCCC1=CC2=C(OCO2)C=C1COCCOCCOCCCC	
fluroxypyr	C7H5Cl2FN2O3	FC1=C(Cl)C(N)=C(Cl)C(OCC(O)=O)=N1	
difenacoum	C31H24O3	O=C(C(C1CC(C2=CC=C(C3=CC=CC=C3)C=C2)CC4=C1C=CC=C4)=C5O)OC6=C5C=CC=C6	
Chlorophacinone	C23H15ClO3	O=C1C(C(C(C2=CC=CC=C2)C3=CC=C(Cl)C=C3)=O)C(C4=C1C=CC=C4)=O	

- **ChemBioDrawUltra :** est un logiciel de représentation de structures pour spécialistes chevronnés, offrant des outils de prévision perfectionnés et une intégration Web complète par le biais du complément ChemBioDraw ActiveX / Plugin. La suite ChemBioDraw Ultra comprend également ChemDraw / Excel, ChemBio3D Pro, ChemFinder Ultra avec BioViz, E-Notebook Pro et les bases de données ChemInfo .

La méthode SMILES permet d'échanger numériquement des descripteurs de molécules arbitrairement complexes par le biais d'un fichier texte classique. Ces descripteurs peuvent êtres calculées par d'autres logiciels tels que DRAGON. Ce dernier est une application utilisée pour le calcul des descripteurs moléculaires à l'origine développé par la Chimio-métrie Milano et QSAR Research Group .

Ces descripteurs peuvent être utilisés pour évaluer la structure moléculaire-activité ou structure-propriété des relations, ainsi que pour l'analyse de similitude et de dépistage des bases de données moléculaire. La représentation SMILES étant très compacte, elle permet de stocker simplement des fichiers de plusieurs centaines de milliers de molécules (www.bkchem. zirael. org/index. Html) (Consulté le 06/09/2010).

II.10 Conclusion

La relation quantitative structure activité(QSAR) est depuis quelques années utilisée de façon systématique dans la conception de nouveaux médicaments et permet de prédire l'activité thérapeutique de molécules avant même de les avoir synthétisées. Cette prédiction, bien que non totale, a l'avantage d'économiser une partie de l'expérimentation initiale. La relation structure-activité est devenue partie intégrante d'un arbre de décision.

La même démarche a été utilisée pour prévoir une toxicité déterminée. Peut-être moins rationnelle ou plus approximative, cette méthode a donné dans le passé des résultats décevants en toxicologie. A partir de ce constat, des travaux récents impliquant des toxicologues, des chimistes et des informaticiens ont montré que le respect d'un certain nombre de règles simples permettait d'obtenir des résultats comparables à ceux obtenus lors du développement de molécules thérapeutiques.

SMARTS est une extension de SMILES qui permet, en plus des définitions classiques, de remplacer des atomes ou des liaisons par des jokers. Ceci est utilisé pour spécifier des motifs chimiques utilisés pour des recherches dans des banques de données de molécules (chimio thèques). Cela permet en particulier la recherche de sous-structures chimiques communes à deux molécules. Les logiciels développés facilitent les calculs des descripteurs et les propriétés des molécules sans essais pratiques.

CHAPITRE III

LES RESEAUX DE NEURONES ARTIFICIELS

III.1 Introduction

Le terme de réseaux de neurones "formels" (ou "artificiels") fait rêver certains, et fait peur à d'autres. La vérité est à la fois plus prosaïque et plus rassurante ; les réseaux de neurones constituent maintenant une technique de traitement de données bien comprise et maitrisée, qui devrait faire partie de la boite à outils de tout ingénieur et chercheur soucieux de tirer le maximum d'informations pertinentes des données dont ils disposent : faire des prévision, élaborer des modèles, reconnaitre des formes ou des signaux, etc.

Les réseaux de neurones artificiels (RNA), qui doivent tout autant leur essor considérable récent à la biologie, montrent de plus en plus leurs importance dans la plupart des domaines d'ingénierie pas uniquement pour leur capacité d'apprentissage tout à fait impressionnante, mais aussi pour leur malléabilité et leur bonne performance, par rapport aux techniques courantes. Les RNA sont, en effet, capables d'approximer toute fonction complexe et non linéaire. Elles sont dites : approximateurs universels.

Ainsi, aujourd'hui, ce n'est pas le paradigme neurobiologique qui contribue au développement des réseaux de neurones formels : au contraire, ce sont les réseaux de neurones formels qui contribuent, de plus en plus fréquemment, à la compréhension des systèmes neuronaux vivants.

Avec l'évolution de l'informatique et l'augmentation de la capacité de calcul offertes par les nouveaux ordinateurs, les réseaux neurones artificiels ont connu une diversification dans leurs applications, ils sont devenus capables d'apporter des solutions efficaces et élégantes : notamment dans le domaine de la reconnaissance de forme, le contrôle non destructif, le filtrage d'information textuelle, la bio- ingénierie, la formulation de matériaux nouveaux, la modélisation des procédés industriels, contrôle de l'environnement, l'interprétation des spectres, la classification des images (satellitaires), etc.

III.2 Historique

L'histoire de doter une machine d'intelligence est une idée qui fascine l'humanité depuis très longtemps. Dès l'Antiquité, les Grecs anciens rêvaient de robots. Selon la légende, Héphaïstos, le dieu du feu, aurait construit des tables à trois pieds munies de roulettes qui pouvaient aller et venir toutes seules dans les palais des dieux.

Pour les réseaux de neurones artificiels, tout commence en 1943, lorsque deux scientifique : le neuropsychiatre Warren S. McCulloch (1898-1972) et le mathématicien Walter Pitts (né en 1924) de l'Université de Chicago dans l'Illinois, conçoivent le premier modèle du neurone biologique, baptisé neurone formel ou automate à seuil (McCulloch W.S. et Pitts W., 1943). Un peu plus tard, en 1949, le neurophysiologiste Donald O. Hebb (1904-1985) publia ses travaux sur la « plasticité synaptique » qui sont à la base des mécanismes d'apprentissage. Il présente dans son ouvrage (*The Organization of Behavior*) une règle d'apprentissage qui porte encore son nom. En 1957, Frank Rosenblatt développa le *Perceptron* au Laboratoire d'Aéronautique de l'Université de Cornell dans le but de comprendre la mémoire humaine, l'apprentissage et les processus cognitifs. Le perceptron est un réseau de neurones artificiels inspiré du système visuel, possédant une couche de neurones "perceptive" et une couche de neurones "décisionnelle" (Rosenblatt F. 1962). Sa mise au point (en 1962) fut le départ des premières machines adaptatives. Ce réseau, qui parvient à apprendre à identifier des formes simples et à calculer certaines fonctions logiques, constitue le premier système artificiel exhibant une faculté jusque là réservée au vivant, la capacité d'apprendre par l'expérience (l'apprentissage). Dans la même période, en 1960, Bernards Widrow, de Stanford, conçut *l'Adaline* (Adaptative Linear Element), une machine basée sur le principe des composants électriques et utilisée pour le traitement du signal. Ce modèle sera par la suite le modèle de base des réseaux multicouche.

En 1969, Marvin Minsky et Seymour Papert, du Massachusetts Institute of Technology, publièrent « Perceptrons », une analyse très critique démontrant les limites des machines de type Perceptron et du connexionniste comme modèle de la cognition. Ils montrèrent que les réseaux de neurones comme le Perceptron n'étaient même pas capables de comprendre la fonction logique XOR (le « ou » exclusif). La neuro-informatique vit fondre ses crédits de recherche au profit d'autres domaines comme l'étude des systèmes experts.

Il faut attendre le début des années 80 pour assister à la renaissance de la recherche dans le domaine des RNA. John J. Hopfield, du California Institute of Technology, démontra en 1982 dans une publication de l'Académie des sciences (Hopfield J.J. 1982) l'utilité des réseaux complètement

connectés basés sur la règle de *Hebb*, qui ont permis de définir les notions d'attracteurs et de mémoire associative.

En1984, le Finlandais Tuevo Kohonen proposa un modèle de « cartes autoorganisatrices » reposant sur le principe des cartes cérébrales. En 1986, *D.E.Rumelhart* décrit dans son ouvrage "Parallel distributed processing" (Rumelhart D.E, et al. 1986). l'algorithme d'apprentissage dit de rétro propagation de l'erreur qui fournit un moyen simple d'entrainer les neurones des couches cachées. Vers la fin des années 80, les réseaux de neurones sont devenus un domaine de recherche attractif, qui réunit autour d'une thématique commune des mathématiciens, des psychologues, des physiciens, des biologistes, des ingénieurs et des informaticiens.

III.3 Le Neurone biologique

Le neurone est l'unité fonctionnelle de base du système nerveux. D'une espèce à une autre, il peut présenter des différences d'ordre fonctionnels ou anatomiques, cependant divers points communs subsistent, et sont à la base de la cellule nerveuse.

Le neurone biologique est donc une cellule composée d'un corps cellulaire et d'un noyau, figure III.1. Le corps cellulaire se ramifie pour former ce que l'on nomme les dendrites. Celles-ci sont parfois si nombreuses que l'on parle alors de chevelure dendritique ou d'arborisation dendritique. C'est par les dendrites que l'information est acheminée de l'extérieur vers le soma, corps du neurone. L'information traitée par le neurone chemine ensuite le long de l'axone (unique) pour être transmise aux autres neurones. La transmission entre deux neurones n'est pas directe. En fait, il existe un espace intercellulaire de quelques dizaines d'Angstroms (10^{-9} m) entre l'axone du neurone presynaptique et les dendrites du neurone postsynaptique. La jonction entre deux neurones est appelée la synapse.

Les messages nerveux arrivant à une cellule se manifestent par des modifications locales de potentiels électriques entre l'intérieur et l'extérieur de la cellule. Lorsque ces messages parviennent en quantité suffisante, ils produisent un changement de potentiel (une dépolarisation) qui est transmis le long de l'axone et le message (chimique) est émis à d'autres cellules par les synapses. Le neurone émettant le signal est appelé neurone présynaptique et celui recevant ce signal, neurone postsynaptique. Le cerveau se compose d'environ 10^{12} neurones (mille milliards), avec 1000 à 10000 synapses (connexions) par neurone. Elles forment un gigantesque et inextricable réseau de neurones qui étend ses prolongements à tout notre corps (Hassani D. 2002-2003).

III.3.1 Dendrites (pôles récepteurs)

Cette partie qui présente une sorte de ramification appelée arbre dendritique, recueille, via les synapses, les signaux électrochimiques ou les potentiels d'action venant d'autres cellules ou de l'extérieur et les achemine vers le corps cellulaire ou le soma. Les dendrites mesurent de quelques microns à 1,50 mètres pour un neurone sensoriel de la moelle épinière.

III.3.2 Soma

Le soma, outre son rôle concernant le métabolisme de la cellule, collecte et concentre les informations et en fait une sommation dite «spatio-temporelle». Si le potentiel somatique dépasse un certain seuil, il y a émission d'un potentiel d'action appelé spike, qui correspond à une oscillation électrique très bref (environ 1ms).

III.3.3 Axone (pôle émetteur)

L'axone est un prolongement cellulaire chargé de faire passer l'information nerveuse à d'autres cellules à travers les synapses grâce à son arborisation terminale. Selon le type du neurone, la longueur de l'axone peut varier de quelques microns à 1,50 mètres pour un motoneurone.

III.3.4 Synapses

Les synapses se trouvent au point de contact entre neurones, elles convertissent les impulsions nerveuses en signaux chimiques c'est-à-dire en molécules de neurotransmetteur. En effet, l'arrivée d'un potentiel d'action à l'extrémité d'un axone entraine le passage d'une quantité de neurotransmetteurs dans l'intervalle synaptique. Ces neurotransmetteurs dépolarisent alors la membrane post-synaptique d'une dendrite ou les hyper-polarisent dans le cas des synaptiques inhibiteurs. Ils permettent alors un passage pondéré de l'influx nerveux. Le nombre de synapses par neurone varie aussi considérablement de plusieurs centaines à une dizaine de milliers.

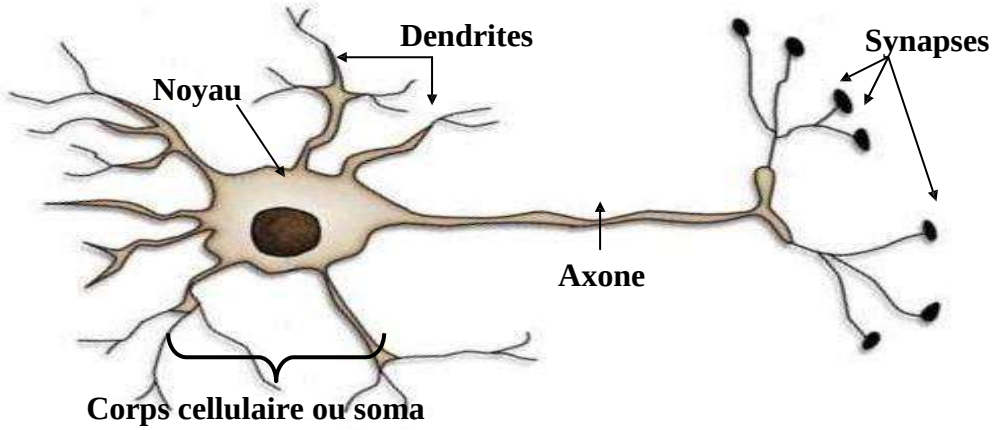


Figure III.1 : Schématisation d'un neurone biologique.(Hassani D. 2002-2003)

III.4 Neurone formel

La première fois qu'est apparu la notion de neurone formel, c'est en 1943 dans l'article " A logical Calculus of the ideas immanent in nervous activity " (McCulloch W.S. et Pitts W. 1943) paru dans le journal Bulletin of mathematical. A l'époque, Mc Culloch étudie la strychnisation du cerveau des chimpanzés. Il se convaint que les circuits d'influx nerveux étaient analogues à la propagation logique d'un raisonnement.

Le fonctionnement est simple si aucune synapse inhibitrice n'est active, le neurone ajoute ses entrées synaptiques et vérifie que la somme atteint ou excède son seuil. On dit alors que le neurone devient actif, sinon le neurone est inactif (McCulloch W.S. et Pitts W. 1943) .
Ainsi un neurone formel ou artificiel, aussi fréquemment appelé neurone de *McCulloch* et *Pitts*, est un opérateur mathématique très simple qui traduit la représentation mathématique d'un neurone biologique. Selon, G. Dreyfus (Dreyfus G et al. 2002) : "un neurone est une fonction algébrique non linéaire, paramétrée, à valeur bornée". Il possède des entrées qui peuvent être les sorties d'autres neurones, ou des entrées de signaux extérieurs, et une sortie. La valeur de la sortie résulte du calcul de la somme des entrées, pondérées par des coefficients (dits poids synaptiques) et du calcul d'une fonction non linéaire bornée de cette somme pondérée (figure III.2) ; cette fonction (dite fonction d'activation) est le plus souvent une fonction sigmoïde, mais peut également être une gaussienne, un échelon, etc (Duprat A, et al. 1998).

Un neurone formel peut être défini par les cinq éléments suivants (Davalo E. et Naïm P. 1993) :

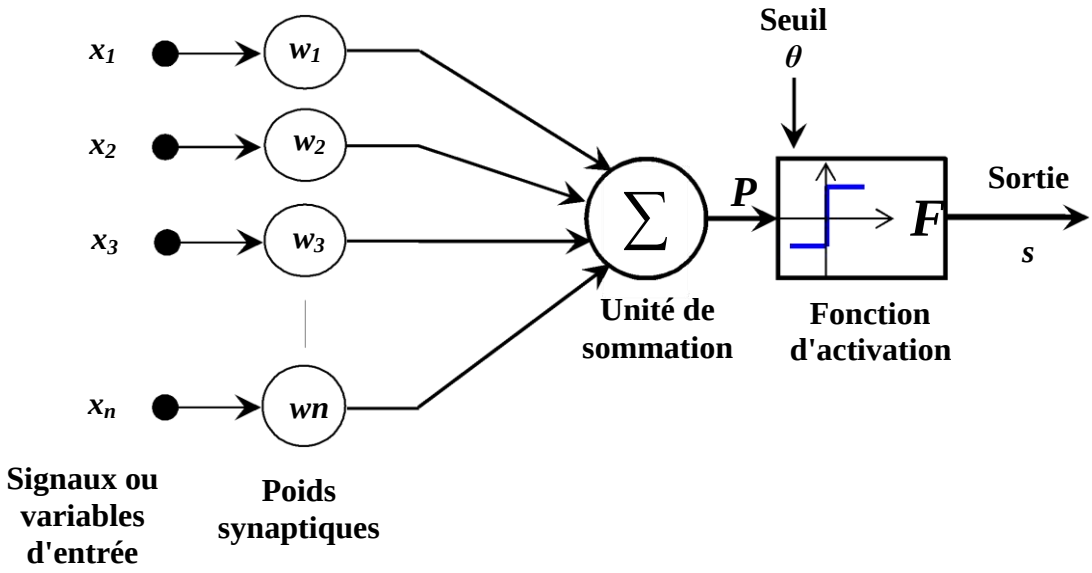


Figure III.2 : Neurone formel.(Duprat A, et al. 1998)

III.4.1 Entrées

Typiquement, les valeurs d'entrée sont des stimuli externes en provenance de l'environnement ou des sorties d'autres neurones artificiels. Elles peuvent être discrétisées grâce à un ensemble tel que {0,1} ou bien être des valeurs réelles (Hassani D. 2002-2003).

III.4.2 Poids synaptiques

Les paramètres w_n (figure III.2) sont souvent désignés sous le nom de «poids» ou, en raison de l'inspiration biologique des réseaux de neurones, «poids synaptiques». Ils sont des valeurs réelles qui déterminent la contribution de chaque entrée. Un algorithme d'apprentissage permet de déterminer le meilleur ensemble de poids pour le problème considéré. Trouver l'ensemble optimal est souvent un compromis entre temps de calcul, minimisation de l'erreur et conservation de la capacité de généralisation pour le réseau.

La première chose que fait un neurone formel est de calculer la somme pondérée de ses entrées. Conformément à l'usage (également inspiré par la biologie), cette somme pondérée est appelée potentiel, à laquelle s'ajoute souvent un terme constant ou «biais» :

$$P = b + \sum_{i=1}^{n-1} w_i x_i \qquad \text{(III.1)}$$

III.4.3 Seuil

Le seuil est un nombre réel qui est soustrait de la somme pondérée des valeurs d'entrée. Pour des raisons de simplicité, le seuil peut être vu comme une entrée supplémentaire associée à un poids, où ($w = \theta$) et ($x = -1$), tel qu'il est illustré par la figure suivante :

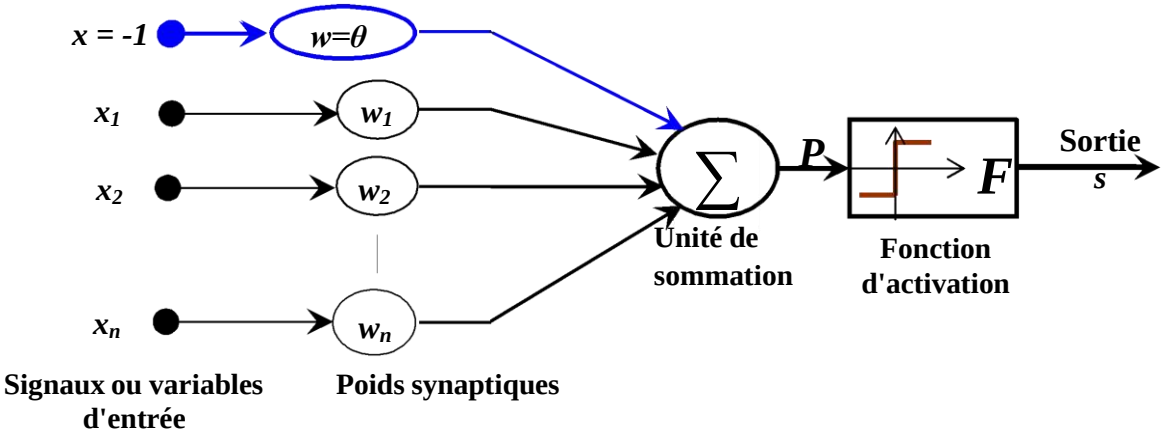


Figure III.3 : Neurone formel modèle étendu.(Hassani D. 2002-2003).

III.4.4 Fonction d'activation

La fonction de transfert ou d'activation calcule la valeur de l'état du neurone. C'est cette valeur qui sera transmise aux neurones avals. Il existe de nombreuses formes possibles pour la fonction de transfert. A 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].

Nous décrivons en ce qui suit, les formes les plus courantes des fonctions de transfert (Davalo E. et Naïm P. 1993).

III.4.4.1 Fonction binaire à seuil

Elle a été utilisée par *McCulloch* et *Pitts* dans leur modèle du neurone formel (McCulloch W.S. et Pitts W. 1943). Elle fait passer l'activation du neurone d'une valeur à une autre des que l'entrée résultante dépasse un certain seuil (équation III.2). Le seuillage introduit une non linéarité dans le comportement du neurone. Cependant, il limite la gamme de ses réponses possibles à deux valeurs (figure III.4). L'inconvénient de cette fonction est qu'elle n'est pas différentiable, ce qui pose un problème pour les algorithmes basés sur le gradient. L'équation qui définit l'état du neurone est :

$$S = F(P - \theta) = F(\sum_{i=1}^{n} w_i x_i - \theta) \qquad \text{(III.2)}$$

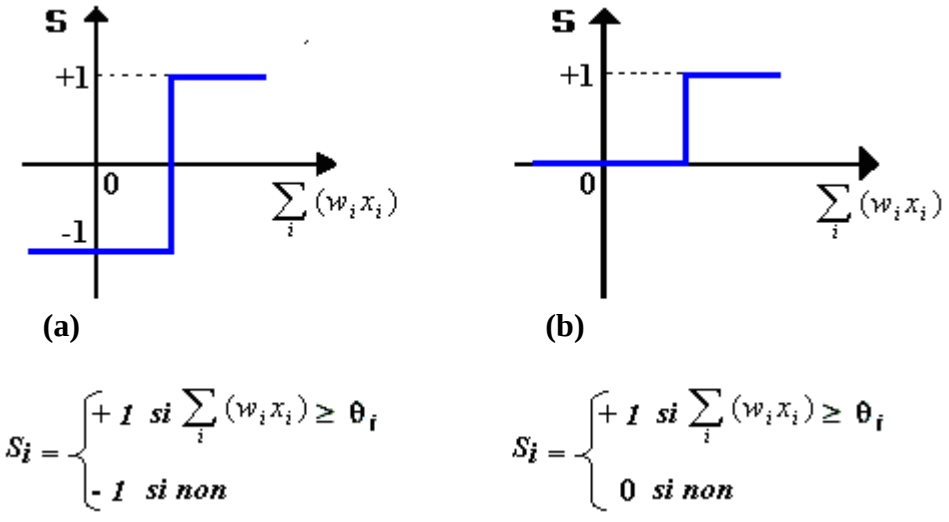


Figure III.4 : Fonction binaire à seuil (McCulloch W.S. et Pitts W. 1943)

III.4.4.2 Fonction identité

La fonction identité dite aussi linéaire dont l'expression est la suivante ;

$$S = F(P = \sum_{i=1}^{n} w_i x_i) \qquad \text{(III.3)}$$

Admet des valeurs non bornées pour les états, ce qui peut entrainer des débordements lors des simulations. Elle est employée souvent pour les neurones de sortie dans les réseaux multicouches. *Kohonen* l'a également utilisé pour construire son modèle de mémoires associatives (Kohonen T. 1984)

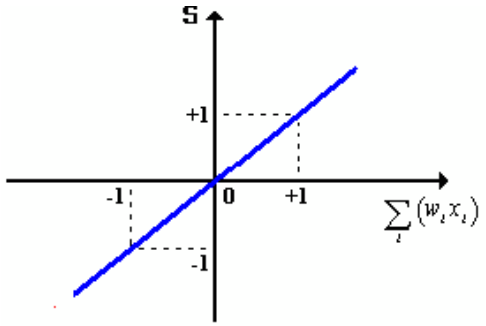


Figure III.5 : Fonction identité(Kohonen T. 1984)

III.4.4.3 Fonction linéaire à seuil

Cette fonction, qui est aussi appelée linéaire par morceaux, représente un compromis entre la fonction linéaire et la fonction seuil : entre ses deux bornes de saturation, elle confère au neurone une gamme de réponses possibles. En modulant la pente de linéarité, on affecte la plage de réponses du neurone (figure III.6).

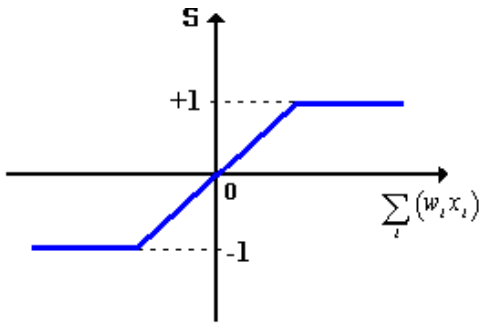


Figure III.6 : Fonction linéaire à seuil (Kohonen T. 1984)

III.4.4.4 Fonction non linéaire

a. Fonction sigmoïde logarithmique

La fonction sigmoïde est l'équivalent continu de la fonction linéaire bornée. Vu ses nombreux avantages (continuité et différentiabilité), ce type de fonction est généralement employé dans le réseau de neurones multicouches (figure III.7). L'équation qui définit la sigmoïde logarithmique est :

$$F(P) = \frac{1}{1 + e^{-P}} \qquad \text{(III.4)}$$

Pour $0 < F(P) < +1$

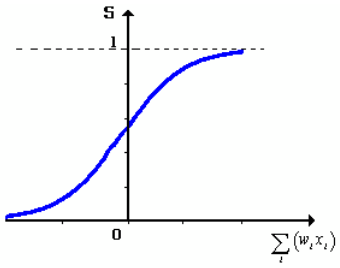


Figure III.7 : Fonction sigmoïde logarithmique (Kohonen T. 1984)

b. Fonction sigmoïde tangentielle

La fonction sigmoïde tangentielle ou tangente hyperbolique est fort utilisée dans les réseaux multicouches, surtout dans les neurones de la couche cachée (figure III.8). L'équation qui définit l'état du neurone est :

$$F(P) = \frac{1 - e^{-2P}}{1 + e^{-2P}} \qquad \text{(III.5)}$$

Pour $-1 < F(P) < +1$

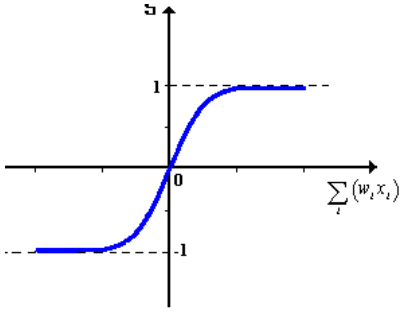


Figure III.8 : Fonction sigmoïde tangentielle (Kohonen T. 1984)

III.4.4.5 Fonction Gaussienne

Cette fonction n'est pas monotone croissante, elle est souvent employé dans les réseaux RBF (Radial Basis Function).

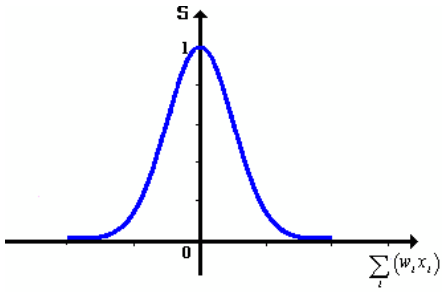


Figure III.9 : Fonction Gaussienne(Kohonen T. 1984)

Dans une étude comparative de fonctions d'activation B. (DasGupta G. et al. 1993), la fonction sigmoïde s'est révélée meilleure en terme de capacité d'approximation. H.N.Mhaskar et C.A.Micchelli (Mhaskar H.N. et al. 1994), ont établi une relation entre la précision de l'approximation, le nombre d'unités cachées dans un réseau à une seule couche cachée et la régularité de la fonction d'activation. D'après leur théorème, plus la fonction d'activation est régulière, plus la précision de l'approximation est bonne. Cependant, Fombellida et al. (Fombellida M.R.J. et al. 1990). ont constaté que l'utilisation de fonctions d'activation non monotones peut présenter certains avantages : accélération de l'apprentissage, réduction du nombre d'unités cachées, etc.

III.4.5 Sortie du neurone "*S*"

La sortie du neurone ni d'autre que l'application de la fonction de transfert à la somme pondérée de ces entrées selon l'équation suivante :

$$S = F\left(\sum_{i=1}^{n} w_i x_i - \theta\right) \qquad \text{(III.6)}$$

La sortie calculée, qui peut être discrète ou réelle selon la fonction d'activation utilisée, passe aux autres neurones ou directement à l'environnement extérieur.

III.5 Étude comparative

La figure suivante fait le parallélisme entre le neurone formel ou artificiel et le neurone biologique. On remarque dans cette description schématique, que chaque composant ou constituant du neurone formel trouve son inspiration biologique.

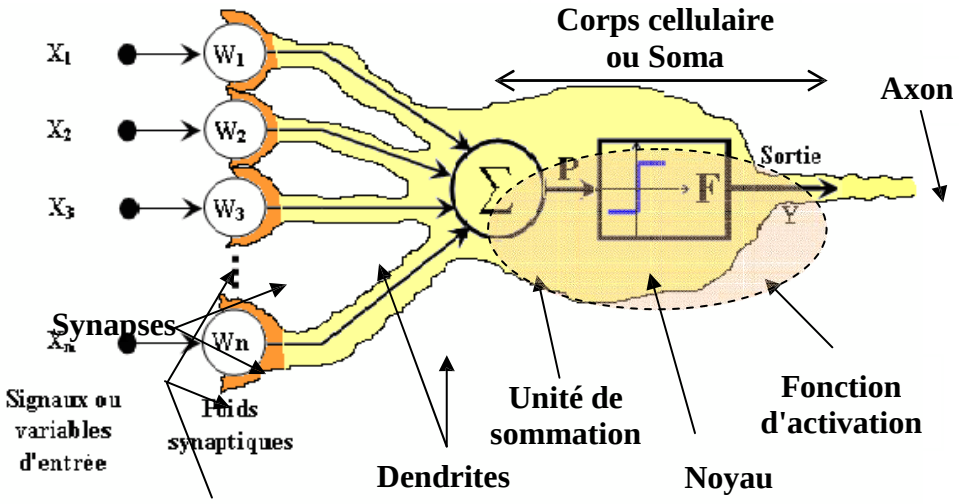


Figure III.10 : comparaison entre un neurone biologique et un neurone formel.

III.6 Architecture et connexionnisme des réseaux de neurones

Comme nous l'avons vu plus haut, un neurone réalise simplement une fonction non linéaire, paramétrée, de ses variables d'entrée. L'intérêt des neurones réside dans les propriétés qui résultent de leur association en réseaux, c'est-à-dire de la composition des fonctions non linéaires réalisées par chacun des neurones.

Les connexions entre les neurones qui composent le réseau décrivent la « topologie » du modèle. Le plus souvent, cette topologie fait apparaitre une certaine régularité de l'arrangement des neurones ; cependant, celui-ci peut être quelconque.

Selon le mode de connexion des neurones, On distingue les réseaux suivants :

III.6.1 Réseau de neurones non bouclés

La définition de ce type de réseaux selon G. Dreyfus (Dreyfus G et al. 2002) est : "Un réseau de neurones non bouclé réalise une (ou plusieurs) fonction algébriques de ses entrées, par composition des fonctions réalisées par chacun des neurones".

Un réseau de neurones non bouclé est donc représenté graphiquement par un ensemble de neurones connectés entre eux, l'information circulant des entrées vers les sorties sans retour en arrière : si l'on se déplace dans le réseaux, à partir d'un neurone quelconque, en suivant les connexions, on ne peut pas revenir au neurone de départ. Les neurones qui effectuent le dernier calcul de la composition des fonctions sont les neurones de sortie ; ceux qui effectuent des calculs intermédiaires sont les neurones cachés.

Il convient d'insister sur le fait que le temps ne joue aucun rôle fonctionnel dans le réseau de neurones non bouclé : si les entrées sont constantes, les sorties le sont également. Le temps nécessaire pour le calcul de la fonction réalisée par chaque neurone est négligeable et, fonctionnellement, on peut considérer ce calcul comme instantané. Pour cette raison, les réseaux non bouclés sont souvent appelés "réseau statique", par opposition aux réseaux bouclés ou dynamique que nous introduirons plus loin.

Le domaine d'application par excellence des réseaux non bouclé, dont l'apprentissage est souvent supervisé, est la modélisation statique type boite noire. En effet, si on dispose d'un ensemble de variables mesurées et d'un ensemble de mesures d'une grandeur relative à un processus de nature quelconque (physique, chimique, économique, financier…) et en supposant qu'il existe une relation (suffisamment régulière) entre le vecteur des variables et la grandeur en question, alors il est possible d'approximer cette relation par un réseau de neurone non bouclé.

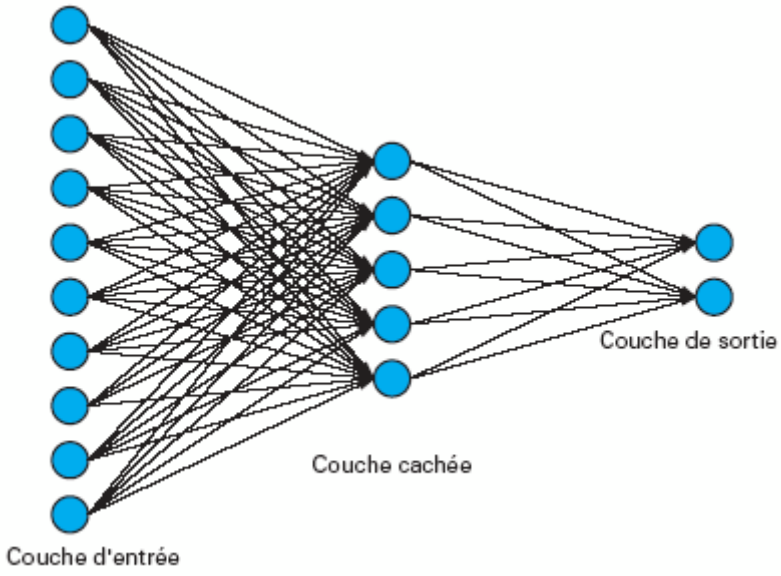


Figure III.11 Réseau multicouche non bouclé classique (Dreyfus G et al. 2002)

Selon le nombre de couches dans un réseau et la fonction d'activation, on distingue :

III.6.1.1 Perceptron

C'est l'un des premiers réseaux de neurones, conçu en 1958 par F.Rosenblatt (Rosenblatt F., 1962). Il est linéaire et représente l'unité élémentaire d'un réseau de neurones. Le Perceptron a été inspiré du système visuel. La première couche ou couche d'entrée représente la rétine, et la couche de sortie est équivalente aux cellules de décision.

Un Perceptron calcule à partir d'un vecteur de *n* entrées réelles [x_1 x_2... x_n], une combinaison linéaire de ces entrées ; le résultat *P* de cette combinaison est ensuite utilisé comme argument d'une fonction *F* d'une variable pour délivrer la valeur de sortie *s* (Figure III.12). On a donc la relation :

$$s = F(P) = F\left(b + \sum_{i=1}^{n} x_i w_i\right) \qquad \text{(III.7)}$$

dans laquelle les (ω_i) $i \leq n$ appelés poids représentent la contribution des entrées (x_i) $i \leq n$ sur la sortie du perceptron. La fonction *F* est appelée fonction d'activation du neurone.

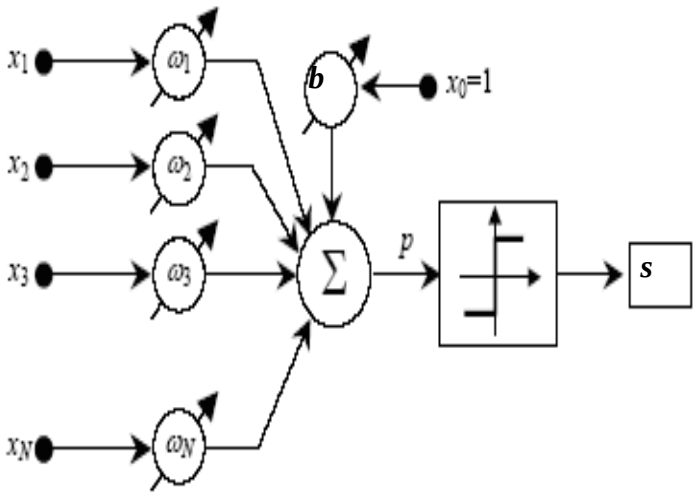


Figure III.12 : Principe d'un perceptron Multicouche (Rosenblatt F. 1962)

III.6.1.2 Perceptron multicouche (MLP)

Les réseaux neurones multicouches contiennent en plus des couches d'entrée et de sortie, un certain nombre (supérieur ou égal à un) de couches intermédiaires ou cachées dont les neurones ont une fonction d'activation sigmoïde.

Les liens n'existent évidemment qu'entre les cellules d'une couche avec les cellules de la couche suivante (réseau non bouclé). De telles architectures sont habituellement désignées dans la littérature sous le nom "**M**ulti-**L**ayer **P**erceptron" MLP (figure III.11) (Dreyfus G et al. 2002).

Les MLP possédant un seul neurone de sortie linéaire, peuvent approcher toute fonction bornée suffisamment régulière, avec une précision arbitraire, dans un domaine fini de l'espace de ses variables (Hornik K. 1991). Néanmoins, cette propriété, qui n'est qu'un théorème d'existence et ne donne pas la méthode de trouver les paramètres du réseau, n'est pas spécifique aux réseaux de neurones. C'est la propriété de parcimonie qui leur est particulière et fait tout leur intérêt.

Les MLP sont, en effet, des approximateurs parcimonieux, c'est-à-dire ; que le nombre des paramètres de pondération ou poids, pour une précision donnée, croit linéairement avec le nombre des variables d'entrées du modèle. La parcimonie, qui est une propriété fondamentale des MLP, est donc d'autant plus importante que le nombre des entées d'un modèle est grand (Dreyfus G et al. 2002)

III.6.1.3 Réseaux RBF (Radial Basis Function)

Ces réseaux ont été introduits par M.Powel (Powe M.l. 1985). et D.Broomhead et al. (Broomhead D. et Lowe D. 1988). Il s'agit d'une "spécialisation" d'un perceptron multicouche (MLP). Les neurones de ces réseaux sont actifs pour une zone restreinte de l'espace d'entrée qui est définie par des fonctions radiales. Cette technique consiste à utiliser la non linéarité d'une fonction (par exemple une gaussienne) pour ponctuer la réponse des neurones des couches cachées (figure III.13) (Hassani D. 2002-2003, et al. 1995).

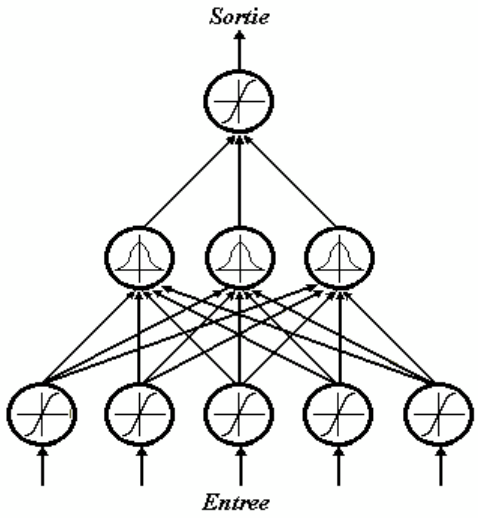


Figure III.13 : Réseau Radial Basis Function (RBF) (Hassani D. 2002-2003, et al. 1995)

III.6.2 Réseaux de neurones bouclés

Un réseau de neurone et dit bouclé ou récurrent lorsque des sorties de neurones d'une couche sont connectées aux entrées d'une couche d'indice inférieur**.** Autrement dit, lorsqu'on se déplace dans le réseau en suivant le sens des connexions, il est possible de trouver au moins un chemin qui revient à son point de départ. La sortie d'un neurone du réseau peut donc être fonction d'elle-même ; cela n'est évidemment concevable que si la notion de temps est explicitement prise en considération. Raison pour la quelle ce type de réseaux sont appelés dynamique.

Ainsi, à chaque connexion d'un réseau de neurones bouclé est attaché, outre un poids comme pour les réseaux non bouclé, un retard, multiple entier de l'unité de temps choisie (Dreyfus G et al. 2002). Une grandeur, à un instant donné, ne pouvant pas être fonction de sa propre valeur au même instant, tout cycle du graphe du réseau doit avoir un retard non nul.

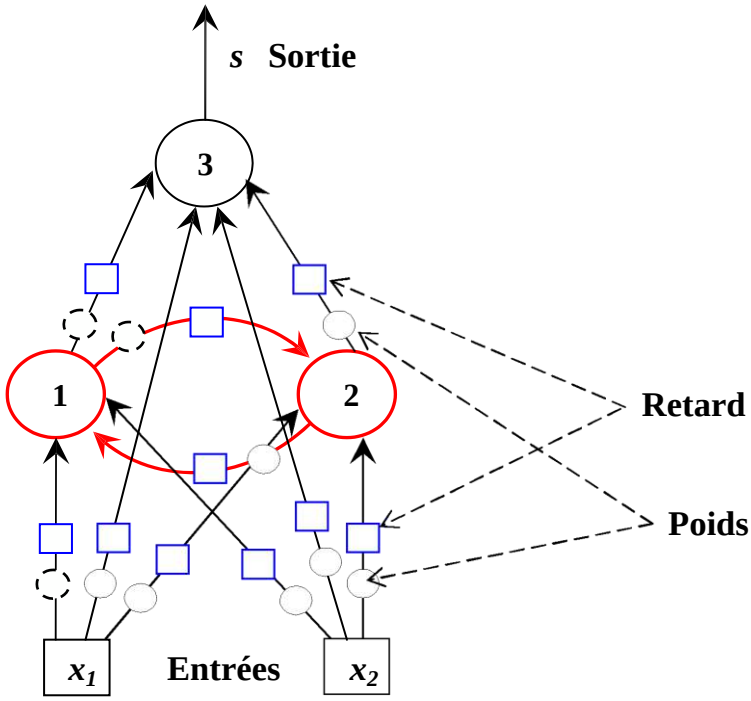


Figure III. 14 : Un réseau de neurone bouclé à deux entrées contenant une boucle ente les neurones (1) et (2). (Dreyfus G et al. 2002).

Les réseaux de neurones bouclés sont mise en œuvre essentiellement pour concevoir des modelés dynamiques. Ils peuvent être utilisés à titre d'exemples ; comme modèle de simulation pour prévoir l'évolution d'un processus quelconque dans le temps ou comme prédicteur à très court terme (une période d'échantillonnage) afin de l'intégrer à un système de commande(Dreyfus G et al. 2002).

Selon la complexité de son connexionnisme, on distingue :

III.6.2.1 Modèle de Hopfield

Le modèle élaboré par *J.Hopfield* (Hopfield J.J. 1982), est totalement interconnecté c'est le réseau bouclé par excellence (figure III.30). Il n'y a plus de notions de couche, du fait que chaque neurone du réseau est a la fois une entrée et une sortie, le réseau implantant ainsi une mémoire associative (adressables par le contenu), c'est-à-dire qu'on lui donne des exemples et il doit être capable de les reconnaitre lorsqu'ils sont bruités. La question qui se pose alors est : quelle est la capacité de stockage d'un tel réseau ? Des recherches ont prouvé que le rapport optimal entre le nombre d'exemples et le nombre de neurones est de 0,14 (Buniet L. 1997).

Dés que le nombre d'exemples devient trop grand, le réseau oublie tout ce qu'il a appris. Le processus d'apprentissage fait appel à la règle de *Hebb* (Bottou L. 1991 et Hebb D, 1949). qui renforce la connexion entre deux neurones s'ils sont actifs simultanément. Ce type de réseau est utilisé essentiellement dans la reconnaissance de la parole et la reconnaissance des formes (Bottou L. 1991) .

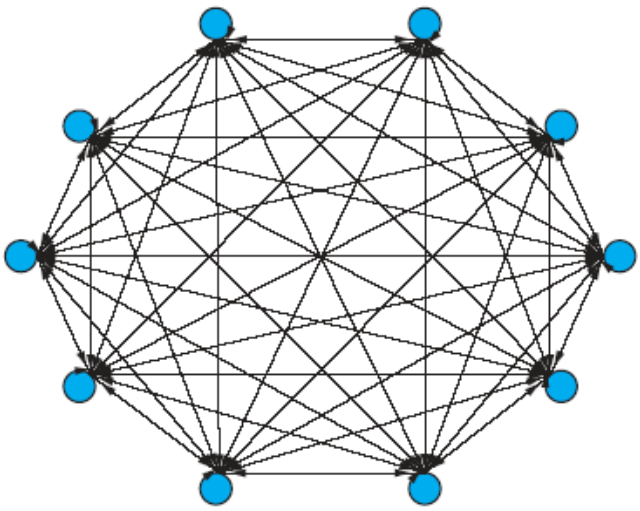


Figure III.15 : Modèle de Hopfield (Bottou L. 1991)

III.6.2.2 Modèle de Boltzmann

La Machine de Boltzmann qui a été introduite en 1984 par *G.E.Hinton et al*.(Hinton G.E., Sejnowski T.J. et Ackley D.H. 1984). est le premier modèle connu apte à traiter de manière satisfaisante les limitations recensées dans le cas du perceptron. Il s'agit d'un réseau fortement récurrent au même titre que les réseaux de *Hopfield*. Mais à la différence de ces derniers, le réseau est divisé en trois parties distinctes et les entrées et sorties sont clairement séparées et identifiées(Abdi H. 1994).

L'utilisation pratique de ces réseaux s'avère difficile, la convergence de l'algorithme étant extrêmement longue (les temps de calcul sont considérables).

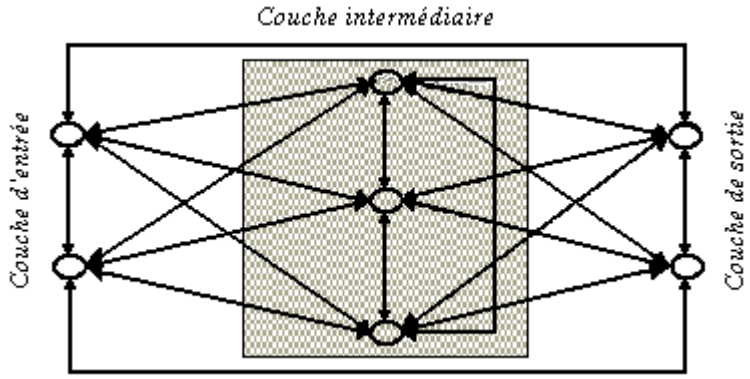


Figure III.16: Réseau de type machine de Boltzmann.**(**Hassani D. 2002-2003)

Il est souvent employé dans le domaine de reconnaissance de la parole ou de la vision (Bottou L. 1991)

III.6.2.3 Modèle de Kohonen

La carte de Kohonen (Kohonen T. 1984). Est en général à deux dimensions. Chaque neurone de la couche d'entrées est relié à chaque neurone de la carte de Kohonen (lien 1 vers tous). Chaque neurone de la carte de Kohonen est relié à tous les neurones de la carte. Les poids sont déterminés en fonction de la distance selon une fonction "chapeau mexicain" ou DOG (Difference of Gaussians). Ces réseaux fonctionnent dans leur très grande majorité dans un mode non supervisé. Il est donc relativement difficile de lui faire apprendre nos bases (Davalo E. et Naïm P. 1993).

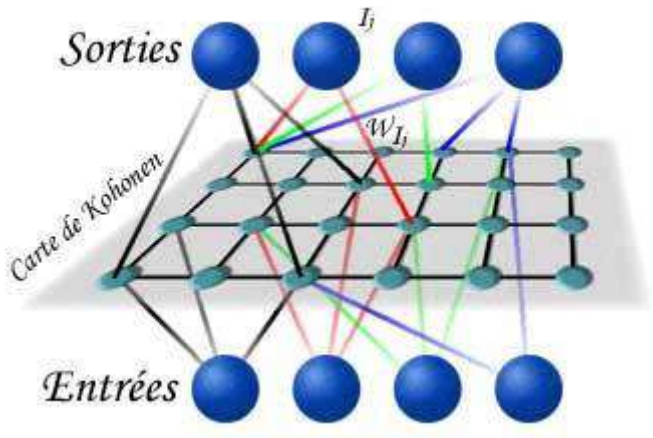


Figure III.17 : Architecture d'un modèle de Kohonen (Davalo E. et Naïm P. 1993)

Les réseaux de *Kohonen* trouvent leurs champs d'application dans plusieurs domaine tels : la classification, le traitement de l'image, la reconnaissance de la parole et l'optimisation (Bottou L. 1991)

III.7 Apprentissage

On appelle apprentissage des réseaux de neurones la procédure qui consiste à estimer les paramètres (poids) des neurones du réseau, afin que celui-ci remplisse au mieux la tâche qui lui est affectée(Dreyfus G et al. 2002). 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. L'apprentissage fait appel à des exemples de comportement du processus à modéliser (Hinton G. 1989).

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

III.7.1 Apprentissage supervisé

L'apprentissage supervisé admet l'existence d'un expert (ou professeur) qui possède une connaissance innée de l'environnement. Le rôle de l'expert est de fournir les informations relatives à l'environnement nécessaires à l'apprentissage du réseau, sous forme d'un ensemble d'exemples composés de stimuli auxquels sont associées des réponses désirées.

L'apprentissage supervisé exige donc que les exemples ou les données soient constitués de couples de valeurs du type : (valeur d'entrée, valeur de sortie désirée). Tout le problème de

l'apprentissage supervisé consiste, étant donné un vecteur d'apprentissage de n couples *(x, d) i = 1, 2, ..., n*, à déterminer le vecteur des poids des neurones capables de prédire le même vecteur de sortie à partir du même vecteur d'entrée.

Un apprentissage supervisé force donc le réseau à converger vers un état final précis, en même temps qu'on lui présente un motif (données).

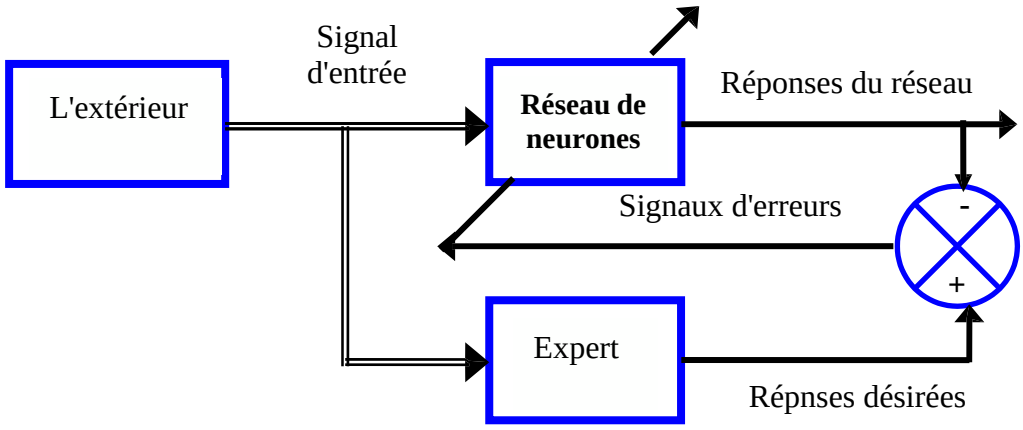


Figure III.18 : Illustration de l'apprentissage supervisé (Davalo E. et Naïm P. 1993)

III.7.2 Apprentissage non supervisé

Contrairement à l'apprentissage supervisé, lors d'un apprentissage non-supervisé, le réseau est laissé libre de converger vers n'importe quel état final lorsqu'on lui présente un motif. Il n'y a donc pas de professeur qui dirige le réseau vers une situation souhaitée, c'est au réseau de découvrir les ressemblances entre les éléments de la base de données, et de les traduire par des réponses adéquates. L'apprentissage est qualifié donc de non supervisé lorsque seules les valeurs d'entrée sont disponibles. 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.

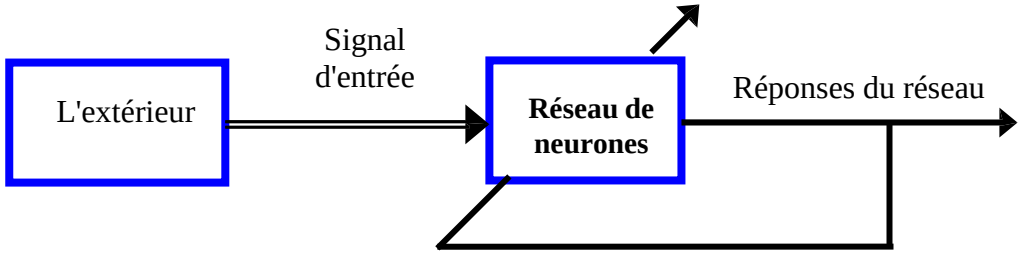


Figure III.19 : Illustration de l'apprentissage non supervisé (Davalo E. et Naïm P. 1993)

III.8 Règles d'apprentissage

III.8.1 Règle d'apprentissage de Hebb

Les réseaux neuronaux se sont développés grâce à leur capacité d'apprentissage. Le psychologue Hebb fût le premier à présenter un apprentissage à partir des neurones artificiels. "Lorsqu'un axone de la cellule A est suffisamment proche pour pouvoir exciter la cellule B et qu'il prend part de manière répétitive ou persistante à cette excitation, alors on doit trouver soit un phénomène de croissance, soit un changement métabolique dans l'une ou l'autre des cellules tel que l'efficacité de la cellule A pour exciter B doit être accrue." (Hebb D. 1949).

En d'autres termes, la synapse entre la cellule A et la cellule B doit faciliter la transmission de l'influx nerveux lorsque les deux sont excités en même temps, et au contraire si les cellules A et B ne sont pas activés sur la même excitation, la synapse doit inhiber cette transmission.
La règle de Hebb est utilisée dans un apprentissage supervisé. Son algorithme modifie de façon itérative les poids pour adapter la réponse obtenue à la réponse désirée. Il comprend les points suivants (Touzet C, 1992).

1. Présentation des matrices d'entrée X et des vecteurs de sortie désirée D de la base d'apprentissage;

2. Initialisation des poids w_i et du seuil θ à des valeurs petites, choisies au hasard, ainsi que μ le pas de modification des poids;

3. Calcul de la sortie S obtenue pour cette entrée :

$$S = F(x, w) = F\left(\sum_i (w_i x_i) - \theta\right) \tag{III.8}$$

$$F(x, w) = \begin{cases} +1 & si \quad \left(\sum_i (w_i x_i) - \theta\right) > 0 \\ -1 & si \quad non \end{cases} \tag{III.9}$$

4. Si la sortie S est différente de la sortie désirée D alors modification des poids :

$$w_i^{t+1} = w_i^t + \mu(x_i . d_i) \tag{III.10}$$

5. Tant que tous les exemples de la base d'apprentissage ne sont pas traités correctement, retour à l'étape 3.

Le perceptron est donc capable de réaliser une partition de son espace d'entrée en deux classes (1 et 2), selon la valeur de sa sortie (+1 ou -1). La séparation entre ces deux classes étant donnée par un hyperplan (figure III.20). L'équation de la droite séparatrice appelée encore droite de décision est (Touzet C. 1992) :

$$\sum_i (w_i x_i) - \theta = 0 \qquad \text{(III.11)}$$

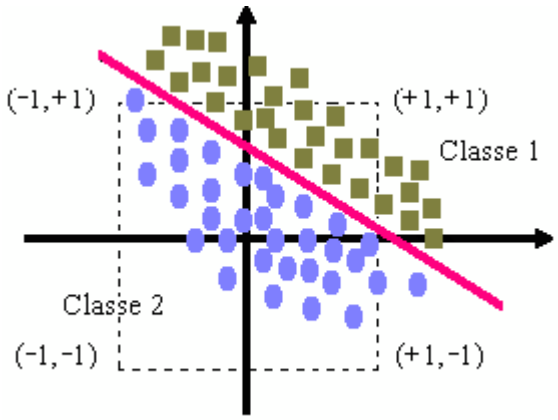


Figure III.20 : Séparation des exemples d'apprentissage par la droite de décision (Touzet C. 1992)

III.8.2 Algorithme de rétro propagation du gradient

L'algorithme de rétro propagation du gradient est certainement à la base des premiers succès des réseaux de neurones multicouches. Il figure parmi les algorithmes d'apprentissage les plus utilisés. Il s'agit toujours de minimiser l'erreur quadratique, ce qui est assez simple quand on utilise une fonction d'activation dérivable (la sigmoïde par exemple). On propage la modification des poids de la couche de sortie jusqu'à la couche d'entrée (Elie F. 1997).

L'apprentissage par rétro propagation du gradient d'un perceptron multicouches peut donc se décomposer principalement en quatre phases :

- la propagation du signal de la couche d'entrée jusqu'à la couche de sortie,
- le calcul de l'erreur en sortie,

- la rétro propagation de l'erreur;
- correction des poids synaptiques.

Avant tout apprentissage, il est indispensable de normaliser et de centrer toutes les variables d'entrées. Il faut aussi, qu'au lancement de l'apprentissage, les valeurs initiales des poids doivent être différentes de zéro pour que l'algorithme de rétro propagation puisse fonctionner. D'autre part, l'utilisation de valeurs élevées peut provoquer un phénomène de saturation prématurée qui contribue à diminuer la vitesse de convergence de l'apprentissage (Lee Y, et al. l 991).

III.8.3 Autres règles d'apprentissage

Plusieurs autres règles d'apprentissage ont été développées. Nous en citons quelque unes :

- Règle d'apprentissage de Windrow-Hoff (ADALINE)
- Règle d'apprentissage du perceptron de F. Rosenblatt
- Méthode de Newton
- Méthode de quasi-Newton
- Méthode Levenberg-Marquardt

III.9 Théorie de la généralisation

La généralisation concerne la tâche accomplie par le réseau une fois son apprentissage achevé (Marquardt D.W. 1963). Elle peut être évaluée en testant le réseau sur des données qui n'ont pas servi à l'apprentissage. Elle est influencée essentiellement par quatre facteurs : la complexité du problème, l'algorithme d'apprentissage, la complexité de l'échantillon (le nombre d'exemples et la manière dont ils représentent le problème) et enfin la complexité du réseau (nombre de poids). La complexité du problème est déterminée en partie par sa nature même : on peut parler de "complexité intrinsèque". Par ailleurs, l'algorithme d'apprentissage influe sur la généralisation par son aptitude à trouver un minimum local assez profond, sinon le minimum global (Hassani D. 2002-2003).

Les courbes données par la figure III.21 représentent la modélisation de *g* par trois modèles qui différent par leur nombre de paramètres. On peut constater que le modèle ayant très peu de paramètres n'a pas assez de flexibilité pour réaliser un apprentissage correct des exemples d'apprentissage. Les erreurs d'apprentissage et de test sont toutes deux importantes : c'est la situation de sous-apprentissage (figure III.21.c). En revanche, le modèle constitué de nombreux paramètres, lisse parfaitement les exemples d'apprentissage. Il commet donc une erreur faible sur ces données, mais probablement une erreur plus importante sur les données de test. C'est la situation de sur-apprentissage (figure III.21.b). Finalement, le modèle possédant un nombre de paramètres modérés réalise un bon compromis entre précision d'apprentissage et bonne généralisation (figure III.21.a).

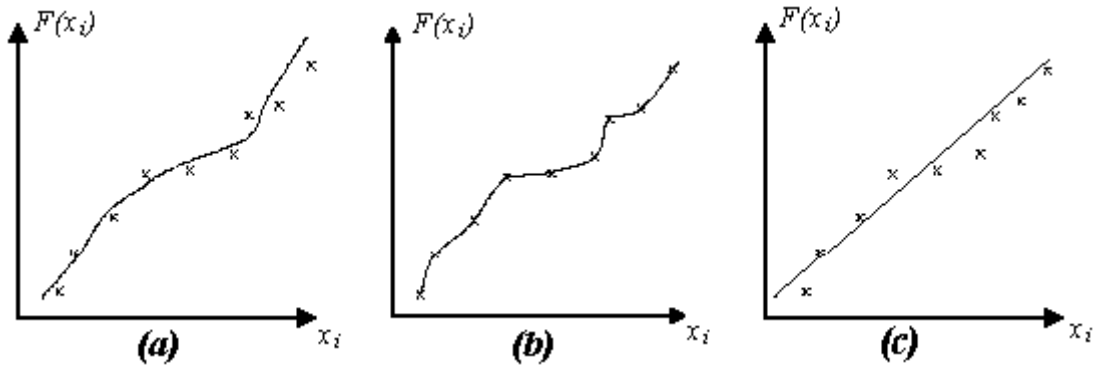


Figure III.21 : a. Apprentissage correct **b.** Sur-apprentissage

c. Sous-apprentissage **(**D. Hassani. 2002-2003)

Le problème de la généralisation est souvent vu sous deux perspectives différentes. Dans la première, la taille du réseau est fixée (en accord avec la complexité du problème) et la question est : combien d'exemples d'apprentissage sont nécessaires pour atteindre une bonne généralisation ? Cette perspective est intéressante dans les applications où l'on a la possibilité d'acquérir autant d'exemples que l'on veut. Dans le second cas, le nombre d'exemples d'apprentissage est fixé et la question est : quelle taille du réseau donne la meilleure généralisation pour ces données ? On est conduit à adopter ce point de vue lorsque l'on est limité dans la possibilité d'acquérir des données d'apprentissage : il importe alors de déterminer quelle est la taille permettant au réseau de décrire au mieux les données en notre possession (Hassani D. 2002-2003). Cependant une manière d'évaluer la faculté de généraliser la représentation construite par le réseau de neurones à toutes les données, y compris celles n'appartenant pas à l'ensemble d'apprentissage, consiste à mesurer les performances du réseau sur des données représentatives du problème non apprises.

Il s'agit d'une évaluation de l'erreur de généralisation. La différence entre l'erreur d'apprentissage et l'erreur de généralisation (figure III.26) représente une mesure de la qualité de l'apprentissage effectué (Thiria S, et al. 1997).

III.10 Critère d'arrêt pour l'apprentissage

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

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

Les méthodes qui évaluent l'erreur de généralisation sont presque toutes basées sur la partition de l'ensemble des données qu'on possède en plusieurs sous-ensembles. Par exemple un ensemble utilisé pour l'apprentissage et un ensemble de validation. L'ensemble de validation est utilisé pour contrôler et mesurer la généralisation du réseau. Pendant l'apprentissage l'erreur d'apprentissage décroît continuellement, tandis que sur l'ensemble de validation elle ne diminue que jusqu'à un certain point au-delà duquel elle augmente (Marquardt D.W. 1963).

A partir de ce point le réseau apprend par cœur les données de l'ensemble d'apprentissage et l'apprentissage doit être stoppé (figure III.22). Ces deux ensembles servent à déterminer l'architecture la plus appropriée pour le problème à traiter : pour différentes architectures (nombre de neurones cachés variable), on contrôle l'erreur de validation et on choisit l'architecture pour laquelle elle est minimale.

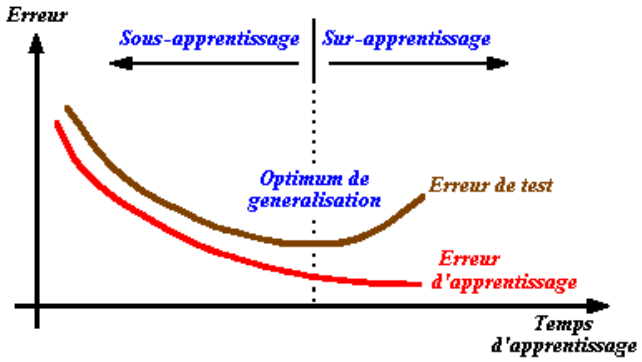


Figure III.22 : Evolution des erreurs d'apprentissage et de test au cours du temps (Hassani D. 2002-2003)

Parmi les méthodes utilisant cette méthodologie, on peut citer les méthodes appelées "split-samples","Cross-validation" : la ``split-samples'' réserve un troisième ensemble de données appelé ensemble de test, pour tester le réseau sur des données qui n'ont jamais été utilisées ni pour l'apprentissage ni pour la validation.

Pour la cross-validation l'ensemble des données de départ est découpé en *k* parties de taille égale. Le réseau est entrainé *k* fois, chaque fois en utilisant *(k-1)* parties pour l'apprentissage et la dernière pour la validation et le calcul des erreurs commises en généralisation. L'ouvrage de *R.Tibshirani* (Boucheron S. 1992) donne une comparaison de ces méthodes.

III.11 Procédure de modélisation neuronale

Le modèle boite noire constitue la forme la plus primitive de modèle mathématique : il est réalisé uniquement à partir de données expérimentales ou d'observations ; il peut avoir une valeur prédictive, dans un certain domaine de validité, mais il n'a aucune valeur explicative (Dreyfus G et al. 2002).

A l'opposé, un modèle de connaissance, que l'on pourrait qualifier de boite blanche, est issu de l'analyse des phénomènes physiques, chimiques, biologiques, etc., dont résulte la grandeur que l'on chercher à modéliser. Ces phénomènes sont mis en équation à l'aide des connaissances théoriques disponibles au moment de l'élaboration du modèle. La démarche scientifique s'efforce, dans la mesure du possible, de construire des modèles de connaissance : l'élaboration d'un modèle de connaissance nécessite l'existence d'une théorie, alors qu'un modèle boite noire requiert seulement la disponibilité de mesures.

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

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

Un modèle statique est régi exclusivement par des équations algébriques ; un modèle dynamique, en revanche, obéit à des équations différentielles et à des équations algébriques (Dreyfus G et al, 2002). Nous examinerons en bref, exclusivement, dans ce qui suit la conception

d'un modèle statique type boite noire, réalisé à l'aide de réseaux de neurones bouclés (MLP) du fait qu'il a un lien direct avec le thème de ce travail.

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

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

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

III.11.1 Sélection des données

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

- que la représentation des entrées soit aussi compacte que possible,
- que la représentation des entrées contienne tous les facteurs pertinents par rapport à la grandeur à modéliser, mais seulement ceux-ci. En effet, la présence de variables qui ont, sur la grandeur à modéliser, moins d'influence que les perturbations ou le bruit, entraine celle de paramètres superflus et introduit des variations non significatives sur les variables d'entrées, qui constituent autant de sources d'erreur pour le modèle (Dreyfus G et al. 2002).

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

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

III.11.2 Collecte des données

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

III.11.3 Séparation des bases de données

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

III.11.4 Normalisation des données

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

III.11.5 Choix du type de réseau de neurones

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

III.11.6 Choix du nombre de neurones dans la couche cachée

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

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

III.11.7 Choix de la fonction de transfert

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

III.11.8 Estimation des paramètres ou apprentissage

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

Le critère d'arrêt de l'apprentissage est souvent calculé à partir d'une fonction de coût, caractérisant l'écart entre les valeurs de sortie obtenues et les valeurs désirées. Ce qui permet un arrêt adéquat de l'apprentissage pour obtenir de bonnes performances de généralisation.
L'obtention des paramètres architecturaux optimaux se fera par comparaison des performances obtenues pour différentes architectures de réseaux de neurones (Elie F. 1997).

III.11.9 Validation et généralisation

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

de départ. Elle permet de savoir si le réseau a développé une capacité de généralisation assez performante.

III.12 Conclusion

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

CHAPITRE IV

MODELISATION NEURONALE DE LA TOXICITE DES INSECTICIDES

IV.1 Introduction

Dans ce chapitre nous présentons les résultats préliminaires de la modélisation de la toxicité des insecticides par les réseaux de neurones artificiels. Ce chapitre est divisé en deux parties : la première partie concerne l'élaboration de la base de données. Dans la deuxième partie on s'intéressera à la modélisation de la toxicité par les réseaux de neurones de type perceptron multicouches. Enfin on conclura ce chapitre en analysant et en interprétant les différents résultats obtenus.

IV.2 Méthodologie de calcul

La méthodologie de calcul par réseaux de neurones de la toxicité est illustrée par l'organigramme de la figure IV.1.

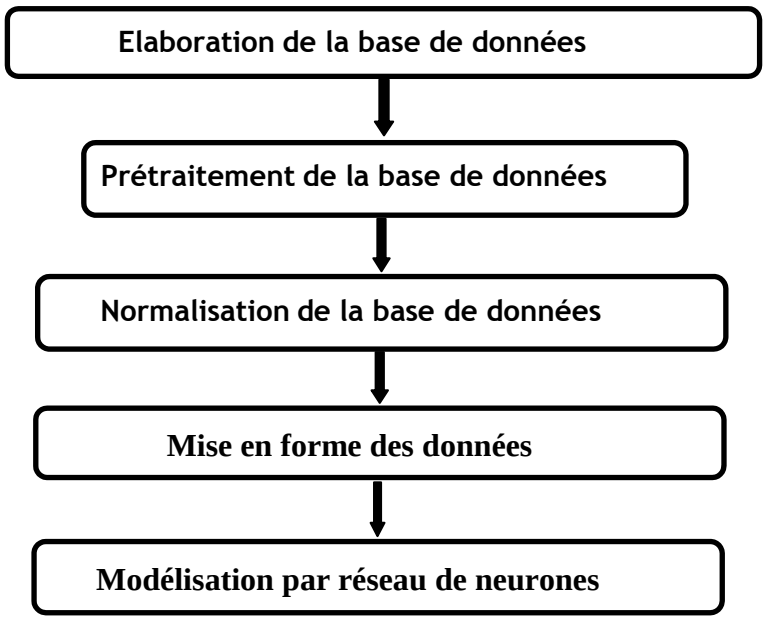


Figure IV.1 : Méthodologie de calcul de la toxicité par les réseaux de neurones artificiel

IV.2.1 Elaboration de la base de données

Lors de cette étape, nous avons sélectionné 141 molécules d'insecticides qui appartiennent à différentes classes de toxicité (Tableau IV.1).

Tableau IV.1 : Liste des insecticides utilisés dans cette étude

INSECTICIDES	NOTATION SMILES
Acéphate	O=C(NP(=O)(OC)SC)C
Acetamipride	Clc1ncc(cc1)CN(\C(=N\C#N)C)C
Acethion	P(SCC(OCC)=O)(=S)(OCC)OCC
Alanycarbe	O=C(OCC)CCN(SN(C(=O)O\N=C(/SC)C)C)Cc1ccccc1
Aldicarbe	O=C(O\N=C\C(SC)(C)C)NC
Aldoxycarbe	O=S(=O)(C(\C=N\OC(=O)NC)(C)C)C
Aldrine	Cl\C3=C(/Cl)C4(Cl)C2C(C1/C=C\C2C1)C3(Cl)C4(Cl)Cl
Allethrine	O=C2\C(=C(\C)C(OC(=O)C1C(/C=C(/C)C)C1(C)C)C2)C\C=C
Allyxycarb	c1(N(CC=C)CC=C)c(cc(OC(NC)=O)cc1C)C
Alphacypermethrine	Cl/C(Cl)=C/C3C(C(=O)OC(C#N)c2cccc(Oc1ccccc1)c2)C3(C)C
Alphaendosulfan	Cl\C1=C(/Cl)C3(Cl)C(Cl)(Cl)C1(Cl)C2C3COS(=O)OC2
Alphahexachlorocyclohexane	Cl[C@H]1[C@@H]([C@@H]([C@@H](Cl)[C@@H]([C@H]1Cl)Cl)Cl)Cl
Amidithion	P(SCC(NCCOC)=O)(OC)(OC)=S
Aminocarbe	O=C(Oc1cc(c(cc1)N(C)C)C)NC
Amitraze	N(=C/N(\C=N\c1ccc(cc1C)C)C)\c2ccc(cc2C)C
Aramite	c1(ccc(C(C)(C)C)cc1)OC[C@@H](O[S@@](OCCCl)=O)C
Azametiphos	O=P(OC)(OC)SCN1c2ncc(Cl)cc2OC1=O
Azidithion	S=P(SCc1nc(nc(n1)N)N)(OC)OC
Azinphos-ethyl	S=P(OCC)(OCC)SCN1\N=N/c2ccccc2C1=O
Azinphos-methyle	S=P(OC)(OC)SCN1\N=N/c2ccccc2C1=O

INSECTICIDES	NOTATION SMILES
Azoxybenzene	c1(\[N+](=N\c2ccccc2)[O-])ccccc1
Bamphéchlore	ClC2(Cl)CC1C(C2(C(Cl)C1Cl)CCl)(C(Cl)Cl)CCl
Bendiocarbe	O=C(Oc1cccc2OC(Oc12)(C)C)NC
Benfuracarbe	O=C(OCC)CCN(SN(C(=O)Oc2cccc1c2OC(C1)(C)C)C)C(C)C
Bensultap	O=S(=O)(SCC(N(C)C)CSS(=O)(=O)c1ccccc1)c2ccccc2
Betacypermethrine	Cl/C(Cl)=C/[C@H]3[C@@H](C(=O)O[C@H](C#N)c2cccc(Oc1ccccc1)c2)C3(C)C
Bifenthrine	O=C(OC(C)C)NNc1cc(ccc1OC)c2ccccc2
Bioallethrine	O=C2\C(=C(\C)C(OC(=O)C1C(/C=C(/C)C)C1(C)C)C2)C\C=C
Bromophos	Clc1cc(OP(=S)(OC)OC)c(Cl)cc1Br
Bromophos-éthyl	CCOP(=S)(OCC)OC1=CC(=C(C=C1Cl)Br)Cl
Butocarboxime	O=C(O\N=C(/C(SC)C)C)NC
Butonate	CCCC(=O)OC(C(Cl)(Cl)Cl)P(=O)(OC)OC
Butoxycarboxime	CC(C(=NOC(=O)NC)C)S(=O)(=O)C
Cadusafos	O=P(SC(CC)C)(OCC)SC(CC)C
Carbanolate	c1(c(c(c(C)cc1)C)Cl)OC(NC)=O
Carbaryl	c12c(OC(NC)=O)cccc1cccc2
Carbofuran	CC2(C)CC1=C(C(=CC=C1)OC(NC)=O)O2
Carbophenothion	COP(=S)(OC)SCSC1=CC=C(C=C1)Cl
Carbosulfan	c12c(CC(O2)(C)C)cccc1OC(N(SN(CCCC)CCCC)C)=O
Cartap	CN(C)C(CSC(=O)N)CSC(=O)N
Chlordane	Cl/C2=C(\Cl)C3(Cl)C1CC(Cl)C(Cl)C1C2(Cl)C3(Cl)Cl
Chlordecone	ClC54C(=O)C1(Cl)C2(Cl)C5(Cl)C3(Cl)C4(Cl)C1(Cl)C2(Cl)C3(Cl)Cl
Chlordimeforme	Clc1cc(c(/N=C/N(C)C)cc1)C
Chlorethoxyfos	CCOP(=S)(OCC)OC(C(Cl)(Cl)Cl)Cl
Chlorfénapyr	Clc2ccc(c1c(C#N)c(Br)c(n1COCC)C(F)(F)F)cc2

INSECTICIDES	NOTATION SMILES
Chlorfensulphide	c1(c(cc(Cl)c(c1)Cl)Cl)\N=N\Sc1ccc(Cl)cc1
Chlorfenviphos	CCOP(=O)(OCC)OC(=CCl)C1=C(C=C(C=C1)Cl)Cl
Chlorfluazuron	C1=CC(=C(C(=C1)F)C(=O)NC(=O)NC2=CC(=C(C(=C2)Cl)OC3=C(C=C(C=N3)C(F)(F)F)Cl)Cl)F
Chlorméphos	ClCSP(=S)(OCC)OCC
Chlorobenzilate	Clc1ccc(cc1)C(O)(c2ccc(Cl)cc2)C(=O)OCC
Chloromethiuron	c1cc(Cl)cc(C)c1NC(=S)N(C)C
Chlorpyriphos	Clc1c(OP(=S)(OCC)OCC)nc(Cl)c(Cl)c1
Chlorpyriphos-methyl	Clc1c(OP(=S)(OC)OC)nc(Cl)c(Cl)c1
Chlorthiophos	Clc1c(OP(=S)(OC)OC)ccc(SC)c1Cl
Cloethocarb	c1(c(cccc1)OC(NC)=O)O[C@@H](CCl)OC
Closantel	c1([C@@H](c2ccc(Cl)cc2)C#N)c(cc(NC(c2c(c(cc(c2)I)I)O)=O)c(c1)C)Cl
Crotoxyphos	O=P(OC)(OC)O\C(=C\C(=O)OC(c1ccccc1)C)C
Cyanophos	S=P(Oc1ccc(C#N)cc1)(OC)OC
Cyhalothrine	Cl\C(=C/C3C(C(=O)OC(C#N)c2cccc(Oc1ccccc1)c2)C3(C)C)C(F)(F)F
Cypermethrine	Cl/C(Cl)=C/C3C(C(=O)OC(C#N)c2cccc(Oc1ccccc1)c2)C3(C)C
Cyphenothrin	N#CC(OC(=O)C1C(/C=C(/C)C)C1(C)C)c3cccc(Oc2ccccc2)c3
Ddt	Clc1ccc(cc1)C(c2ccc(Cl)cc2)C(Cl)(Cl)Cl
Deltamethrine	Br/C(Br)=C/[C@H]3[C@@H](C(=O)O[C@H](C#N)c2cccc(Oc1ccccc1)c2)C3(C)C
Demeton-O-methylsulfone	S(CCOP(=S)(OC)OC)(=O)(=O)CC
Demeton-S-methyle	O=P(OC)(SCCSCC)OC
Demeton-S Methylsulfone	P(SCCS(CC)(=O)=O)(OC)(OC)=O
Diafenthiuron	S=C(NC(C)(C)C)Nc2c(cc(Oc1ccccc1)cc2C(C)C)C(C)C
Dialiphos	CCOP(=S)(OCC)SC(CCl)N1C(=O)C2=CC=CC=C2C1=O
Diazinon	S=P(OCC)(OCC)Oc1nc(nc(c1)C)C(C)C

INSECTICIDES	NOTATION SMILES
Diethyltoluamide	O=C(c1cc(ccc1)C)N(CC)CC
Diflovidazin	n1nc(nnc1c1c(cccc1F)F)c1c(cccc1)Cl
Dimefox	FP(=O)(N(C)C)N(C)C
Dimethoate	O=C(NC)CSP(=S)(OC)OC
Dimethylvinphos	Clc1cc(Cl)ccc1C(OP(=O)(OC)OC)=[C@H]Cl
Dioxabenzophos	c12c(CO[P@@](O1)(OC)=S)cccc2
Dioxacarbe	O=C(Oc1ccccc1C2OCCO2)NC
Dioxathion	S=P(SC1OCCOC1SP(=S)(OCC)OCC)(OCC)OCC
Disulfoton	S=P(OCC)(SCCSCC)OCC
Endosulfan	Cl\C1=C(/Cl)C3(Cl)C(Cl)(Cl)C1(Cl)C2C3COS(=O)OC2
Endothion	O=C1/C=C(\O/C=C1/OC)CSP(=O)(OC)OC
Endrine	Cl\C1=C(/Cl)C2(Cl)C3C4CC(C3C1(Cl)C2(Cl)Cl)C5OC45
Epn	S=P(Oc1ccc(cc1)[N+]([O-])=O)(OCC)c2ccccc2
Esfenvalerate	Clc1ccc(cc1)[C@H](C(=O)O[C@@H](C#N)c3cccc(Oc2ccccc2)c3)C(C)C
Ethiofencarbe	O=C(Oc1ccccc1CSCC)NC
Ethion	S=P(SCSP(=S)(OCC)OCC)(OCC)OCC
Ethiprole	FC(F)(F)c2cc(Cl)c(n1nc(C#N)c(c1N)S(=O)CC)c(Cl)c2
Ethoate-methyle	P(SCC(NCC)=O)(OC)(OC)=S
Fenazaflor	FC(F)(F)c2nc1cc(Cl)c(Cl)cc1n2C(=O)Oc3ccccc3
Fenchlorphos	Clc1cc(OP(=S)(OC)OC)c(Cl)cc1Cl
Fenitrothion	S=P(Oc1cc(c(cc1)[N+]([O-])=O)C)(OC)OC
Fenobucarb	O=C(Oc1ccccc1C(C)CC)NC
Gamma-cyhalothrine	C1([C@@H]([C@@H]1\C=C(\C(F)(F)F)Cl)C(=O)O[C@@H](c1cc(ccc1)Oc1ccccc1)C#N)(C)C
Halfenprox	BrC(F)(F)Oc1ccc(cc1)C(C)(C)COCc3cc(Oc2ccccc2)ccc3
Halofenozide	CC(C)(C)N(C(=O)C1=CC=CC=C1)NC(=O)C2=CC=C(C=C2)Cl

INSECTICIDES	NOTATION SMILES
Heptenophos	COP(=O)(OC)OC1=C(Cl)C2C=CCC12
Icaridin	O=C1/C=C(\ON1)C
Imiprothrin	O=C(OCN1C(=O)CN(C1=O)CC#C)C2C(\C=C(/C)C)C2(C)C
Indoxacarbe	FC(F)(F)Oc1ccc(cc1)N(C(=O)OC)C(=O)N3/N=C4/c2c(cc(Cl)cc2)C[C@@]4(OC3)C(=O)OC
Iodofenphos	Clc1cc(I)c(Cl)cc1OP(=S)(OC)OC
Isobenzan	Cl/C1=C(\Cl)C2(Cl)C(Cl)(Cl)C1(Cl)C3C2C(Cl)OC3Cl
Isoprocarbe	O=C(Oc1ccccc1C(C)C)NC
Isoxathion	S=P(Oc2noc(c1ccccc1)c2)(OCC)OCC
Kelevan	C12([C@@]3([C@@]4(C([C@@]5(Cl)C4(Cl)C(C3(Cl)[C@@]15Cl)(Cl)Cl)(Cl)C2(CC(CCC(OCC)=O)=O)O)Cl)Cl)Cl
Lambda-cyhalothrine	Cl\C(=C/[C@H]3[C@@H](C(=O)OC(C#N)c2cccc(Oc1ccccc1)c2)C3(C)C)C(F)(F)F
Leptophos	c1([P@@](Oc2c(cc(Br)c(c2)Cl)Cl)(OC)=S)ccccc1
Malathion	O=C(OCC)C(SP(=S)(OC)OC)CC(=O)OCC
Mécarbam	O=C(N(C(=O)OCC)C)CSP(=S)(OCC)OCC
Méphospholan	O=P(OCC)(/N=C1\SCC(S1)C)OCC
Methacrifos	S=P(OC)(OC)O\C=C(/C)C(=O)OC
Methamidophos	O=P(OC)(SC)N
Methidathion	O=C1SC(=N/N1CSP(=S)(OC)OC)\OC
Methiocarb	O=C(Oc1cc(c(SC)c(c1)C)C)NC
Methomyl	C/C(=N\OC(=O)NC)SC
Mexacarbate	O=C(Oc1cc(c(c(c1)C)N(C)C)C)NC
Mirex	ClC53C1(Cl)C4(Cl)C2(Cl)C1(Cl)C(Cl)(Cl)C5(Cl)C2(Cl)C3(Cl)C4(Cl)Cl
Morphothion	P(SCC(N1CCOCC1)=O)(OC)(OC)=S
Naled	ClC(Br)(Cl)C(Br)OP(=O)(OC)OC
Nitenpyram	C(\[N+]([O-])=O)=C(\N(Cc1cnc(cc1)Cl)CC)NC
Nithiazine	C1(\NCCCS1)=C/[N+](=O)[O-]

INSECTICIDES	NOTATION SMILES
Oxamyl	O=C(C(=N\OC(=O)NC)\SC)N(C)C
Parathion	S=P(Oc1ccc(cc1)[N+]([O-])=O)(OCC)OCC
Phenkapton	Clc1ccc(Cl)cc1SCSP(=S)(OCC)OCC
Phenthorate	O=C(OCC)C(SP(=S)(OC)OC)c1ccccc1
Phorate	S=P(OCC)(SCSCC)OCC
Phosalone	Clc2ccc1c(OC(=O)N1CSP(=S)(OCC)OCC)c2
Phosmet	COP(=S)(OC)SCN1C(=O)C2=CC=CC=C2C1=O
Phosnichlor	P(OC)(Oc1cc(c(cc1)Cl)[N+](=O)[O-])(OC)=S
Phosphamidon	O=P(O/C(=C(\Cl)C(=O)N(CC)CC)C)(OC)OC
Plifenate	Clc1ccc(cc1Cl)C(OC(=O)C)C(Cl)(Cl)Cl
Prallethrin	C1([C@H]([C@@H]1C(O[C@@H]1C(=C(CC#C)C(C1)=O)C)=O)\C=C(\C)C)(C)C
Profenophos	CCCSP(=O)(OCC)OC1=C(C=C(C=C1)Br)Cl
Prothiofos	Clc1cc(Cl)ccc1OP(=S)(OCC)SCCC
Pymetrozine	O=C2N(/N=C/c1cccnc1)C\C(=N/N2)C
Pyridafenthion	S=P(OCC)(OCC)O/C2=N/N(c1ccccc1)C(=O)/C=C2
Pyrimicarbe-	O=C(Oc1nc(nc(c1C)C)N(C)C)N(C)C

L'élaboration de la base de données est décrite par le diagramme de la figure IV.2

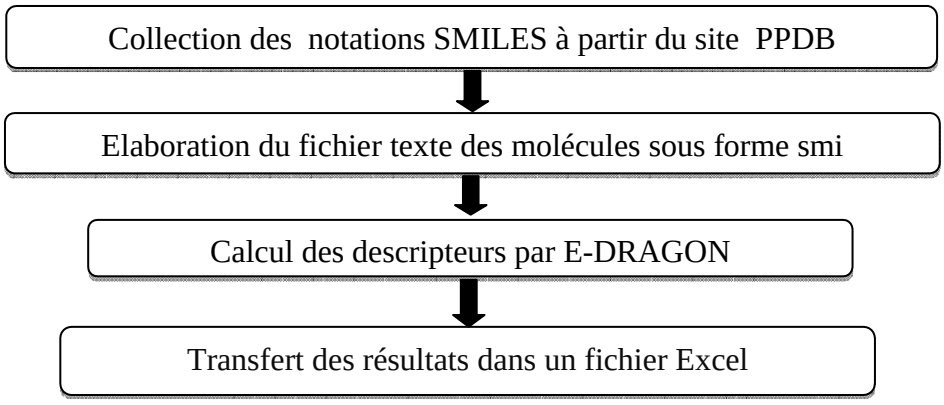
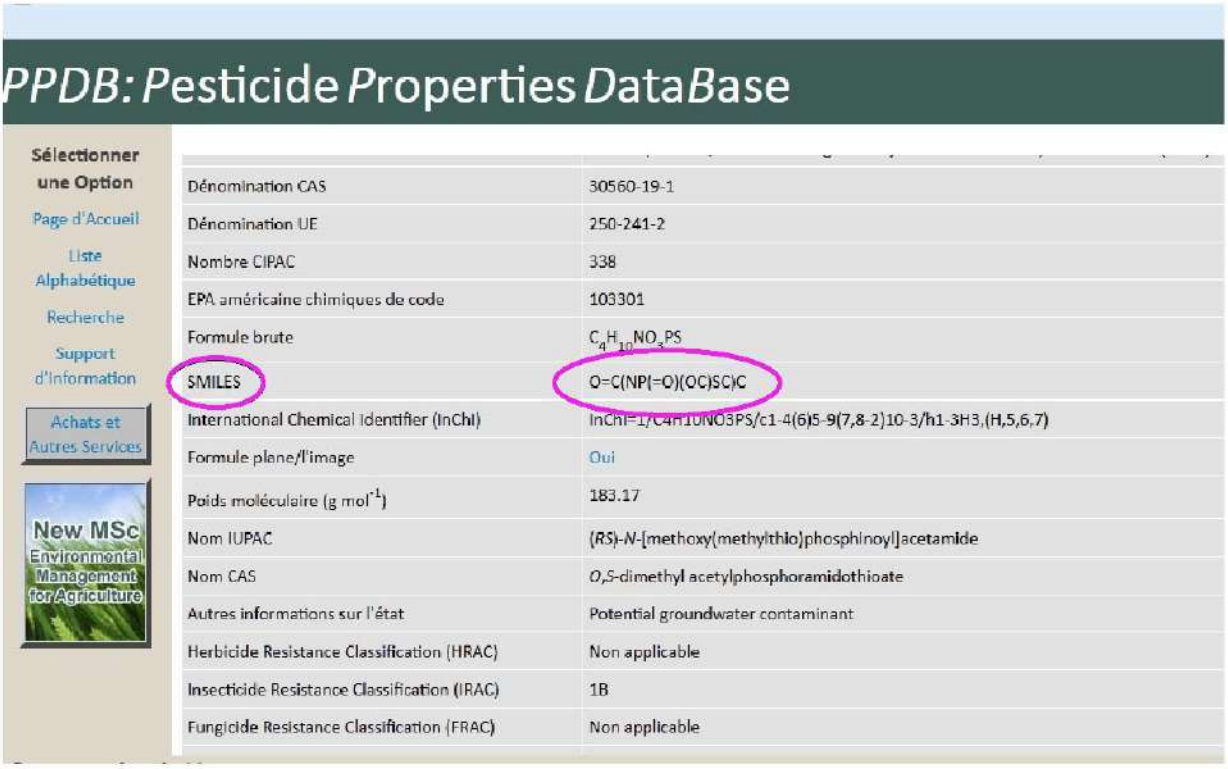


Figure IV.2 : Procédure d'élaboration de la base de données.



PPDB: Pesticide Properties DataBase

Sélectionner une Option

Page d'Accueil

Liste Alphabétique

Recherche

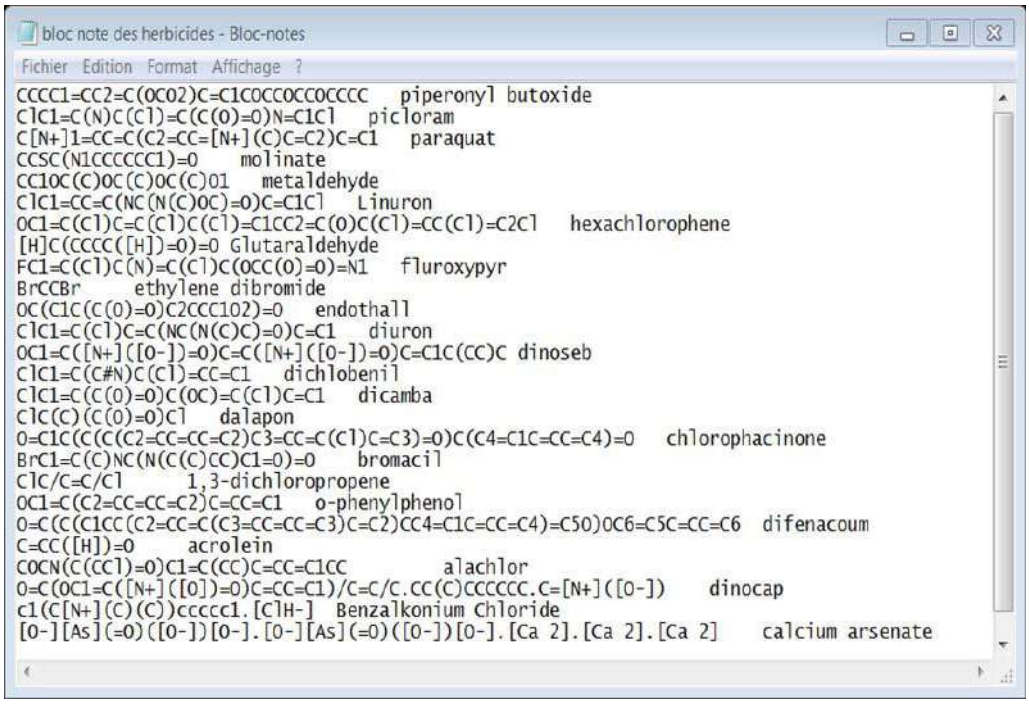
Support d'Information

Achats et Autres Services

New MSc Environmental Management for Agriculture

Dénomination CAS	30560-19-1
Dénomination UE	250-241-2
Nombre CIPAC	338
EPA américaine chimiques de code	103301
Formule brute	$C_4H_{10}NO_3PS$
SMILES	O=C(NP(=O)(OC)SC)C
International Chemical Identifier (InChI)	InChI=1/C4H10NO3PS/c1-4(6)5-9(7,8-2)10-3/h1-3H3,(H,5,6,7)
Formule plane/l'image	Oui
Poids moléculaire (g mol^{-1})	183.17
Nom IUPAC	(*RS*)-*N*-[methoxy(methylthio)phosphinoyl]acetamide
Nom CAS	*O,S*-dimethyl acetylphosphoramidothioate
Autres informations sur l'état	Potential groundwater contaminant
Herbicide Resistance Classification (HRAC)	Non applicable
Insecticide Resistance Classification (IRAC)	1B
Fungicide Resistance Classification (FRAC)	Non applicable

Figure IV.3 : Collection des notations SMILES sur le site PPDB.



bloc note des herbicides - Bloc-notes

Fichier Edition Format Affichage ?

```
CCCC1=CC2=C(OCO2)C=C1COCCOCCOCCCC   piperonyl butoxide
ClC1=C(N)C(Cl)=C(C(O)=O)N=C1Cl   picloram
C[N+]1=CC=C(C2=CC=[N+](C)C=C2)C=C1   paraquat
CCSC(N1CCCCCC1)=O    molinate
CC1OC(C)OC(C)OC(C)O1   metaldehyde
ClC1=CC=C(NC(N(C)OC)=O)C=C1Cl   Linuron
OC1=C(Cl)C=C(Cl)C(Cl)=C1CC2=C(O)C(Cl)=CC(Cl)=C2Cl   hexachlorophene
[H]C(CCCC([H])=O)=O Glutaraldehyde
FC1=C(Cl)C(N)=C(Cl)C(OCC(O)=O)=N1   fluroxypyr
BrCCBr    ethylene dibromide
OC(C1C(C(O)=O)C2CCC1O2)=O   endothall
ClC1=C(Cl)C=C(NC(N(C)C)=O)C=C1   diuron
OC1=C([N+]([O-])=O)C=C([N+]([O-])=O)C=C1C(CC)C dinoseb
ClC1=C(C#N)C(Cl)=CC=C1   dichlobenil
ClC1=C(C(O)=O)C(OC)=C(Cl)C=C1   dicamba
ClC(C)(C(O)=O)Cl   dalapon
O=C1C(C(C(C2=CC=CC=C2)C3=CC=C(Cl)C=C3)=O)C(C4=C1C=CC=C4)=O   chlorophacinone
BrC1=C(C)NC(N(C(C)CC)C1=O)=O    bromacil
ClC/C=C/Cl      1,3-dichloropropene
OC1=C(C2=CC=CC=C2)C=CC=C1   o-phenylphenol
O=C(C(C1CC(C2=CC=C(C3=CC=CC=C3)C=C2)CC4=C1C=CC=C4)=C50)OC6=C5C=CC=C6  difenacoum
C=CC([H])=O     acrolein
COCN(C(CCl)=O)C1=C(CC)C=CC=C1CC         alachlor
O=C(OC1=C([N+]([O])=O)C=CC=C1)/C=C/C.CC(C)CCCCCC.C=[N+]([O-])    dinocap
c1(C[N+](C)(C))ccccc1.[ClH-]  Benzalkonium Chloride
[O-][As](=O)([O-])[O-].[O-][As](=O)([O-])[O-].[Ca 2].[Ca 2].[Ca 2]    calcium arsenate
```

Figure IV.4 : Elaboration du fichier texte des molécules sous forme Smiles.



Figure IV.5 : Première étape pour le calcul des descripteurs par E Dragon1.0. (1)



Figure IV.5 : Deuxième étapes pour le calcul des descripteurs par E Dragon1.0. (2)

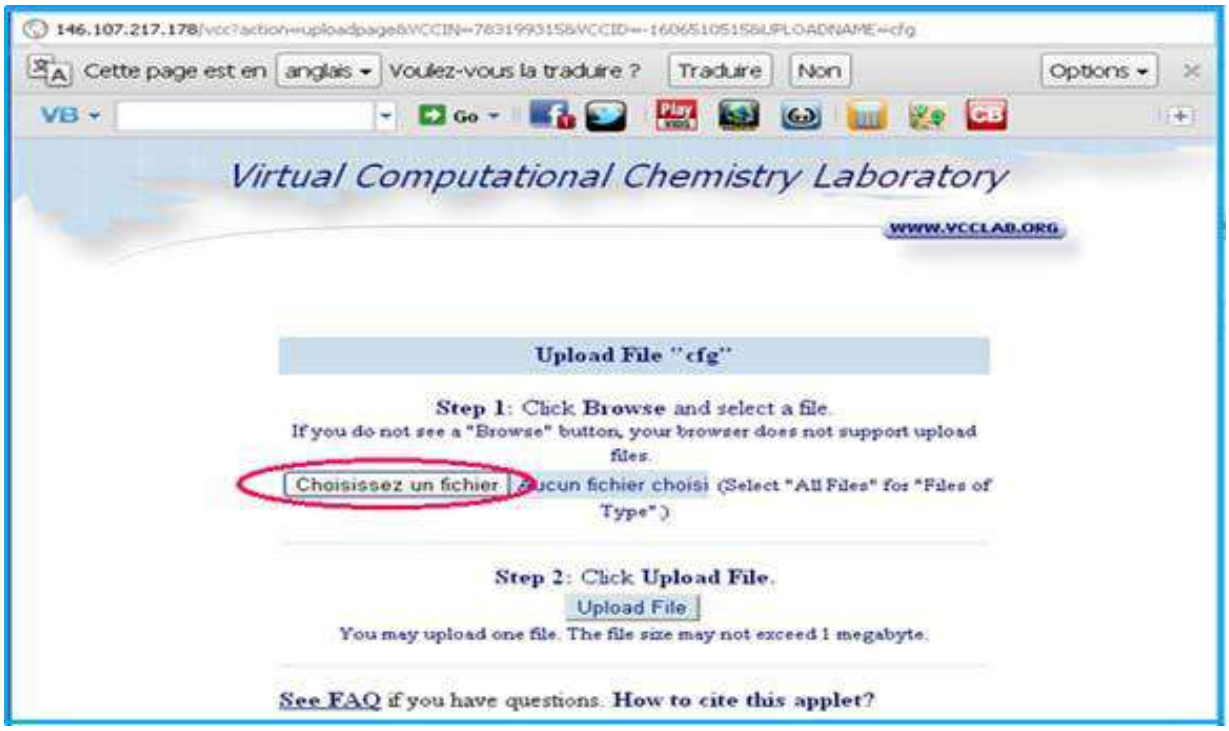


Figure IV.5 : Troisième étape pour le calcul des descripteurs par E Dragon1.0. (3)

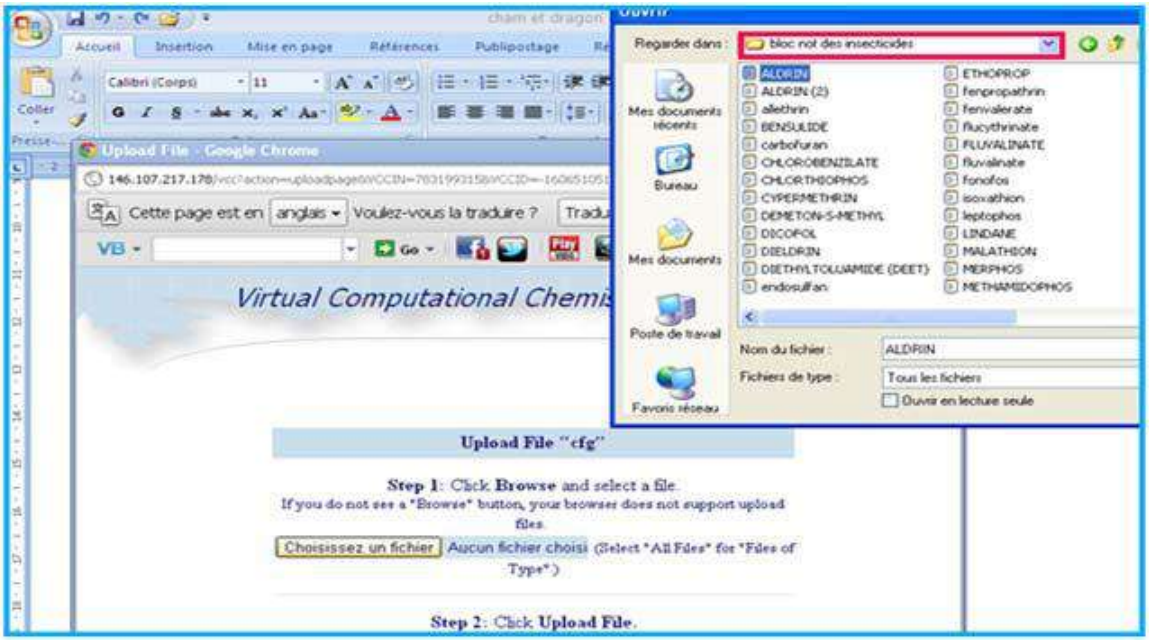


Figure IV.5 : Quatrième étape pour le calcul des descripteurs par E Dragon1.0. (4)

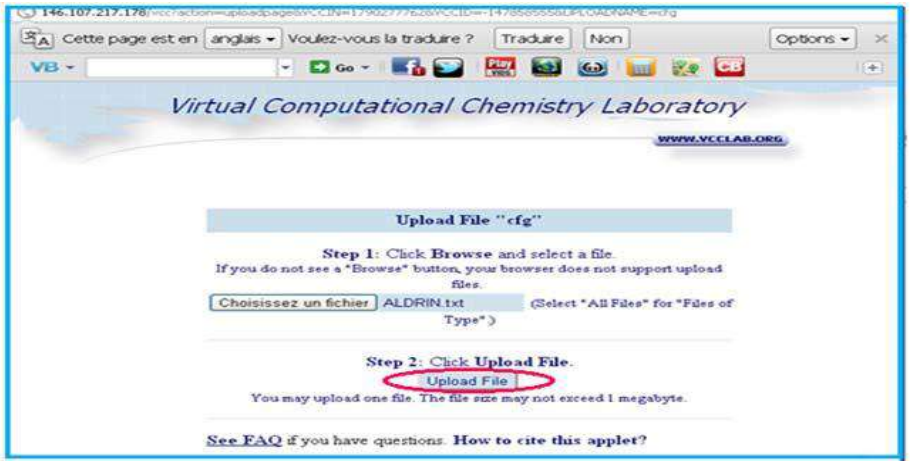


Figure IV.5 : Cinquième étape pour le calcul des descripteurs par E Dragon1.0. (5)

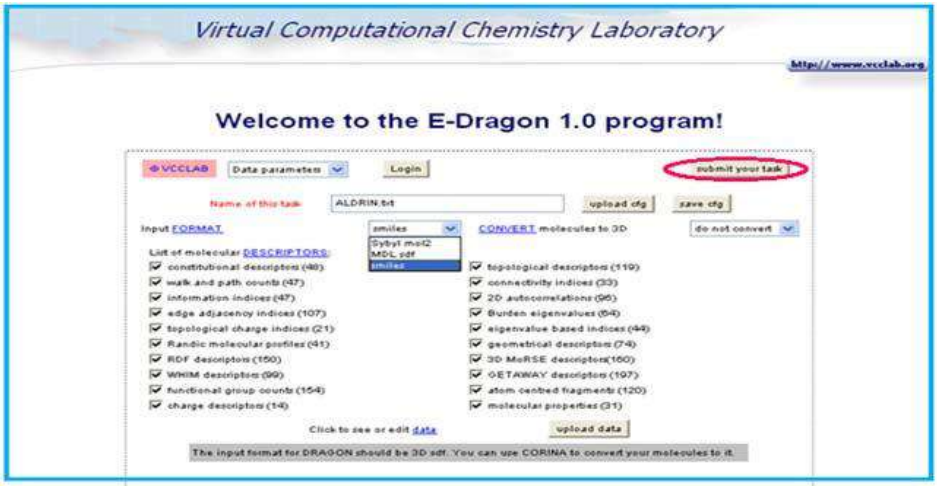


Figure IV.5 : Sixième étape pour le calcul des descripteurs par E Dragon1.0. (6)

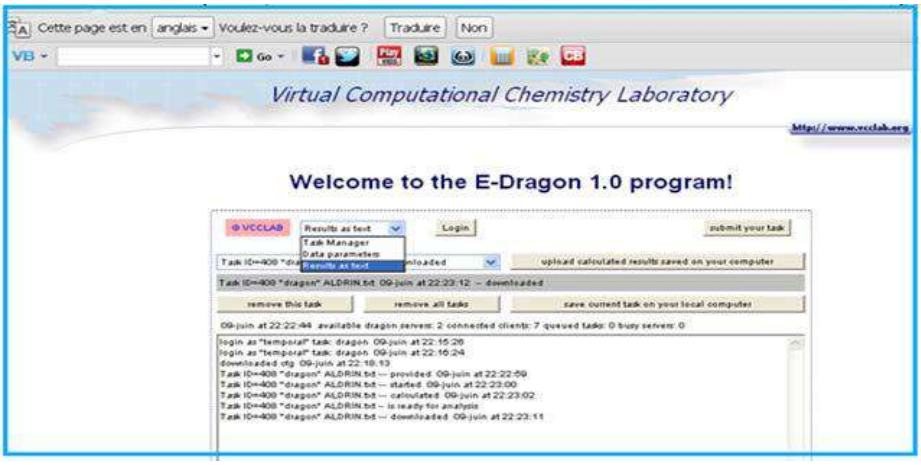


Figure IV. 5 : Septième étape pour le calcul des descripteurs par E Dragon1.0. (7)



Figure IV.5 : Huitième étape pour le calcul des descripteurs par E Dragon1.0. (8)

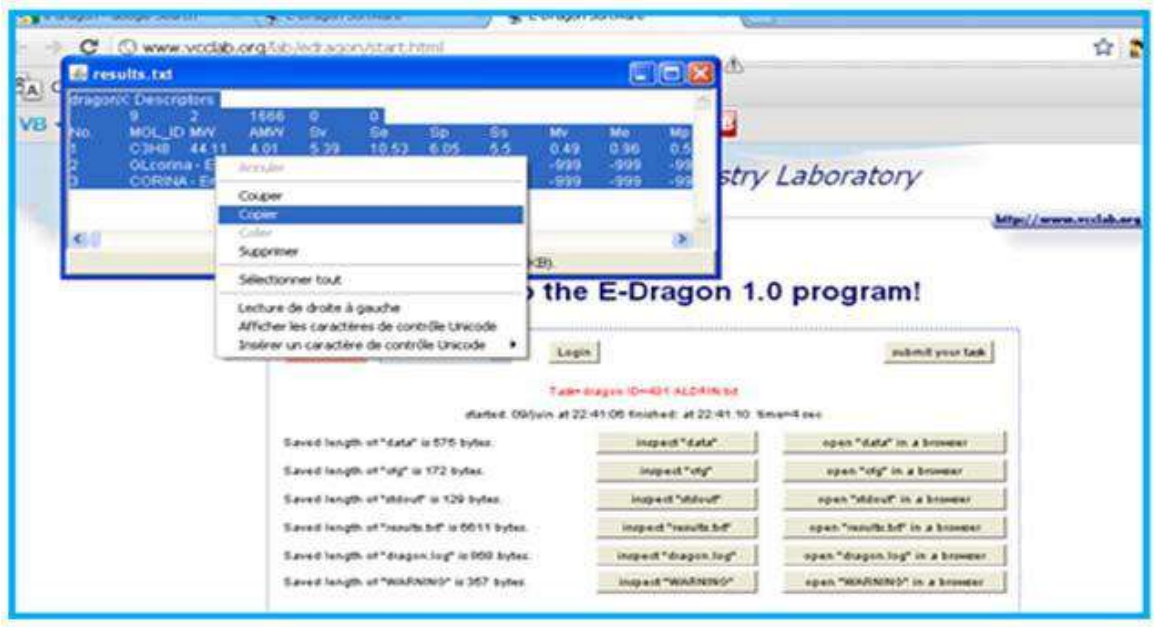
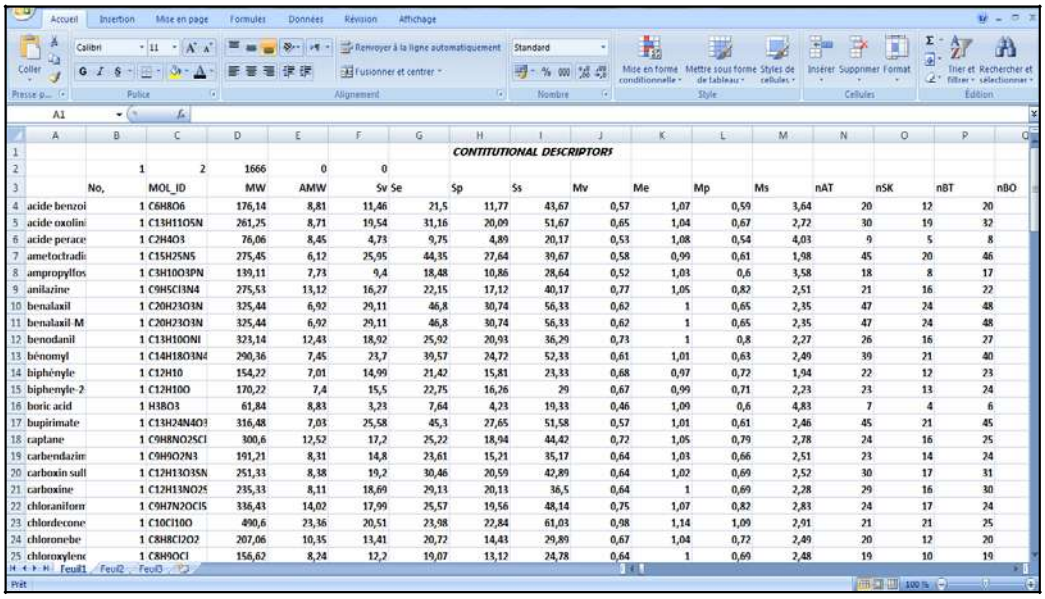


Figure IV.5 : Neuvième étape pour le calcul des descripteurs par E Dragon1.0 (9)



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1								CONTITUTIONAL DESCRIPTORS									
2			1 2	1666	0	0											
3		No,	MOL_ID	MW	AMW	Sv Se		Sp	Ss	Mv	Me	Mp	Ms	nAT	nSK	nBT	nBO
4	acide benzoi	1	C6H8O6	176,14	8,81	11,46	21,5	11,77	43,67	0,57	1,07	0,59	3,64	20	12	20	
5	acide oxolin	1	C13H11O5N	261,25	8,71	19,54	31,16	20,09	51,67	0,65	1,04	0,67	2,72	30	19	32	
6	acide perace	1	C2H4O3	76,06	8,45	4,73	9,75	4,89	20,17	0,53	1,08	0,54	4,03	9	5	8	
7	ametoctradi	1	C15H25N5	275,45	6,12	25,95	44,35	27,64	39,67	0,58	0,99	0,61	1,98	45	20	46	
8	ampropylfos	1	C3H10O3PN	139,11	7,73	9,4	18,48	10,86	28,64	0,52	1,03	0,6	3,58	18	8	17	
9	anilazine	1	C9H5Cl3N4	275,53	13,12	16,27	22,15	17,12	40,17	0,77	1,05	0,82	2,51	21	16	22	
10	benalaxil	1	C20H23O3N	325,44	6,92	29,11	46,8	30,74	56,33	0,62	1	0,65	2,35	47	24	48	
11	benalaxil-M	1	C20H23O3N	325,44	6,92	29,11	46,8	30,74	56,33	0,62	1	0,65	2,35	47	24	48	
12	benodanil	1	C13H10ONI	323,14	12,43	18,92	25,92	20,93	36,29	0,73	1	0,8	2,27	26	16	27	
13	bénomyl	1	C14H18O3N4	290,36	7,45	23,7	39,57	24,72	52,33	0,61	1,01	0,63	2,49	39	21	40	
14	biphényle	1	C12H10	154,22	7,01	14,99	21,42	15,81	23,33	0,68	0,97	0,72	1,94	22	12	23	
15	biphenyle-2	1	C12H10O	170,22	7,4	15,5	22,75	16,26	29	0,67	0,99	0,71	2,23	23	13	24	
16	boric acid	1	H3BO3	61,84	8,83	3,23	7,64	4,23	19,33	0,46	1,09	0,6	4,83	7	4	6	
17	bupirimate	1	C13H24N4O3	316,48	7,03	25,58	45,3	27,65	51,58	0,57	1,01	0,61	2,46	45	21	45	
18	captane	1	C9H8NO2SCl	300,6	12,52	17,2	25,22	18,94	44,42	0,72	1,05	0,79	2,78	24	16	25	
19	carbendazim	1	C9H9O2N3	191,21	8,31	14,8	23,61	15,21	35,17	0,64	1,03	0,66	2,51	23	14	24	
20	carboxin sull	1	C12H13O3SN	251,31	8,38	19,2	30,46	20,59	42,89	0,64	1,02	0,69	2,52	30	17	31	
21	carboxine	1	C12H13NO2S	235,33	8,11	18,69	29,13	20,13	36,5	0,64	1	0,69	2,28	29	16	30	
22	chloranilorn	1	C9H7N2OCl5	336,43	14,02	17,99	25,57	19,56	48,14	0,75	1,07	0,82	2,83	24	17	24	
23	chlordecone	1	C10Cl10O	490,6	23,36	20,51	23,98	22,84	61,03	0,98	1,14	1,09	2,91	21	21	25	
24	chloronebe	1	C8H8Cl2O2	207,06	10,35	13,41	20,72	14,43	29,89	0,67	1,04	0,72	2,49	20	12	20	
25	chloroxylenc	1	C8H9OCl	156,62	8,24	12,2	19,07	13,12	24,78	0,64	1	0,69	2,48	19	10	19	

Figure IV.5 : Dixième et dernière étape coller les descripteurs calculés par E Dragon1.0. (10)

La base de données de la figure VI.5 (10) contient 141 lignes qui représentent les molécules, et 1666 colonnes qui représentent leurs propriétés (descripteurs, voir Figure IV.6) et une colonne réservée à la toxicité de chaque molécule.

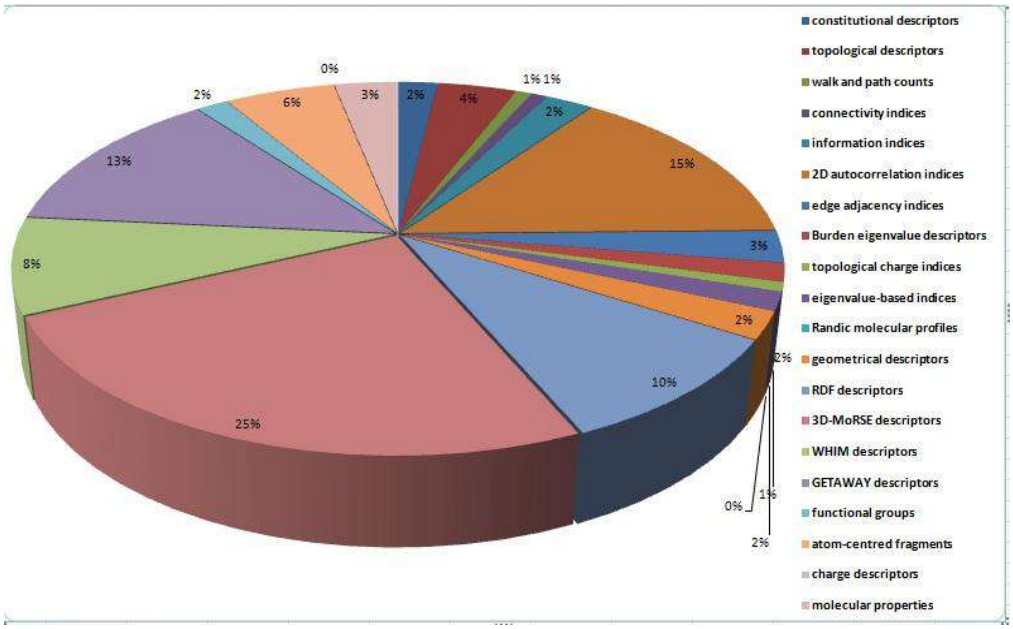


Figure IV.6 : Pourcentage des descripteurs avant prétraitement

IV.2.2 Prétraitement de la base de données

Le prétraitement de la base de données a pour objectif d'éliminer les descripteurs qui ne sont pas pertinents. Elle a lieu en plusieurs étapes, décrites ci-dessous :

Étape 1 : On calcule le minimum et le maximum pour chaque descripteur a l'aide d'un programme de MATLAB on enlève les descripteurs qui ont le maximum et le minimum égaux;

Étape 2 : On élimine le descripteur qui possède la même valeur pour plus de 75% des échantillons

Étape 3 : On calcule la déviation standard relative (RSD) pour chaque descripteur et on élimine ceux ayant des valeurs de RSD inférieures à 0.05.

$$RSD(i) = \frac{\text{std}(\text{data}(:,i))}{\text{mean}(\text{data}(:,i))}$$

Étape 4 : On obtient alors une matrice diagonale qui représente la corrélation entre la toxicité et les descripteurs retenus. On classe les descripteurs selon la valeur décroissante du coefficient de corrélation. On prend le descripteur ayant la plus grande corrélation et on le compare aux autres descripteurs de la matrice. Ceux dont la valeur du coefficient de corrélation est supérieure à 0.75 sont éliminés à leur tour. On répète la même chose avec le descripteur classé juste après le premier et ainsi de suite. Les résultats obtenus sont représentés sur la Figure IV.7.

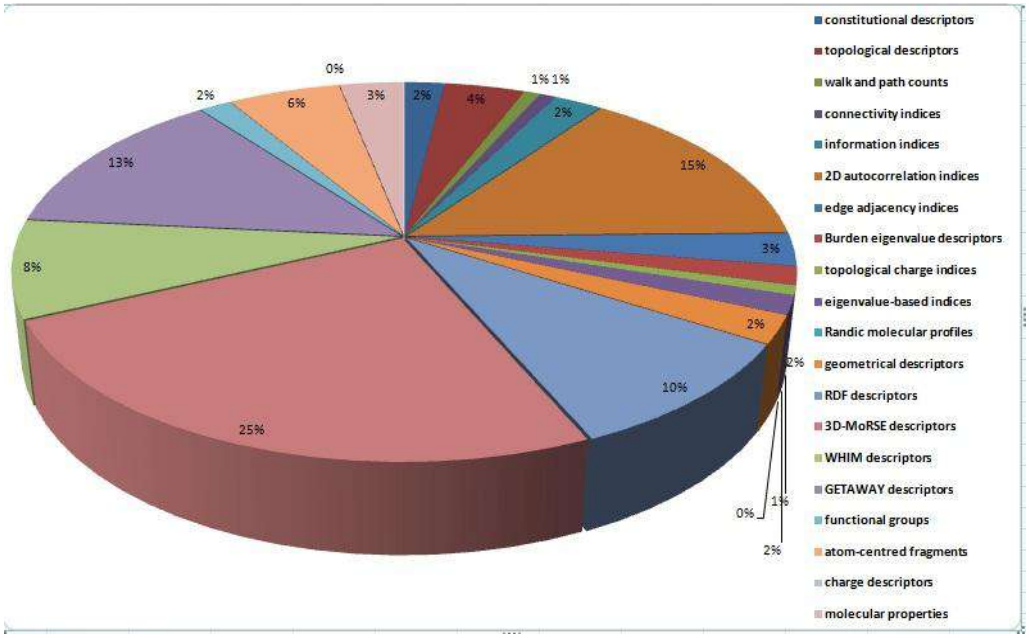


Figure IV.7 : Pourcentage des descripteurs après prétraitement.

Etape 5 : On élabore un programme basé sur la méthode stepwise pour la sélection des descripteurs les plus pertinents parmi ceux obtenus à l'étape 4. Les descripteurs sélectionnés et retenus (au nombre de 8) à l'issue de cette étape sont :

- **Mor11u** : 3D-MoRSE - signal 11 / unweighted
- **Mor26e** : 3D-MoRSE - signal 26 / weighted by atomic Sanderson electronegativities
- **Mor19p** : 3D-MoRSE - signal 19 / weighted by atomic polarizabilities
- **Mor26p** : 3D-MoRSE - signal 26 / weighted by atomic polarizabilities
- **H6m** : H autocorrelation of lag 6 / weighted by atomic masses
- **H6v** : H autocorrelation of lag 6 / weighted by atomic van der Waals volumes
- **R6m+** : R maximal autocorrelation of lag 6 / weighted by atomic masses
- **nArX** : number of X on aromatic ring

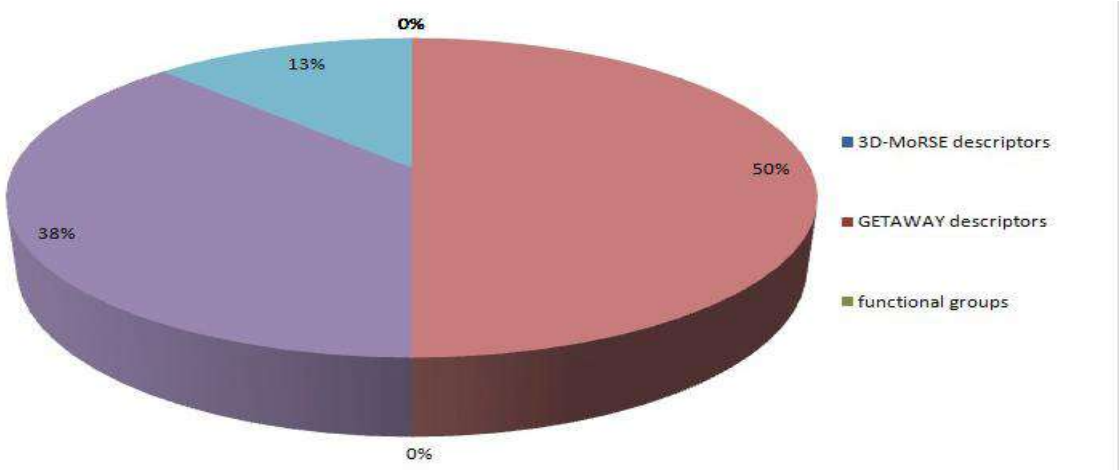


Figure Iv.8. Pourcentage des descripteurs retenus par la méthode stepwise

IV.2.3 Normalisation de la base de données partielle

La normalisation est une étape importante dans le processus de l'élaboration des données. Elle devient dans la plupart des cas, une exigence pour les matrices des inputs, du fait qu'on utilise souvent des fonctions de transfert bornées de type sigmoïdes dans les modélisations statique.

Dans notre travail, on a utilisé la base de données qui est représentée en 8 lignes, correspondant respectivement aux descripteurs ((3D-MoRSE, GETAWAY, groupe fonctionnels) et par la toxicité. Les valeurs numériques ont été normalisées afin d'améliorer l'optimisation. La normalisation se fait par ligne (***mapminmax MATLAB version : 7.5***). Chaque paramètre des entrées et de sortie (x_i) est normalisé (x_{in}) par l'expression suivante:

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

IV.2.4 Mise en forme des données

Dans cette étape les données sont subdivisées en deux sous-ensembles :

- Un ensemble pour l'apprentissage (80%) ;
- Un ensemble pour le test (20%).

IV.2.5 Conception et optimisation

La conception et l'optimisation du réseau a était réalisée en élaborant un programme en MATLAB. Le nombre de neurones dans la couche cachée à été varié de 1 à 20. On a choisi comme fonction de transfert la fonction tangente sigmoïde « tansig » pour la couche cachée et la fonction identité « purelin » pour la couche de sortie et comme algorithme d'apprentissage l'algorithme de Levenberg-Marquardt « trainlm ».

IV.3. Structure de réseau de neurone optimale obtenue (RNSO)

La structure de base retenue est un perceptron multicouche à trois couches : une couche d'entrée, une couche cachée, une couche de sortie. Le tableau IV.3 récapitule les différents éléments structurels du réseau de neurones optimal.

IV.3.1 L'algorithme d'apprentissage et la fonction d'activation

L'algorithme d'apprentissage utilisé est Levenberg-Marquardt (la fonction trainlm de MATLAB®) avec l'option d'apprentissage atténué pour éviter le sur-apprentissage. La procédure commence par deux fonctions de transfert sigmoïde tangent hyperbolique (fonction tansig de MATLAB®) pour la couche cachée et la fonction linéaire (fonction purelin de MATLAB®) pour la couche de sortie.

Tableau IV.2 : Structure des RNSO utilisés dans la modélisation de la toxicité.

Couche d'entrée	**Couche cachée**		**Couche de sortie**	
Nombre de neurones	Nombre de neurones	Fonction d'activation	Nombre de neurones	Fonction d'activation
8	10	Tangente sigmoïde « tansig »	1	Identité « purelin »

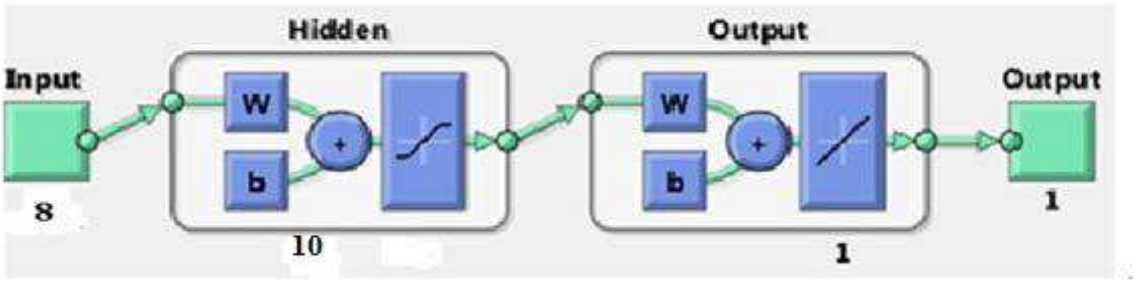


Figure IV.9 : Schéma simplifié de l'architecture du RNSO.

IV.4 Validation des valeurs calculées par RNA avec les valeurs expérimentales

Sur le tableau IV.4 sont donnés les différentes valeurs des erreurs (REQM et ERAM) et les coefficients de régression **(R).**

$$ERAM = (\sum | y_{exp} - y_{cal} | / y_{exp}) / (nbre\ de\ données) \times 100 \quad (IV.1)$$

$$REQM = \sqrt{1/N(y_{exp} - y_{cal})} \quad (IV.2)$$

$$R = \frac{\sum_{k=1}^{N}(y_{exp} - \bar{y}_{exp})(y_{cal} - \bar{y}_{cal})}{\sqrt{\sum_{k=1}^{N}(y_{exp} - \bar{y}_{exp})^2 (y_{cal} - \bar{y}_{cal})^2}} \quad (IV.3)$$

où **N** est le nombre d'observations (nombre d'insecticides); $\mathbf{y_{exp}}$ et $\mathbf{y_{cal}}$ sont respectivement les valeurs expérimentales (ou observées) et les valeurs calculées par notre modèle. $\bar{y}_{exp}$ et $\bar{y}_{cal}$ sont respectivement les valeurs moyennes pour les valeurs expérimentales (ou observées) et les valeurs calculées.

Le coefficient de régression, désigné par **R** (ou parfois r) est une mesure de l'intensité de la relation linéaire ou linéaire entre les deux variables. Ce coefficient prend des valeurs comprises entre +1 et -1. Les points suivants sont les lignes directrices reconnues pour interpréter le coefficient de corrélation: R = 0 indique l'absence de relation linéaire, R = 1 ou R = -1 indique une relation linéaire parfaite respectivement positive ou négative. Les valeurs de R entre 0,7 et 1,0 indiquent une forte relation linéaire positive.

L'exactitude des prévisions est, quant à elle, estimée par utilisation de l'indicateur ERAM (ou MAPE : Mean Absolute Percentage Error).

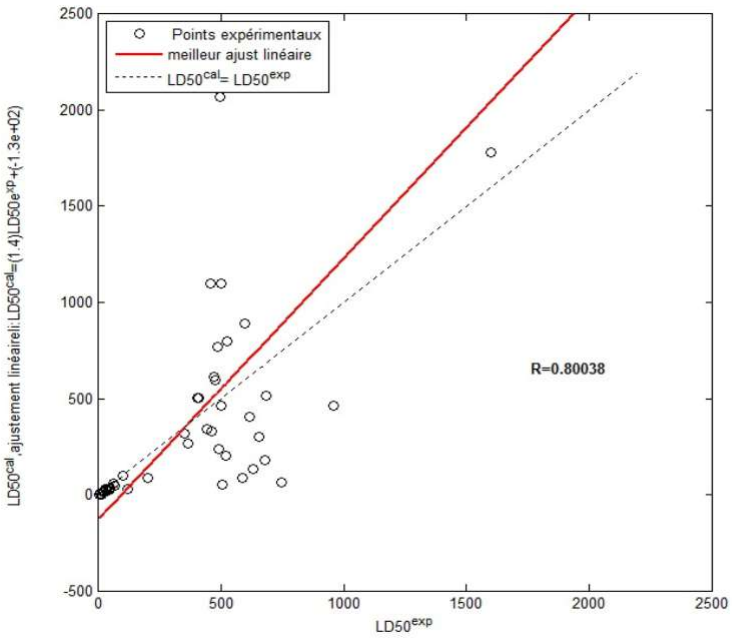
Tableau IV.3 : Récapitulatif des erreurs (REQM, ERAM) et des coefficients de régression (R) obtenues.

	Test	**Apprentissage**	**Total**
REQM (mg/kg)	825.23	1.63 e 03	1.417e 03
ERAM (%)	1.33	2.61	3.95
R	0.80	0.85	0.84

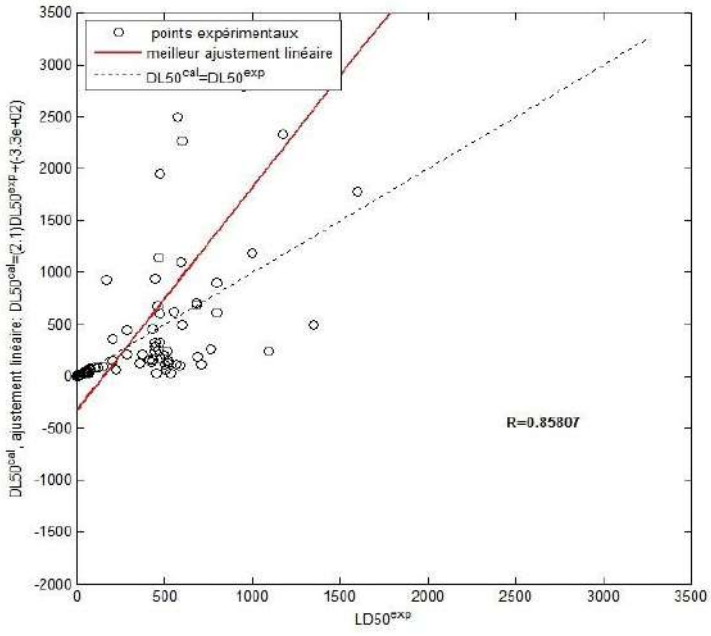
D'après les valeurs de R obtenues, lesquelles se situent entre 0.7 et 1.0, on peut estimer que le modèle obtenu arrive à faire une assez bonne prédiction de la toxicité des insecticides. Cependant,

Il serait souhaitable que ce coefficient se rapproche de la valeur 1.0 afin d'avoir une bonne prédiction. Pour arriver à cela, d'autres paramètres entrant dans la recherche du modèle optimal (comme l'algorithme d'apprentissage, la nature du réseau, … etc.) doivent être envisagés.

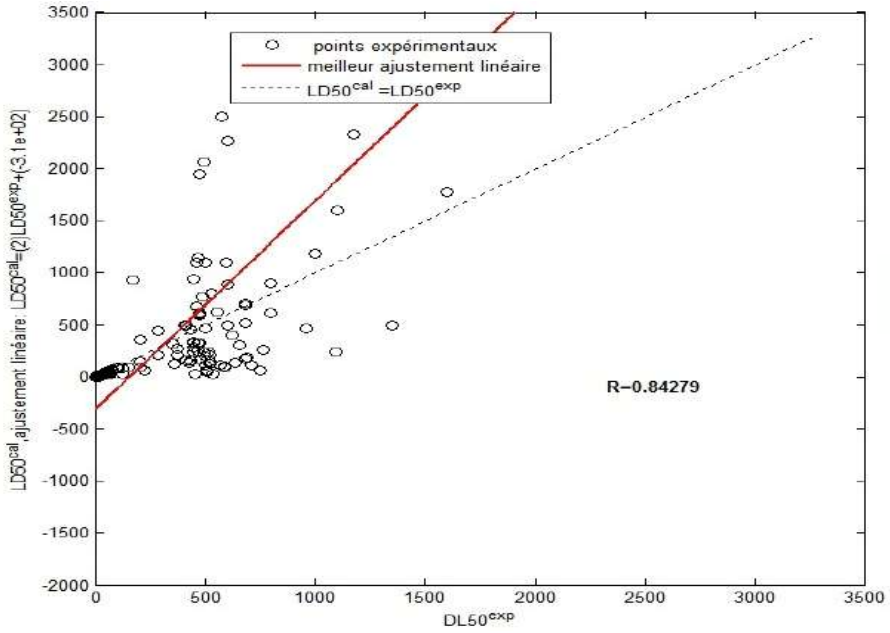
La validation des valeurs calculées par les réseaux de neurone avec les valeurs expérimentales est schématisée sur la figure IV.7 (a, b et c).



(a)



(b)



(c)

Figure IV.10 : Validation des valeurs calculées par RNSO avec les valeurs expérimentale : (a) test, (b) apprentissage et (c) total.

IV.5 Comparaison avec d'autres travaux

Pour mieux situer les résultats de ce travail, nous avons essayé de les comparer avec ceux obtenus par la même méthode (Stepwise) et ce pour les fongicides (Boura-Kessarine. 2013) et les herbicides (Touati-Debieb. 2013).

Tableau IV.4 : valeurs du coefficient de régression R pour les fongicides, herbicides et Insecticides

Composés étudiés	Valeurs du coefficient de régression R		
	Test	**Apprentissage**	**Total**
fongicides	0.97	0.97	0.96
herbicides	0.83	0.97	0.93
insecticides	0.80	0.85	0.84

D'après le tableau IV.5 ci-dessus, nous remarquons que les résultats obtenus dans le cas des fongicides sont de loin les plus importants et les plus significatifs. Ceci peut éventuellement s'expliquer par la similitude des structures des fongicides étudiés, contrairement aux herbicides mais surtout par rapport aux insecticides étudiés dans ce travail.

IV.6 Conclusion

Dans ce chapitre nous avons abordé le problème de prédiction de la toxicité des insecticides, en utilisant des programmes élaborés sur MATLAB. Nous avons montré dans ce travail qu'il est possible de relier la toxicité à la topologie de la molécule après le calcul des descripteurs (3D-MoRSE, GETAWAY, groupe fonctionnels) par le logiciel E-DRAGON.

CONCLUSION GENERALE

Les résidus de pesticides constituent aujourd'hui un veritable problème de société. Depuis près de cinquante ans, ils ont été mis en évidence dans tous les compartiments de l'environnement, dans les eaux des rivières et des nappes phréatiques, dans l'air et dans les eaux de pluie. On les rencontre aussi dans les fruits, les légumes, les céréales et les produits d'origine animale (viande, poisson, lait et œuf). Ils existent sous leur forme initiale (substances actives) mais ils peuvent aussi être existés sous une forme partiellement dégradée. Ces pesticides regroupent un grand nombre de spécialités de toxicité variable pour l'homme et l'environnement.

La modélisation pour la prédiction et où la corrélation de la toxicité des insecticides continue d'être un sujet d'actualité scientifique aussi bien du point de vue académique qu'économique et pratique.

Cette étude traite la modélisation de la toxicité des insecticides par la méthode des réseaux de neurones artificiels qui fait partie des nouvelles techniques sous l'égide de l'intelligence computationnelle ou des sciences cognitives. A cet égard, nous avons présenté les étapes nécessaires au calcul des descripteurs QSAR et leur utilisation pour la prédiction de la toxicité qui contribue à la réduction de l'utilisation des animaux et en fournissant des données pour faire des classifications positive de la toxicité.

Le développement d'un modèle QSAR débute par la collecte de données expérimentales fiables. Il s'agit ensuite de développer une série de descripteurs qui caractérisent les structures moléculaires des composés de la base de données en vue de les relier à la propriété expérimentale étudiée (toxicité). Le calcul des descripteurs est assuré par le logiciel (E Dragon) et est effectué après avoir établi les notations SMILES pour chaque molécule. Un grand nombre de descripteurs a été éliminé par la méthode de prétraitement.

Le choix des descripteurs les plus pertinents a été fait par la méthode dite Stepwise. Les classes de descripteurs les plus pertinents obtenus alors dans cette étude sont (3D-MoRSE, GETAWAY et les groupes fonctionnels).

A la lumière des résultats obtenus, nous pouvons tirer les conclusions suivantes :

- La modélisation par réseau de neurones est une méthode qui est basée seulement sur des données expérimentales, aussi la qualité et la quantité des données influent sur l'exactitude et la robustesse du modèle optimal.
- La modélisation neuronale peut servir comme une alternative fiable aux méthodes traditionnelles de modélisation de la prédiction de la toxicité des insecticides.
- De nombreux facteurs vont intervenir sur la capacité prédictive. Nous citons les plus importants:
    - La nature des molécules sélectionnées dans la base de données et la présence de ressemblances entre les structures des molécules testées (celles ci peuvent être considérées comme des facteurs très importants pour l'amélioration des capacités prédictives du modèle).
    - La taille de la base de données.
    - Le nombre et le type de descripteurs moléculaires.
    - La méthode statistique de sélection des descripteurs les plus pertinents.
    - Et finalement les propriétés du réseau de neurones optimisé.

En perspective, tout en estimant avoir atteint notre objectif tracé pour cette étude, nous proposons à ce que la modélisation neuronale puisse être aussi appliquée en association avec de nouvelles techniques de représentations des structures moléculaires (les graphs machines) qui ne demandent pas le calcul préalable des descripteurs moléculaires et qui peuvent éventuellement améliorer la précision de la prédiction.

CHAPITRE 1

A.Garrido-Frenche; J.J. Martinez Vidal; T. Lopez, S. Cortés Agluado; I.J. Martinez-Salvador ; A. Chrom "Monitoring multi-class pesticide residues in fresh fruits and vegetables by liquid chromatography with tandem mass spectrometry".(2004). (vu le 09 Mars 2013).

Association De Coordination Technique Agricole. "Index phytosanitaire ACTA ".(2006). (vu le 05 Mars 2013).

Children in the New Millennium: Environmental Impact on Health; Possible Health Effects of Pesticide Exposure; UNEP, UNICEF and WHO. (2002). (vu le 30 Mars 2013).

D.A. Lindquist : Les pesticides:la chimie au service de la survie. (vu le 24 Mars 2013).

Directive européenne du conseil du 15 juillet 1991 concernant la mise sur le marché des produits phytosanitaires. (Journal officiel). (91/414/CEE); (JO L 230 du 19. 8. 1991, p. 1). (vu le 11 Mars 2013).

Emballages des préparations pesticides. Arrêté du 28 Mars 1989 relatif aux conditions de classement.

Employment bysector, ILO (2007). (vu le 26 Mars 2013).

Evaluer les risques environnementaux des pesticides. Exemple du désherbage des cultures résistantes ou non au glyphosate pdf. (vu le 07 Mars 2013).

Gunnell and Eddelston :Suicide by intentional ingestion of pesticides: a continuing tragedy in developing countries, International Journal of Epidemiology .32:902-909. (2003). (vu le 09 Mars 2013).

J. Jeyaratnam; acute Pesticide Poisoning: A Major Global Health Problem, World Health
J. M. Bertolote et al; Deaths from pesticide poisoning: a global response, The British Journal of Psychiatry 189: 201-203; http://bjp.rcpsych.org/cgi/content/full/189/3/201. (2006). (vu le 09 Mars 2013).

Jack Weinberg : un guide pour les ONG sur les pesticides dangereux et la SAICM .Par Jack Weinberg, Conseiller principal en politiques International, POPs Elimination Network Conseiller principal en politiques International POPs Elimination Network Un cadre d'action pour la protection de la santé humaine et de l'environnement contre les pesticides dangereux .pdf . (vu le 18 Mars 2013).

Les pesticides en milieu agricole : état de la situation environnementale et initiatives prometteuses . (vu le 20 Mars 2013).

M.A. Watts ; Pesticides : Sowing Poison, Growing Hunger, Reaping Sorrow; 2nd Edition (2009). (vu le 01 Avril 2013).

M.A.Watts : World Development Report: Agriculture for Development. (2009). (vu le 13 Mars 2013).

Statistics Quarterly; Vol. 43. No. 3, pages 139-139-44. (1990). (vu le 02 Avril 2013).

UNEP et al(2002). (vu le 22 Mars 2013).

www. environnement.gov.ml/uploads/pasp/Dangers-risques-pesticides-Final. pdf. (vu le 13 Mars 2013).

www. observatoirepesticides. gouv. Fr . (vu le 19 Mars 2013).

www. senat, fr/rap/102-215-2/102-215-21pdf . (vu le13 Mars 2013).

www.communityipm.org/toxictrail/ Documents/Jeryaratnam-WHO1990.pdf. (vu le 11 Mars 2013).

www.en.wikipedia.org/wiki/Ibid. (vu le14 Mars 2013).

www.en.wikipedia.orgwikiToxicity_Class.08. (vu le 08.Mars.2013).

www.era-orleans.org/AFIP/FR/pesticides.html. (vu le 11 Mars 2013).

www.fao.orgagAGP...printfriendly.html . (vu le 17 Mars 2013).

www.fr.wikipedia.org/wiki/Pesticide#.C3.89tymologie. (vu le 03 Mars 2013) .

www.fr.wiktionary.org/wiki/ibid.The WHO Recommended Classification of Pesticides by Hazard; 2004. (vu le 29 Mars 2013).

www.notre-planete.info/ecologie/alimentation/pesticides_0.php. (vu le12 Mars 2013).

www.linkedin.com/in/stephanielynnwatts. (vu le 14 Mars 2013).

www.who.int/ipcs/ publications/pesticides_hazard_rev_3.pdf. (vu le 29 Mars 2013).

www.who.int/nmh/donorinfo/msd_ preventing_suicide.pdf. (vu le 10 Mars 2013).

Z. Fergani et N. Mami : mémoire de master ‘’prédiction neuronale de la toxicité des pesticides’’, Universite Dr. Yahia Fares de Medea ; Algérie. (2011).

CHAPITRE 2

D. Rouvray ; Bonchev; Danail « Chemical graph theory: introduction and fundamentals ». Tunbridge Wells; Kent; England: Abacus Press. ISBN 0-85626-454-7. (1991). (vu le 16 Avril 2013).

D. Weininger “SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules“. Journal Chemistry Information Computational Science; 28. pp. 31-36. (1988) (vu le 17 Avril 2013).

EK . Freyhult; K. Andersson ; MG. Gustafsson "Structural modelin extends QSAR analysis of antibody-lysozyme interactions to 3D-QSAR" Biophysical Journal 84 (4): 2264–72. doi:10.1016/S0006-3495(03)75032-2. PMC 1302793. PMID 12668435. (April 2003). (vu le 12 Avril 2013).

G. David ; O. Richard; W. Peter ; Hart; Strok; Pattern classification. Chichester: John Wiley & Sons. ISBN 0-471-05669-3. (2001). (vu le 15 Avril 2013).

G. Veith; S. Broderius; In QSAR in Environmental Toxicology II; Kaiser, K.L.E; ed; D. Reidel: Dordrecht; Netherlands; p.385. (1987). (vu le 13 Avril 2013).

H. Könemann; Toxicol. 19.209. (1981). (vu le 16 Avril 2013).

H. Waterbeemd; S. Rose; "Quantitative approaches to structure-activity relationships"; in *Book* "Quantitative approaches to structure-activity relationships". Elsevier. 351-367. (2003). (vu le 17 Avril 2013).

H. Corinne ; " Recherche D'inhibiteurs De Haute Affinite De L'adn Polymerase Beta Par Criblage Virtuel Et Rmn " Thèse de doctorat. Université De Toulouse (France).(2008). (vu le 09 Avril 2013).

J. Dearden; G. Schüürmann, « Environ. Toxicol. Chem ». in press. Received: June (2001). (vu le Avril 2013).

J. Dearden; M. Barratt; R. Benigni; D. Bristol, R. Combes; Cronin, M.T.D; Judson; P.N; Payne; M.P; Richard; A.M; Tichý, M; Worth; A.P.; Yourick, J.J; ATLA. 25. 223. (1997). (vu le 12 Avril 2013).

J. Dearden; Quant. Struct-Act. Relat. Cronin; M.T.D. 14. 1. (1995). (vu le 11 Avril 2013).

J. Hermens; J. Bruijn; J. Pauly; W. Seinen, W. « In QSAR in Environmental Toxicology » II; Kaiser; K.L.E; ed; D. Reidel: Dordrecht, Netherlands, p.135.(1987). (vu le 14 Avril 2013).

M. Nendza; « Structure-Activity Relationships in Environmental Sciences »; Chapman & Hall: London; (1998) . (vu le 18 Avril 2013).

N. Greene; « Drug Delivery » Rev. 54. 417.(2002). (vu le 07 Avril 2013).

O. Mekenyan; G. Veith; « SAR QSAR Environ » Res. 1, 335.(1993). (vu le 18 Avril 2013).

R. Fraczkiewicz ;"In Silico Prediction of Ionization". Comprehensive medicinal chemistryII. Amsterdam: Elsevier. ISBN 0-08-044518-7. (2007). (vu le 11 Avril 2013).

S. Bradbury; T. Henry; R. Carlson« In Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environment Chemistry and Toxicology »;W. Karche; J. Devillers, eds; Kluwer Academic Publishers: Dordrecht; Netherlands; 1990; p. 295; Sar Qsar Environ. Res. 2. 89. (1994). (vu le 09 Avril 2013).

T. Hendrik ; R. Todeschini ; C. Viviana ; M. Raimund ; K. Hugo; « Handbook of Molecular Descriptors ». Weinheim: Wiley-VCH. ISBN 3-527-29913-0.(2002). (vu le 28 A vril 2013).

T. Schultz; Cronin; M.T.D; « Chem. Res. Toxicol ». 14. 1284. (2001). (vu le 11 Avril 2013).

www.bkchem. zirael. org/index. html. (vu le 14 Avril 2013).

www.fr.wikipedia.org/wiki/Relation_quantitative_structure_%C3%A0_activit%C3%A9. (vu le 25 Avril 2013).

www.neurones.espci.fr/Theses_PS/These%20AG-convertie-PASTEL.pdf . (vu le 25 Avril 2013).

Y. Ami; N. Melzi: mémoire de master 'Prédiction neuronale de l'effet toxicologique des fongicides.'; Universite Dr. Yahia Fares de Medea; Algérie _2011).

CHAPITRE 3

A . Dybs and R.V. Edwars; ″ Fundamentals of Transport Phenomena in Porous Media″ edited by J. Bear and Y. Corapcioglu ; Martinus Nishoff Publishers; p.199 (1984) . (vu le 09 Mai 2013).

A . J . Karabelas; T . H . Wegner and T . J . HANRATTY Chem . Eng . Sci; 26. 1581. (1971) . (vu le 15 Mai 2013).

A . Gaunand et F . Coeuret ; Electochim . Acta ; 23 . 1197. (1978) . (vu le 05 Mai 2013).

A . S. Gupta and G. Thodods; A.I.Ch.E.J.9.751. (1963). (vu le 09 Mai 2013).

A . S. Gupta and G. Thodods; Ind. Eng. Chem. Fund; 3.218. (1964). (vu le 09 Mai 2013).

A. Duprat; T. Huynh; G. Dreyfus; Towards a Principled Methodology for Neural Network Design and Performance Evaluatioon in QSAR ; Application to the Prediction of LogP, Journal of Chemical Information an Computer Sciences; 38. 586-594. (1998). (vu le 01 Mai 2013).

A. Firoozabadi and D.L. Katz ; ″ An Analysis of High Velocity Gas Flow Through Porous Media ″; Draft Paper; oct (1976). (vu le 05 Mai 2013).

A. Kumar; G.N. Tiwari; A.I Journal of Food Engineering; doi: 10.1016/j.jfoodeng. 04.026. (2006). (vu le 10 Mai 2013).

A. Tsutsumi; Wie Chen; Neural Network for Prediction of the Dynamic Heat-Transfer Rate in Bubble Columns; Ind. Eng. Chem. Res. 40. 5358-5361. (2001). (vu le 01Mai 2013).

A.E. Scheidegger "The Physics of Trough Porous Media". University of Toronto (1960) .(vu le 27 Mai 2013).

B . J . Gaffney And T . B . Drew; Ind . Eng . Chem ; 42 . 1120. (1950) . (vu le 10 Juin 2013).

B . P . Le Clair and A . E . Hamielec,Ind . Eng . Chem . Fund ; 7 . 542 . (1968) . (vu le19 Mai 2013).

B . P . Le Clair and A . E . Hamielec ; J. Can . Chem . Eng . 49 . 713. (1971) . (vu le 08 Mai 2013).

B. DasGupta; G. Schnitger; The Power of Approximating: a Comparison of Activation Functions; In S. J. Hanson, J.D. Cowan, and C.L. Giles editors, Advances in Neural Networks Information Processing Systems; 5. 615-622. (1993). (vu le 04 Mai 2013).

B. Hassibi et D.G. Stork; Second order Derivatives for Network Pruning : Optimal Brain Surgeon, In S. J. Hanson; J. D. Cawnan and C.L. Giles editors, Advances in Neural Information Processing Systems 5. 164-171. (1993). (vu le 04 Mai 2013).

B. Widrow et M.E. Hoff; Adaptive Switching Circuits; Ire Wescon; New York. (1960), 96-104.
C . L . Kusik and J . Happel; Ind . Eng . Chem . Fund ; 1 . 163 . (1962) . (vu le 11 Mai 2013).

C. Dryden; D. Strang; and A. Withrow; Chem. Eng . Prog ; 49. 191. (1953) . (vu le 02 Mai 2013).

C. Touzet; Les Réseaux de Neurones Artificiels : Introduction au Connexionnisme; (1992). (vu le O7 Juin 2013).

D. A. Nield et A. Betan.« Convection in porous media ». Spriger-Verlag;(1992) . (vu le 29 Mai 2013).

D. Broomhead et D. Lowe; Multivariable Functional Interpolation and Adaptive Networks; Complex Systems. 322-355. (1988). (vu le 31 Mai 2013).

D. Dawson and O. Trass ; Int . J . Heat Mass Transfer ; 15 . 1317. (1972) . (vu le 18 Mai 2013).

D. Hassani; Méthodes des Réseaux de Neurones pour le Calcul de la Dispersion Axiale dans les Lits Fixes; Mémoire de Thèse de Magister; Centre Universitaire de Médéa; (2002-2003) . (vu le 31 Mai 2013).

D. Hebb; The Organization of Behavior; Wiley; New York. (1949) . (vu le 30 Avril 2013).

D.E. Rumelhart; G.E. Hinton et R.J. Williams; Learning Internal Representations by Error Propagation; In D.E. Rumelhart and J. McClelland; Editors; Parallel Distributed Processing. MIT Press; Cambridge; (1986). (vu le 30 Avril 2013).

D.W. Marquardt; an Algorithm for Least-Squares Estimation of Non-linear Parameters; Journal of the Society of Industrial and Applied Mathematics 11. 431- 441. (1963).

Dwivedi and Upadhyay; Ind.Eng.Chem.Process Des.Develop;16.1657.(1977) . (vu le 06 Mai 2013).

E . J . Wilson and C . J . Geankoplis; Ind . Eng . Chem . Fund ; 5 . 9. (1963) . (vu le 21 Mai 2013).

E. Davalo et P. Naïm; Des Réseaux de Neurones; Eyrolles; 2ème Edition, Mars. (1993) . (vu le 01 Mai 2013).

E.J. Wilson and C.J. Geankoplis : Ind. Eng. Chem. Fundam; 5. 9. (1966) . (vu le 21 Mai 2013).

F . Coeuret et A . Storck ; ″ Elements de Genie Electrochimique ″ ; Technique et Documentation (Lavoisier); Paris (1984) . (vu le 11 Mai 2013).

F . Coeurt ; These d'état ; I . N . P . L ; Nancy (1974) . (vu le 27 Mai 2013).

F . Coeurt; Electochim . Acta; 21 . 185 . (1976) . (vu le 28 Mai 2013).

F. A. L. Dullien; ″Porous Media : Fluid Transport and Pore Structure ″ Academic Press New York (1979). (vu le 07 Mai 2013).

F. Elie; Conception et Réalisation d'un Système Utilisant des Réseaux de Neurones pour l'Identification et la Caractérisation; à Bord de Satellites; de Signaux Transitoires de type Sifflement " Thèse de Doctorat, Université d'Orléans" (1997) . (vu le 13 Juin 2013).

F. Rosenblatt; Principles of Neurodynamics; Spartan Books; New York. (1962) . (vu le 01 Mai 2013).

F. Kreith; Principles of Heat Transfer International Textbook Company Scranton; Pensylvania (U.S.A.)
G . C . Evans and C . A . Gerald; Chem . Eng . Proc. 49 . 135 . (1953) . (vu le 13 Juin 2013).

G. Burel; Réseaux de Neurones en Traitement d'Image : des Modèles Théoriques aux Pratiques Industrielles; Thèse de doctorat de l'Université de Bretagne Occidentale. (1991) . (vu le 12 Mai 2013).

G. Delaunay ; A. Stork ; A. Laurent ; et J. Charpentier . Proc. Des . Dev . 19 . 514 . (1980) . (vu le 14 Mai 2013).

G. Dreyfus et al. ; Réseaux de Neurones Méthodologie et Applications; Editions Eyrolles; Paris. (2002) . (vu le 11Mai 2013).

G. Hinton; Connectionnist Learning Procedures; Artificial Intelligence. (1989) . (vu le 02 Mai 2013).

G.E. Hinton; T.J. Sejnowski et D.H. Ackley; Boltzmann Machines : Constraint Satisfaction Networks that Learn; Rapport Technique TR-CMU-CS-84-199; Université Carnegie Mellon; Pittsburgh USA. (1984) . (vu le 10 Juin 2013).

H . E . Hoelscher; A. I . Ch . J . 4 . 300 . (1958) . (vu le 31 Mai 2013).

H .Olive et G . Lacoste ; Electrochim . Acta . 24 . 1109 . (1979) . (vu le 09 Mai 2013).

H . Rumpf and A . R . Gupte; Chem . Ing . Tech . 43 . 367. (1971) .(vu le 28 Mai 2013).

H. Abdi : Les réseaux de Neurones. Sciences et Technologies de la Connaissance; PUG; (1994) . (vu le 30 Mai 2013).

H.N. Mhaskar; C.A. Micchelli; How to Choose an Activation Function; In J.D. Cowan; G. Tesauro and J. Alspector; editors; Advances in Neural Information Processing Systems; 6 . 319-326. (1994). (vu le 29 Mai 2013).

H.P.C. Darcy "Les fontaines publiques de la ville de DIJON" V. Dalmont; Paris (1956) . (vu le 24 Mai 2013).

Hines and Maddox ; Mass Transfer : Fundamentals and Applications; Prentice Hall; (1985) . (vu le 02 Juin 2013).

I. F . Macdonald; M.S. El Sayed; K. Mow and F.A.L. Dullien . Ind. Eng. chem. Fundamentals; 6. 393. (1979) . (vu le 30 Mai 2013).

J . E . Williamson ; K . E . Bazaire and C . J . Geankoplis; Ind . Eng . Chem . Fund . 2 . 126. (1963) . (vu le 20 Mai 2013).

J . Happel and H . Brenner ″ Low Reynolds Hydrodynamic ; with Special Application to Particulate Media ″ Prentice Hall; Englewoods Cliffs ; New Jersy. (1965) . (vu le 03 Mai 2013).

J . J . Carberry; A . I . Ch . E . J . 6 . 460 . (1960) . (vu le 10 Juin 2013).

J . S . B . Kozeny Akad ; Wiss . Wien Abt IIa . 136 . 271 . (1927) . (vu le 11 Mai 2013).

J. Gosse ; Guide technique de thermique. Dunod . (1981). (vu le 01 Juin 2013).

J. Padet ; fluide en écoulement; méthodes et modèles; Masson. Et : Echangeurs Thermiques . (1991) . (vu le 22 Mai 2013).

J.A. Freeman et D.M. Skapura; Neural Networks Algorithms; Applications and Programming Techniques; Addison-Wesley Publishing Company; New Yor. (1992) . (vu le 02 Mai 2013).

J.J. Hopfield; Neural Networks and Physical Systems with Emergent; Collective Computational Abilities; *Proc. Natl. Acad. Sci*. 79. 554-2558. (1982). (vu le 02 Mai 2013).

K. A . Jolls and T . J. Hanratty; A . I . Ch . E . J . 15. 199. (1969) . (vu le 13 Mai 2013).

K. Hornik; Approximation Capabilities of Multilayer Feedforward Networks; Neuronal Network; 4. p. 251-257. (1991). (vu le 04 Mai 2013).

K. Levenberg; A Method for the Solution of Certain Non-linear Problems in Least Squares; Quartely Journal of Applied Mathematics II; 164-168. (1944) . (vu le 13 Mai 2013).

L . K . Mac Kune and R . H . Wilhelm ; Ind . Chem . 41 . 1124. (1949) . (vu le 30 Mai 2013).

L. Bottou; Une Approche Théorique de l'Apprentissage Connexionniste; Applications à la Reconnaissance de la Parole" Thèse de Doctorat de l'Université de Paris Sud, Centre d'Orsay". (1991) . (vu le 12 Mai 2013).

L. Buniet; Traitement Automatique de la Parole en Milieu Bruité : Etude de Modèles Connexionnistes Statiques et Dynamiques"; Thèse de Doctorat; Université Henri Poincaré Nancy 1. (1997) . (vu le 08 Mai 2013).

M . A . Enriquez Granados; D . Hutin et A . Storck , Electochim . Acta. 27. 303. (1982) . (vu le 11Juin 2013).

M . Lebouche –Transfert de matière en régime de couche limite bidimensionnelle et à grand nombre de Schmidt C.R Acad. Sciences Paris ; t. 270.pp1757 -1760. (1970) . (vu le 19 Mai 2013).

M . Lebouche and M . Martin - Convection forcée autour d'un cylindre ; sensibilité aux pulsations de l'écoulement externe . Int . J .of Heat and Mass Transfer , vol . 18. pp .1161-1175. (1975). (vu le 26 Mai 2013).

M . M . El Kaissy and G . M . Homsy; Ind . Eng . Chem . Fund . 12 . 82 (1973) . (vu le 11 Juin 2013).

M . Hobson and G .Thodos ; Chem . Eng . Prog . 45 . 517 . (1949) . (vu le 02 Juin 2013).

M. Powel; Radial Basis Function for Multivariable Interpolation; In IMA Conference on Algorithms for the Approximation of Functions and Data; RCMS Shrivenham; 143-167. (1985). (vu le 27 Mai 2013).

M.R. Berthold et J. Diamond; Boosting the Performance of RBF Networks with Dynamic Decay Adjustment, Forschungszentrum Informatik, Uni Karlsruhe; and Intel Corporation. (1995) . (vu le 28 Mai 2013).

M.R.J. Fombellida; M.J.M. Minsoul et J.L.O. Destine; Perceptrons Multi-Couches et Fonctions d'Activation Non-Monotones; In Proc. Neuronimes'90, EC2 & Cie, Nîmes. (1990) . (vu le 31 Mai 2013).

N . Wakao and T . Furnazkri Chem . Eng . Sci . 33 . 1375. (1978) . (vu le 19 Mai 2013).

N. Fröessling; Physicochemical Hydrodynamics; Prentice-Hall; Englewood Cliffs; N.J. (1962). (vu le 24 Mai 2013).

O. Reynolds; ″Papers on Mechanical and Physical Subjects″ Cambridge University Press (1900). (vu le 10 Mai 2013).

P . C . Carman; Trans . Inst . Chem . Eng . 15 .50 . (1937) . (vu le 10 Mai 2013).

P . N . Dwivedi and S . N . Udadhyay; Ind . Eng . Chem . Proc . Des . Dev . 16 . 157 . (1977) . (vu le 08Mai 2013).

P . S . Fedkiw and J . Newman; Int . J . Heat Mass Transfer ; 25 . 935. (1982) . (vu le 01 Juin 2013).

P.H . Forcheimer ; Z. Ver. Deutch . Ing . 45. 1781 . (1901) . (vu le 04 Juin 2013).

P.Y . Polubarinova –Kochina; ″Teorie dvizheniya gruntovykh vod Gosudarstv″; Izdat . (vu le 11 Mai 2013).

R . E . Sioda; J . Appl, J . Electrochem ;5 . 221. (1975) . (vu le 24 Mai 2013).

R . P . Pfeffer and J . Happel ; A . I . Ch . E . J . 10 . 605 .(1964) . (vu le 11 Mai 2013).

R . Alkire ; B . Gracon , T . Grueter ; J . Marek and P . Blackburn ; J . Electochem . Soc . 127 . 1085 . (1980) . (vu le 01 Juin 2013).

R. L. Steinberger and R. E. Treybal, A. I. Ch.E.J. 6. 227. (1960) . (vu le 02 Mai 2013).

R. Tibshirani; A Comparison of Some Error Estimates for Neural Network Models" Neural Computation" 8. 152-163. (1996). (vu le 02 Mai 2013).

R. Zemouri; D. Racoceanu; et N. Zerhouni; Application des Réseaux de Neurones Récurrents (RBF) à la Détection Dynamique de Dégradations; Les Journées de travail du GRP GRENOBLE; LAG – ENSIEG 21 - 22 mars (2002) . (vu le 14 Mai 2013).

R.H. Notter and C.A. Sleicher: Chem.Eng.Sci. 26. 161. (1971) . (vu le 22 Mai 2013).

R.L. Steinberger and R.E. Treybal: AIChE J. 6. 227. (1960) . (vu le 26 Mai 2013).

R.L. Watrous; Learning Algorithms for Connectionist Networks: Applied Gradient Methods for Nonlinear Optimization; In Proceedings of the 1st International Conference on Neural Networks; San Diego; IEEE; Press; New York; 2. 619-627. (1987). (vu le 12 Mai 2013).

R. Welty,Charles James ; E. Wicks; Robert E.Wilson ; Fundamentals of Momentum; Heat; and Mass transfer, Corvallis, Oregon. (1983) . (vu le 02 Mai 2013).

S . Kumar; S .N . Uparhyay and V . K . Mathur ; Ind . Eng . Chem . Proc . Des . Dev . 16 . 1 . (1977) . (vu le 12 Mai 2013).

S . Upadhyay and G . Tripathi ; J . Chem . Eng . Data; 20 . 20 . (1975) . (vu le 17 Mai 2013).

S .Banerjee ; Chem . Eng . Sci . 26 . 989. (1971) . (vu le 02 Mai 2013).

S .D . Venkateswaran and G .S . Laddha Indian . Chem . Eng . 8 . 33 . (1966) . (vu le 17 Mai 2013).

S. Boucheron; Théorie de l'Apprentissage; Hermès, Paris; (1992) . (vu le20 Mai 2013).

S. Ergun ; Chem. Eng. Prog. 48 . 98. (1952) . (vu le 07 Mai 2013).

S. Thiria; Y. Lechevallier; O. Gascuel et S . Canu, Statistiques et Méthodes Neuronales; Dunod; Paris; (1997) . (vu le 13 Mai 2013).

S. Evnochides and G. Thodos; A.I.Ch.E.J. 6. 178. (1959) . (vu le 19 Mai 2013).

S.Irmay; Trans . Am . Geophys . Un . N° 4 . 39 . 702. (1959) . (vu le 17 Mai 2013).

Skelland ; Diffusional Mass Transfer; Wiley. (1974) . (vu le 26 Mai 2013).

T . Miyauchi ; K . Matsumoto and K. Yoshika ; J. Chem. Eng. Japan. 8. 228. (1975) . (vu le 21 Mai 2013).

T. Kohonen, Self Organization and Associative Memory; Springer-Verlag, New York; (1984) . (vu le 23 Mai 2013).

T. W. Jonshon and D. B. Taliaferro ;U.S. Bur . Mines Tech. Paper. 592. (1938) . (vu le 06 Mai 2013).

T.K. Sherwood; and E.R. Gilliland: Ind.Eng.Chem.26.516. (1934) . (vu le 23 Mai 2013).

T.K. Sherwood; R. L. Pigford; and C. R. Wilke; Mass Transfer; McGraw-Hill Book Company; New York; 1975. (vu le15 Mai 2013).

Tech. Theoret. Lit Moscou (1952). Traduction anglaise de R . De WIEST ″theory of Groundwater Movement″ Princeton Univ. Press . Princeton, New Jersey. (1962) . (vu le 16 Mai 2013).

W . A . Selke; Y . Bard; A . D . Pasternak and S . K . Adita; A . I . Ch . E . J . 2 . 468 . (1956) . (vu le 27 Mai 2013).

W . E . Ranz and W . R . Marshall; Chem . Eng . Proc . 48 . 173 . (1952) . (vu le 12 Mai 2013).

W. E. Dunn; C . F .Bonilla; C. Ferstenberg. and B. Gross; A . I . Ch . E . J . 2 . 184. (1956) . (vu le 04 Mai 2013).

W.H. Linton and T.K. Sherwood: Chem. Eng.Prog. 46. 258. (1950) . (vu le 27 Mai 2013).

W.S. McCulloch et W. Pitts; A logical Calculus for the Ideas Immanent in Nervous Activity. Bulletin of Mathematical Biophysics; 5. 115-133. (1943). (vu le 19 Mai 2013).

Wilson and Geankoplis; Ind. Eng. Chem. Fundam. 5. 9. (1966) . (vu le 21 Mai 2013).

Y . Kawase and J . J . Ulbrecht,Ind . Eng .Chem . Fund . 24 . 115. (1985) . (vu le 07 Mai 2013).

Y . Y . Wang; C . H . Tseng; G . C . Wan and W . Maok; Chem . Eng . Sci. 37. 939 . (1982) . (vu le 20 Mai 2013).

Y . Nishimura and T . Ishit ; Chem . Eng . Sci . 35 . 1206. (1980) . (vu le 22 Mai 2013).

Y. Lee; S. Hoon Oh et Myung Won Kim; The Effect of Initial Weights on Premature Saturation in Back-Propagation Learning, International Joint Conference on Neural Networks. 1. 765-770 . (1991) . (vu le 22 Mai 2013).

Yoshida ; Ramaswami; Hougen; and J. Aiche ;8. 5. (1962) . (vu le 22 Mai 2013).

Introduction

Chapitre I

Chapitre II

Chapitre III

Chapitre IV

Conclusion

Références bibliographiques