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

UNIVERSITE DE MEDEA

Département du Génie des Procédés et Environnement

Laboratoire de Biomatériaux et Phénomènes de Transport (LBMPT)

LBMPT
Biomaterials and transport phenomena laboratory
University of MEDEA
ALGERIA

THESE DE DOCTORAT

Présentée par :

Mabrouk HAMADACHE

à l'Université de Médéa
pour l'obtention du titre de Docteur en Sciences

MODELISATION EN VUE DE LA PREDICTION OU DE LA CORRELATION DE L'ACTIVITE TOXICOLOGIQUE A PARTIR DE LA STRUCTURE MOLECULAIRE

Membres du jury :

Mr Mustapha OUMOUNA	Professeur (Université de Médéa)	Président
Mr Mohamed AMARI	Professeur (USTHB - Alger)	Examinateur
Mme Nadia A. LAOUFI	Maitre de Conférences A (USTHB - Alger)	Examinatrice
Mr Salah HANINI	Professeur (Université de Médéa)	Directeur de thèse
Mr Cherif SI MOUSSA	Maître de Conférences B (U. de Médéa)	Invité

REMERCIEMENTS

Je voudrais, en premier lieu, exprimer mes sincères et ineffables remerciements à Mr le Professeur Salah HANINI qui a assuré l'encadrement de cette thèse et qui a bien voulu m'accueillir au sein de son équipe. Tout au long de cette thèse, il n'a cessé de suivre et de s'intéresser à son avancement. Qu'il trouve ici l'expression de ma reconnaissance et de ma profonde gratitude.

Le Docteur Cherif SI MOUSSA s'est occupé du co-encadrement de ce travail. Son aide précieuse dans la compréhension des éléments de base des réseaux de neurones m'a été bénéfique. Je lui réitère à cette occasion mes sincères remerciements.

Ce serait de l'ingratitude si j'ose oublier l'aide précieuse et désintéressée du Docteur Latifa KHAOUANE. A chacune de mes multiples interrogations, son aide et ses conseils étaient spontanés. Pour toute cette disponibilité, je ne peux que la remercier du fond du cœur.

Je dérogerai à la règle coutumière qui stipule qu'on ne doit pas remercier son frère pour services rendus. Aussi, je tiens à remercier chaleureusement mon frère le Docteur Othmane BENKORTBI pour toute son aide multiforme tout au long de la réalisation de ce projet. A aucun moment, il n'a cessé de suivre et de participer passionnément à la progression de cette thèse.

Je tiens à remercier Mrs le Professeur M.OUMOUNA, le Professeur M.AMARI et le Docteur N.A. LAOUFI pour l'immense honneur qu'ils me font pour avoir accepté de participer au jury de cette thèse.

Je remercie également le Professeur Abdeltif AMRANE de l'ENSC de Rennes pour m'avoir accueilli au sein de son équipe durant les stages de courte durée. Il n'a cessé de s'intéresser à mon projet malgré le fait que le thème du travail ne coïncidait pas avec son profil. Je ne saurais oublier son aide spontanée dans les relectures de la publication à chaque fois que c'était nécessaire.

Mes remerciements les plus particuliers, je les adresse à Monsieur CHEBAIKI Saadane, recteur de l'université de Médéa de 2005 à 2012 pour m'avoir "obligé" à m'inscrire en doctorat alors que je faisais partie du staff administratif de la faculté des Sciences et Technologie. Qu'il sache que je lui serais toujours reconnaissant surtout que cette inscription a vu le jour après trois décennies depuis l'obtention du diplôme de Magister.

Je ne saurais terminer cette page sans remercier l'équipe administrative de la faculté des sciences et de la technologie en l'occurrence Mrs Allel AZIZI, Smail BOUHRAOUA, Billal DOUAKH, Slimane NEGUEB, Mokhtar FEZZANI, Mohamed BOUROUILA, Melles Dalila KADI et Karima HADJERI pour toute leur aide.

Mabrouk HAMADACHE

الملخص

نمذجة من أجل تنبؤ أو ربط النشاط المتعلق بالسمية عن طريق البنية الجزيئية

موضوع هذا البحث مفيد بشكل خاص لدراسة العلاقات بين بنية الجزيئات (مثل مبيدات الآفات) وسميتها على الإنسان والبيئة.

في الواقع، التنبؤ المسبق بسمية مادة دون اللجوء إلى التجارب الشاقة على الحيوانات هو في حد ذاته تحديا. نماذج (**العلاقة الكمية البنية – النشاط: QSAR**) التي تربط البنية والسمية تم اثباتها في أيامنا هذه وهي أدوات هامة للتنبؤ بسمية مادة وضعت حديثا في السوق.

الهدف المعين لهذا العمل هو تطوير نماذج QSAR للتنبؤ بالسمية الفمية الحادة لمبيدات الجرذان وهذا باستخدام الشبكات العصبونية الاصطناعية. لهذه الغاية تم تطوير أربعة نماذج على التوالي: مبيدات الأعشاب ومبيدات الحشرات ومبيدات الفطريات ومبيدات الآفات بشكل عام. مختلف عناصر تحليل النتائج القياسية لهذه النماذج أثبتت موثوقية جيدة وقوة تنبؤ ممتازة. عموما، تم التنبؤ لأكثر من 90٪ من السميات المجربة بشكل صحيح. ومع ذلك، فإن النماذج الأربعة المقدمة، التي تم الحصول عليها لمبيدات الآفات هي الأكثر دقة.

مقارنة مع النماذج القليلة التي نشرت في المقالات العلمية فإن النموذج الذي تم تطويره خلال هذا العمل هو الأحسن خاصة وأنه يلبي التوصيات الصارمة لمنظمة التعاون الاقتصادي والتنمية (OECD) للمصادقة على نماذج QSAR.

كلمات المفاتيح :المبيدات الآفات، التنبؤ، (العلاقة الكمية البنية – النشاط: QSAR)، السمية الحادة، الشبكات العصبونية، (MLR: الانحدار الخطي المتعدد).

Résumé

Modélisation en vue de la prédiction ou de la corrélation de l'activité toxicologique à partir de la structure moléculaire.

Le thème de ce travail de recherche est particulièrement utile pour l'étude de relations entre la structure des molécules (par exemple les pesticides) et leur nocivité sur l'homme et sur l'environnement. En effet, prédire à l'avance la toxicité d'une substance sans recourir aux fastidieuses expérimentations animales est en soi un challenge. Les modèles QSAR mettant en relation la structure et la toxicité se sont avérés, de nos jours, des outils importants pour prédire la toxicité d'une substance nouvellement mise sur le marché.

L'objectif assigné à ce travail est d'élaborer des modèles QSAR pour la prédiction de la toxicité orale aigüe des pesticides sur les rats et ce par utilisation des réseaux de neurones artificiels. A cet effet, quatre modèles ont été élaborés respectivement pour les herbicides, les insecticides, les fongicides et les pesticides en général. Les différents éléments d'analyse de la performance de ces modèles font ressortir leur bonne fiabilité et leur excellent pouvoir de prédiction. Dans l'ensemble, plus de 90% des toxicités ont été correctement prédites. Néanmoins, des quatre modèles élaborés, celui obtenu avec les pesticides est le plus performant.

Par comparaison avec les rares modèles parus dans la littérature scientifiques, celui élaboré lors de ce travail est plus performant d'autant plus qu'il satisfait aux recommandations rigoureuses de l'organisation de coopération et de développement économiques (OCDE) en matière de validation des modèles QSAR.

Mots clés : Pesticides, QSAR, Prédiction, Toxicité aigüe, Réseaux de neurones, MLR.

Summary

Modeling for the prediction and correlation of toxicological activity from the molecular structure.

The topic of this research is particularly useful for the study of relationships between the structure of molecules (eg, pesticides) and their toxicity to humans and the environment. Indeed, to predict in advance the toxicity of a substance without resorting (turning) to the boring animal experiments is in itself a challenge. A QSAR model relating the structure and toxicity is proved, nowadays, important tools for predicting the toxicity of a new marketing substance.

The objective set for this work is to develop QSAR models for the prediction of acute oral toxicity of pesticides in rats by using artificial neural networks. For that purpose , four models were respectively developed for herbicides, insecticides, fungicides and pesticides in general. The various analysis of the performance of these models factors highlight their good reliability and excellent predictive power. Overall, more than 90% of toxicities were correctly predicted. Nevertheless, the four models developed, that obtained with pesticides is the most efficient.

Compared with the few models appeared in the scientific literature, that developed during this work is more successful especially since it satisfies the rigorous recommendations of the Organization for Economic Cooperation and Development (OECD) regarding validation of the QSAR models.

Keywords: Pesticide, QSAR, Prediction, Acute toxicity, Neural network, MLR.

TABLE DES MATIERES

RESUME
LISTE DES TABLEAUX
LISTE DES FIGURES
ACRONYMES

INTRODUCTION GENERALE 1

Chapitre 1

REVUE CRITIQUE DE LA LITTERATURE SUR L'ACTIVITE TOXICOLOGIQUE

Introduction 8
1.1 Pesticides et pollution de l'environnement 9
1.1.1 Contamination de l'air 9
1.1.2 Contamination de l'alimentation 11
1.1.3 Contamination des milieux aquatiques 13
1.1.4 Contamination des sols 14
1.1.5 Contamination des travailleurs agricoles et non agricoles 15
1.2 Impact sanitaires avérés ou suspectes des pesticides 16
1.2.1 Pesticides et troubles de la reproduction 17
1.2.2 Malformations 18
1.2.3 Affection du système immunitaire 18
1.2.4 Effet sur les organismes aquatiques 19
1.2.5 Pesticides et cancer 19
1.2.6 Pesticides et diabète 20
1.2.7 Pesticides et pathologies neurologiques 20
1.2.8 Autisme 21
1.2.9 Affections respiratoires et pesticides 21
1.3 Toxicité et modèles de prediction QSAR 22
1.4 Modèles QSAR et prédiction de la toxicité aigüe des pesticides chez les rats 24
Conclusion 27

Chapitre 2

RELATIONS QUANTITATIVES STRUCTURE – ACTIVITE

2.1 Utilités et applications des méthodes QSAR 31

2.2 Outils de QSAR 32
2.2.1 Données biologiques 32
2.2.2 Les descripteurs moléculaires 32
2.2.2.1 Les descripteurs 1D 33
2.2.2.2 Les descripteurs 2D 33
2.2.2.3 Les descripteurs 3D 33
2.3 Formulation mathématique des QSAR 34
2.4 Methodes d'apprentissage statistiques 35
2.5 La regression linéaire multiple (*RLM*) 35
2.6 Les réseaux de neurones artificiels 36
2.6.1 Le neurone formel 37
2.6.2 Architecture d'un réseau de neurones 38
2.6.3 Perceptron multicouche 39
2.6.4 Expression algébrique de la régression neuronale 40
2.6.5 Optimisation de la structure du réseau 41
2.6.6 Mise en œuvre de la régression neuronale 42
2.6.7 Construction de la base de données et choix des échantillons 42
2.6.8 Conception du modèle non linéaire 43
2.6.9 Apprentissage 44
2.6.10 Algorithme de rétro-propagation 45
2.6.11 Apprentissage et sur-apprentissage 46
2.6.12 Validation du modèle 46
2.6.13 Détermination des performances du modèle 47

Chapitre 3

LES PESTICIDES

3.1 Historique 49
3.2 Classification des pesticides 50
3.2.1 Classement par cible 50
3.2.2 Classement par groupe chimique 51
3.2.3 Classement par mode d'action 51
3.3 Résistance aux pesticides 51
3.4 Toxicologie des pesticides 52
3.4.1 Toxicité aigüe 53
3.4.2 Toxicité chronique 53
3.4.3 Evaluation expérimentale de la toxicité aigüe des pesticides 53
3.5 Utilisation des pesticides en Algérie 54
3.5.1 Usage et importation de pesticides 54
3.5.2 Contamination par les pesticides 55

Chapitre 4

MODELISATION DE LA PREDICTION DE L'ACTIVITE TOXICOLOGIQUE DES PESTICIDES

4.1 Matériel et méthodes 58
4.1.1 Collecte des données de la toxicité orale aigüe 58
4.1.2 Obtention des descripteurs 60
4.1.3 Prétraitement de la base de données 60
4.1.4 Base de données finale 62
4.1.5 Développement des modèles QSAR 62
4.1.5.1 La régression linéaire multiple 63
4.1.5.2 Les réseaux de neurones artificiels 63
4.1.6 Estimation de la performance d'un modèle 63
4.1.6.1 Validation croisée 64
4.1.6.2 Paramètres statistiques 64
4.1.6.3 Domaine d'applicabilité 65
4.1.6.4 Utilisation de l'expression mathématique du modèle 66
4.1.7 Analyse de la contribution des descripteurs 67
4.1.8 Outils informatiques utilises 68
4.2 Prediction de la toxicité des herbicides 69
4.2.1 Résultats et discussion 69
4.2.1.1 Sélection des descripteurs 69
4.2.1.2 Développement des modèles QSAR 72
a) Modèle RLM 72
b) Modèle RNA 77
4.2.1.3 Contribution des descripteurs 79
4.2.2 Comparaison entre les modèles RLM et RNA 80
4.2.3 Domaine d'applicabilité du modèle 81
4.2.4 Expression mathématique du modèle PMC-RNA obtenu 82
4.2.5 Prediction de la toxicité à partir de l'expression mathématique 83
4.2.6 Conclusion 86
4.3 Prediction de la toxicité des insecticides 87
4.3.1 Résultats et discussion 87
4.3..1.1 Prétraitement de la base de données 87
4.3.1.2 Normalisation et mise en forme des données 89
4.3.1.3 Réseau optimal 90
4.3.1.4 Domaine d'applicabilité 93
4.3.1.5 Analyse de la contribution des descripteurs 94

4.3.2 Conclusion 95
4.4 Prediction de la toxicité aigüe des fongicides 96
4.4.1 Résultats et discussion 96
4.4.1.1 Sélection des descripteurs pertinents 96
4.4.1.2 Fongicides utilisés 96
4.4.1.3 Normalisation et mise en forme des données 99
4.4.1.4 Elaboration du modèle neuronal 100
4.4.1.5 Domaine d'applicabilité du modèle 105
4.4.1.6 Analyse de la contribution des descripteurs 106
4.4.2 Conclusion 108
4.5 Modélisation neuronale de la toxicité des pesticides 109
4.5.1 Résultats et discussion 109
4.5.1.1 Collecte des données et prétraitement 109
4.5.1.2 Obtention et selection des descripteurs 109
4.5.1.3 Méthodologie d'élaboration du modèle neuronal optimal 110
a- Sélection de la répartition de la base de données 111
b-Sélection du nombre de couches cachées 112
c- Sélection de la fonction d'activation 112
d- Sélection du nombre de neurones dans la couche cachée 114
e- Sélection de l'algorithme d'apprentissage optimum 114
4.5.1.4 Structures du réseau neuronal optimum obtenu 115
4.5.2 Application du réseau de neurone optimal 116
4.5.3 Domaine d'applicabilité du modèle 119
4.5.4 Expression mathématique du modèle PMC-RNA obtenu 120
4.5.5 Prédiction de la toxicité à partir de l'expression mathématique 120
4.5.6 Analyse de la contribution des descripteurs 123
4.5.7 Conclusion 125
4.6 Comparaison entre modèles 126
4.6.1 Modèles établis dans ce travail 126
4.6.2 Comparaison avec les modèles de la littérature 127

CONCLUSION GENERALE ET PERSPECTIVES 128

BIBLIOGRAPHIE

ANNEXES

LISTE DES ACRONYMES

BFGS algorithm	Broyden–Fletcher–Goldfarb–Shanno algorithm
CAS number	Chemical Abstracts Service
CCA	Curvilinear Component Analysis
ECETOC	European Centre for Ecotoxicology and Toxicology of Chemicals
FAO	Food and Agriculture Organization
IPCS	International Programme of Chemical Safety
LD50	Lethal Dose 50
MLP-ANN	Multilayer Perceptron Artificial Neural Network
MLR	Multiple Linear Regression
PAN	Pesticide Action Network
PCA	Principal Component Analysis
PLS	Partial Least Squares
PMC	Perceptron Multi-Couches
PPDB	Pesticide Properties Data Base
QSAR	Quantitative Structure-Activity Relationship
QSPR	Quantitative Structure-Property Relationship
REACH	Réglement sur l'Enregistrement, l'Evaluation, l'Autorisation et les restrictions des substances CHimiques.
RMSE	Root Mean Square Error
RMSEP	Root Mean Square Error of Prediction test
RMSET	Root Mean Square Error of Training test
RNA	Réseau de Neurones Artificiels
SMILES	Simplified Molecular Input Line Entry Specification
SVM	Support Vector Machine
VIF	Variance Inflation Factor

LISTE DES FIGURES

		Page
Figure IG.1 :	Pourcentage d'animaux utilisés par catégorie en Europe en 2008 (a) et objet des expériences (b) (COM 2010)	4
Figure 1.1 :	Evolution du chiffre des ventes mondiales entre les années 2006 et 2011 (UIPP, 2012)	8
Figure 1.2 :	Nombre de publications traitant la contamination de l'environnement par les pesticides (PubMed consulté le 18/04/2014)	10
Figure 1.3 :	Nombre de publications inhérentes aux effets des pesticides sur la santé (PubMed consulté le 18/04/2014).	17
Figure 1.4 :	Nombre de publications inhérentes aux études QSAR de la toxicité des substances Chimiques (PubMed consulté le 18/04/2014).	23
Figure 2.1 :	Méthodologie du calcul de ε_i	36
Figure 2.2 :	Réseau de neurones artificiel à n variables d'entrée et une sortie	36
Figure 2.3:	Schématisation d'un neurone formel	38
Figure 2.4 :	Exemple d'architecture d'un *PMC* à n variables d'entrée, sept neurones cachés répartis en deux couches, et deux sorties.	40
Figure 2.5 :	Algorithme de rétropropagation du gradient d'erreur	45
Figure 2.6 :	Généralisation optimale	46
Figure 2.7 :	Schéma d'élaboration d'un modèle neuronal	47
Figure 3.1 :	Evolution des importations de pesticides pour les trois pays du Maghreb.	55
Figure 3.2 :	Evolution des importations de fongicides, d'herbicides et d'insecticides en Algérie	56
Figure 4.1 :	Capture d'écran du site intitulé PPDB	59
Figure 4.2 :	Capture d'écran du site intitulé PAN	59
Figure 4.3 :	Schématisation générale et répartition du contenu de la base de données	62

		Page
Figure 4.4 :	Capture d'écran des logiciels STATISTICA et MATLAB	68
Figure 4.6 :	Représentation graphique du modèle RLM de prédiction de la toxicité orale aigüe ($1/DL_{50}$) chez le rat d'une série d'herbicides	71
Figure 4.7 :	Représentation graphique du modèle PMC-RNA de prédiction de la toxicité orale aigüe ($1/DL_{50}$) chez le rat d'une série d'herbicides	79
Figure 4.8 :	Importance des descripteurs dans la prédiction de la toxicité orale aigüe des herbicides	80
Figure 4.9 :	Représentation graphique des résidus en fonction de la toxicité ($1/DL_{50}$) expérimentale.	81
Figure 4.10 :	Diagramme de Williams pour le modèle PMC-RNA de prédiction de la toxicité ($1/DL_{50}$) des herbicides.	82
Figure 4.11 :	Représentation schématique du réseau PMC-RNA optimisé	83
Figure 4.12 :	Courbe de corrélation entre les valeurs expérimentales et prédites par le modèle mathématique	86
Figure 4.13:	Nombre de Descripteurs initiaux (a) et nombre de descripteurs pertinents obtenus après exécution des 5 étapes (b).	88
Figure 4.14 :	Mise en forme des données	89
Figure 4.15 :	Courbe de corrélation des valeurs prédites de DL_{50} en fonction des valeurs expérimentales pour l'ensemble d'apprentissage.	90
Figure 4.16 :	Courbe de corrélation des valeurs prédites de DL_{50} en fonction des valeurs expérimentales pour l'ensemble de validation.	91
Figure 4.17 :	Diagramme de Williams pour le modèle de prédiction de la toxicité des insecticides	94
Figure 4.18 :	contribution relative des descripteurs dans la prédiction de la toxicité orale aigüe des insecticides.	95
Figure 4.19 :	Répartition de la base de données des fongicides	99
Figure 4.20 :	Courbe de corrélation des valeurs prédites de DL_{50} (des fongicides) en fonction des valeurs expérimentales pour l'ensemble d'apprentissage	101
Figure 4.21 :	Courbe de corrélation des valeurs prédites de DL_{50} (des fongicides) en fonction des valeurs expérimentales pour l'ensemble de validation	101
Figure 4.22 :	Résidus des valeurs de DL_{50} obtenus pour les fongicides en fonction de la DL_{50} expérimentale	105
Figure 4.23 :	Diagramme de Williams pour le modèle de prédiction de la toxicité des fongicides	106
Figure 4.24 :	Contribution (en %) à la toxicité orale aigüe des fongicides par type de descripteurs.	107

		Page
Figure 4.25 :	Contribution (en %) à la toxicité orale aigüe des fongicides par catégorie de descripteurs.	107
Figure 4.26 :	Effet de la répartition de la base de données sur l'erreur quadratique moyenne	112
Figure 4.27 :	Erreur quadratique moyenne pour différentes combinaisons des fonctions d'activation pour la couche cachée et la couche de sortie	113
Figure 4.28 :	Effet du nombre de neurones de la couche cachée sur l'erreur quadratique moyenne (EQM)	114
Figure 4.29 :	Erreur quadratique moyenne obtenue par utilisation de 4 types d'algorithme d'apprentissage.	115
Figure 4.30 :	Courbe de corrélation des valeurs prédites de $1/DL_{50}$ (des pesticides) en fonction des valeurs expérimentales pour l'ensemble d'apprentissage	116
Figure 4.31 :	Courbe de corrélation des valeurs prédites de $1/DL_{50}$ (des pesticides) en fonction des valeurs expérimentales pour l'ensemble de validation	117
Figure 4.32 :	Architecture finale du réseau optimal pour la prédiction de la toxicité orale aigüe des pesticides	117
Figure 4.33 :	Résidus des valeurs de $1/DL_{50}$ obtenus pour les pesticides en fonction de $1/DL_{50}$ expérimentale	118
Figure 4.34 :	Diagramme de Williams pour le modèle de prédiction de la toxicité des pesticides	119
Figure 4.35 :	Courbe de corrélation entre les valeurs expérimentales et celles prédites par le modèle mathématique.	123
Figure 4.36 :	Contribution relative des descripteurs dans la prédiction de la toxicité orale aigüe des pesticides	124

LISTE DES TABLEAUX

		Page
Tableau IG.1 :	Informations toxicologiques et écotoxicologiques requises par la réglementation REACH [selon le règlement CE N°1907/2006 UE, 2006] (Barjhoux, 2011).	3
Tableau 1.1 :	Vue d'ensemble des travaux antérieurs sur les modèles QSAR de prédiction de la toxicité orale aigüe pour le rat (Sazonovas et al., 2010).	25
Tableau 3.1:	Mode d'action de certains pesticides	52
Tableau 4.1 :	Nombre de descripteurs calculés	61
Tableau 4.2 :	Etapes de sélection des descripteurs pertinents.	70
Tableau 4.3:	Matrice des coefficients de corrélation des descripteurs sélectionnés.	72
Tableau 4.4 :	Valeurs du VIF, t-test et catégories des descripteurs sélectionnés.	73
Tableau 4.5 :	Valeurs des toxicités aigües expérimentales, prédites par RLM et par RNA	75
Tableau 4.6 :	Architecture du réseau de neurones optimisé	78
Tableau 4.7:	Comparaison des paramètres statistiques des deux modèles RLM et PMC-RNA	80
Tableau 4.8:	Poids et biais du réseau de neurone optimal	84
Tableau 4.9:	Valeurs de $1/DL_{50}$ des herbicides prédites à partir du modèle mathématique	85
Tableau 4.10 :	Descripteurs sélectionnés lors de l'étude des insecticides	87
Tableau 4.11 :	Structure du réseau de neurones optimal obtenu dans la modélisation de la toxicité aigüe.	90
Tableau 4.12 :	Valeurs des paramètres statistiques pour les ensembles d'apprentissage et de validation	91
Tableau 4.13 :	Valeurs de DL_{50} expérimentales et prédites par PMC-RNA des insecticides	92
Tableau 4.14 :	Descripteurs sélectionnés pour la prédiction des fongicides	96
Tableau 4.15 :	fongicides utilisés ainsi que leur écriture SMILES	97

Page

Tableau 4.16a :	Structure du RNA optimisé pour la toxicité aigüe des fongicides	100
Tableau 4.16b :	Valeurs de la toxicité (expérimentale et prédite) aigüe, des résidus normalisés et des points leviers des fongicides.	102
Tableau 4.17 :	Valeurs des paramètres statistiques pour les ensembles d'apprentissage et de validation des fongicides.	104
Tableau 4.18 :	Etapes de la sélection des descripteurs pertinents	110
Tableau 4.19 :	Différents types de répartition de la base de données	111
Tableau 4.20 :	Optimisation du nombre de couches cachées	112
Tableau 4.21 :	Valeurs de l'EQM pour différentes combinaisons de fonctions d'activation	113
Tableau 4.22 :	Algorithmes d'apprentissages utilisés.	115
Tableau 4.23 :	Structure du RNA optimisé pour la toxicité aigüe des fongicides.	116
Tableau 4.24 :	Valeurs des paramètres statistiques pour les ensembles d'apprentissage et de validation	118
Tableau 4.25:	Valeurs de $1/DL_{50}$ des pesticides prédites à partir du modèle mathématique	120
Tableau 4.26 :	Comparaison entre les 4 modèles de notre travail	126
Tableau 4.27 :	Comparaison entre notre modèle et ceux parus dans la littérature	127

INTRODUCTION GENERALE

Le terme pesticide désigne l'ensemble des produits chimiques, naturels ou de synthèse, ayant pour but de repousser ou d'éliminer les organismes jugés indésirables, qu'il s'agisse de plantes, d'animaux, de champignons ou de bactéries et ce durant la production, le stockage ou la commercialisation de produits agricoles, de denrées alimentaires, ou de bois. Ils servent également à combattre les différents vecteurs de maladies humaines ou animales et sont utilisés comme produits d'entretien.Actuellement, la majorité des pesticides est d'origine synthétique, c'est-à-dire obtenus par des procédés de synthèse organique.

L'utilisation généralisée des pesticides en agriculture a été créditée de l'augmentation spectaculaire des rendements des cultures en réduisant les menaces naturelles et est devenue une partie intégrante des pratiques agricoles. L'utilisation sanitaire des pesticides a été, elle aussi, couronnée de succès à la suite de l'éradication de certaines maladies. Par ailleurs, les pesticides ayant des propriétés antiseptiques et désinfectantes ont eu leur part d'utilisation en médecine. La gamme des pesticides disponibles s'est alors rapidement diversifiée et leur usage a connu un très fort développement au cours des décennies suivantes. Petit à petit, du fait de leur facilité d'emploi et de leur efficacité, ils se sont imposés dans les usages domestiques (anti-moustiques par exemple), dans le traitement de la voierie publique, des plantes ornementales des voies ferrés, des routes, des espaces verts et des jardins. Durant la seconde moitié du vingtième siècle, l'utilisation des pesticides a doublé tous les dix ans. De plus, la résistance acquise par certaines espèces vis-à-vis des pesticides a contraint les utilisateurs de ces produits à augmenter la quantité à utiliser afin que l'efficacité d'action soit optimale d'une part et à la diversification des molécules actives (recherche de nouveaux pesticides plus actifs) d'autre part.

Cette large utilisation des pesticides a été et est une arme à double tranchant. En dépit du fait qu'ils ont contribué quantitativement à l'amélioration de la production agricole et du bien-être de l'humanité, les pesticides se retrouvent disséminés dans les divers milieux environnementaux. En effet, de nombreuses études attestent de la contamination des eaux

(douces, souterraines et côtières), de l'air, des sols et desaliments (d'origine animale ou végétale) par les pesticides. La conséquence de cette pollution environnementale sur la santé suscite bien des inquiétudes et demeure une préoccupation de l'heure. En effet, de par leur toxicité aigüe ou chronique et leur persistance, les pesticides demeurent une menace directe pour les agriculteurs et leurs familles, la faune et la flore, mais aussi de façon indirecte pour tout organisme vivant.

Depuis les années 1980, un intérêt particulier a été porté sur l'étude de l'impact environnementale des pesticides. Une multitude de travaux de recherche et d'expertise sur site ont évoqué l'éventuelle implication des pesticides dans plusieurs pathologies, en particulier des pathologies cancéreuses, des maladies neurologiques et des troubles de la reproduction. Des études ont évoqué la présence de pesticides dans le lait maternel. On a rapporté aussi la diminution des effectifs d'insectes, d'oiseaux ou d'abeilles et de mammifères et ce à la suite de l'empoisonnement par les pesticides. La faune aquatique n'est pas en reste, puisqu'on a enregistré une mortalité spectaculaire de poissons. Des espèces aquatiques comme les algues ou les crustacés (Daphnies) sont aussi victimes de la contamination des eaux par les pesticides. A la suite de ces études, réparties sur plus de trente années, certains pesticides ont été interdits alors que l'utilisation d'autres a été réglementée.

Sur la base de la masse croissante de preuves de la recherche en laboratoire, d'études sur la faune, et de l'exposition humaine accidentelle, il est très clair que les pesticides actuellement utilisés constituent des menaces graves pour la santé humaine. Par ailleurs, Il est maintenant établi que les pesticides présentent un certain nombre de risques pour les différents compartiments de l'environnement (eau, air, sols, espèces vivantes), pour la santé des opérateurs et des populations humaines. La littérature consacrée à l'évaluation de l'impact environnemental des pesticides fait ressortir un certain nombre de paramètres qui y sont à l'origine. Nous pouvons les résumer comme suit :

- ✓ La toxicité du pesticide pour chacune des espèces.
- ✓ La quantité de pesticide appliquée et son mode d'application
- ✓ La vitesse de dégradation du pesticide dans les eaux de surface et les eaux souterraines, dans l'air et dans le sol.

Le marché industriel des pesticides étant en constante évolution, l'Union Européenne a mis en place une nouvelle réglementation (en 2006) concernant la production de composés chimiques intitulée REACH. Cette réglementation oblige les industries chimiques à fournir un cahier descriptif sur les composés chimiques qu'elles importent ou produisent. Parmi les informations demandées, l'influence de la molécule sur l'être humain (toxicologique) et sur

l'environnement (écotoxicologique) doit être particulièrement détaillée. En d'autres termes, la toxicité de chaque pesticide produit doit être mentionnée et ce pour quantifier ses effets environnementaux. Le Tableau IG.1 récapitule les informations toxicologiques et écotoxicologiques à renseigner lors de la procédure d'enregistrement des substances auprès de l'Agence Européenne des produits chimiques en fonction du volume annuel produit ou importé.

La dose létale 50 (DL_{50}) est la source d'information sur la toxicité de toute substance chimique, y compris les pesticides. C'est la dose d'une substance administrée (par voie orale, cutanée, intraveineuse ou intramusculaire) en une seule fois, qui cause la mort de 50% d'un groupe d'animaux d'essai. On utilise le plus souvent des rats ou des souris. Cette dose létale, outre le fait qu'elle nous permet de comparer la puissance toxique de différents pesticides, elle nous aidera aussi à prendre diverses mesures nécessaires comme l'établissement des limites d'exposition, l'utilisation des vêtements et d'équipement de protection appropriés, l'élaboration des mesures d'urgence en cas d'accidents,…etc.

Tableau IG.1 : Informations toxicologiques et écotoxicologiques requises par la réglementation REACH [selon le règlement CE N°1907/2006 UE, 2006] (Barjhoux, 2011).

	Tonnage annuel (tonnes)			
	1 à 10	**10 à 100**	**100 à 1000**	**>1000**
Test de toxicité aigüe	oui	--	--	--
Test d'irritation cutanée et oculaire	oui	--	--	--
Etude de toxicité par administration répétée	--	Oui	--	--
Test de reprotoxicité	--	oui	--	--
Etude de toxicité subchronique	--	--	oui	--
Etude de toxicité à long terme	--	--	--	oui
Etude de carcinogénicité	--	--	--	oui

Notons que la détermination expérimentale de la dose létale est une opération qui se heurte à trois inconvénients majeurs. Le premier a trait à la difficulté de réalisation des expériences. Ce sont des manipulations qui exigent beaucoup de soin et une multitude de conditions opératoires à satisfaire (sélection des espèces animales, conditions d'hébergement et d'alimentation des animaux d'essai, préparation des animaux et des doses, suivi et observation des animaux après l'administration des doses,…etc.). Le second inconvénient est que ces tests nécessitent beaucoup de temps, beaucoup d'animaux, donc d'importants

coûts.En effet, les tests prennent au minimum 15 jours et avec la multiplication des expériences, on doit patienter encore plus pour aboutir à des résultats fiables. Le troisième inconvénient est un problème d'éthique. En effet, des dizaines d'associations ayant vu le jour aux Etats Unis et en Europe, dénoncent l'horreur de l'expérimentation animale et militent pour son abolition. A titre d'illustration, le nombre total d'animaux utilisés à des fins expérimentales en 2008 dans les états membres de l'Union Européenne dépasse 12 millions. Les souris et les rats sont de loin les animaux les plus utilisés. Par ailleurs, 8.7% de ces animaux ont été utilisés pour les essais toxicologiques (figure IG.1a et IG.1b).

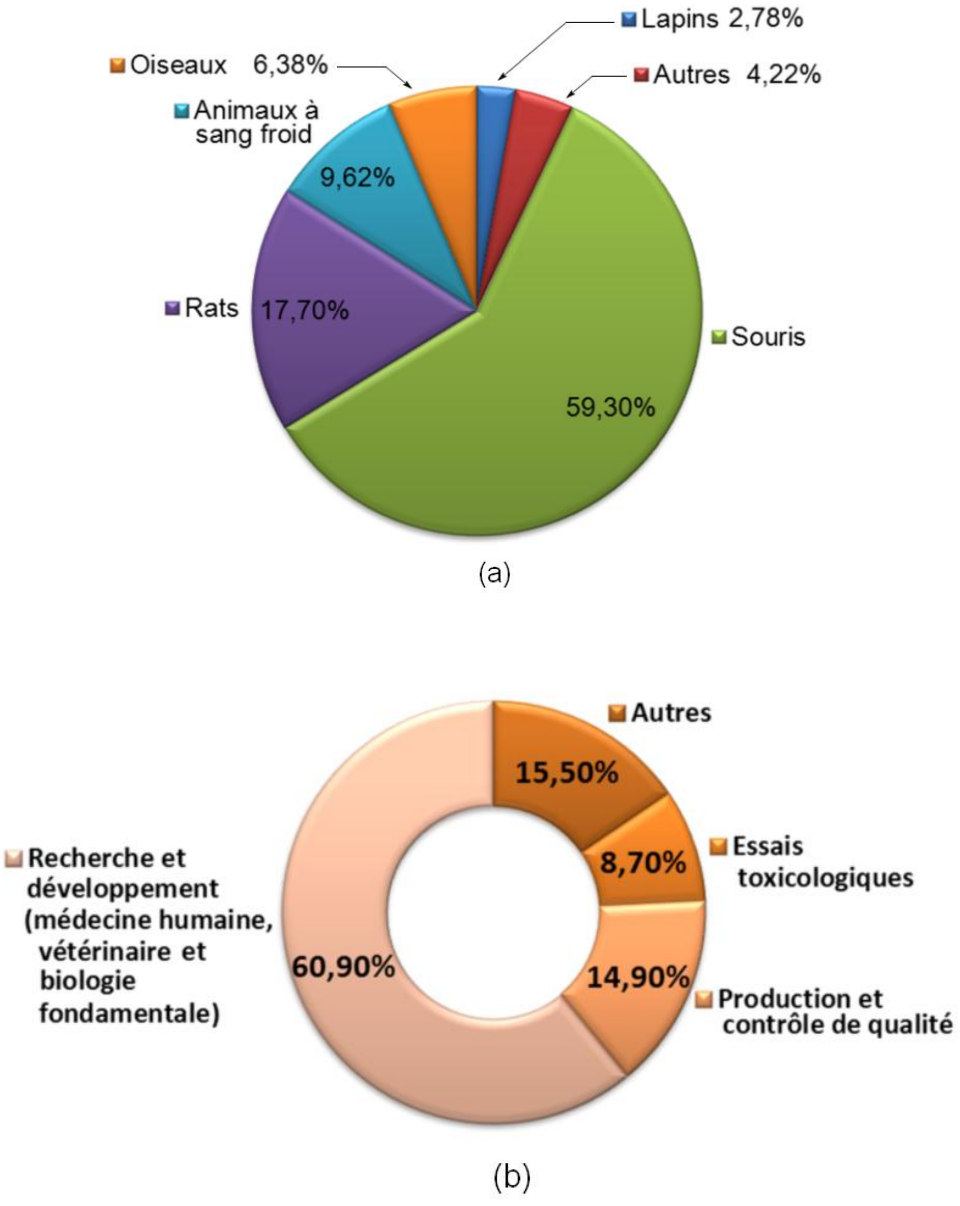


Figure IG.1 : Pourcentage d'animaux utilisés par catégorie en Europe en 2008 (a) et objet des expériences (b) (COM 2010)

Les associations de protection des animaux prônent l'utilisation de méthodes substitutives. Ces méthodes permettraient d'éviter la souffrance et d'épargner la vie de millions de chats, chiens, souris, singes, oiseaux, ovins, caprins, poissons utilisés chaque année dans les laboratoires.Les efforts de toutes ces associations ne sont pas restés vains puisque des méthodes alternatives plus fiables, moins coûteuses et aux résultats plus rapides ne cessent de se développer.Elles présentent aussi l'avantage d'être éthiques et de bénéficier de technologies à la pointe du progrès. On citera à titre d'exemple les cultures de cellules, la biologie moléculaire et la chémoinformatique.

Cette thèse se veut justement une contribution à l'utilisation d'une de ces méthodes alternatives. On s'est assigné comme objectif, l'élaboration d'un modèle qui nous permettra de prédire la toxicité orale aigüe des pesticides au moyen de la chémoinformatique.Cette dernière est une discipline scientifique qui consiste à étudier et résoudre des problèmes relatifs à la chimie en appliquant des méthodes informatiques. Nous nous intéresserons aux modèles "in silico", appelés modèles Quantitative Structure-Activity Relationships (*QSARs*), mettant en relation la structure d'une molécule avec son activité biologique (p.ex. : la toxicité). Ces modèles quantifieront cette relation et mettront en exergue l'impact des propriétés de la molécule sur sa toxicité. En se basant sur le modèle obtenu, on pourra faire la prédiction de la toxicité de tout nouveau pesticide en un temps record et sans recourir à l'expérimentation animale.

Pour atteindre l'objectif assigné à cette thèse, quatre chapitres ont été développés. La présente introduction générale de ce mémoire a été l'occasion de situer le contexte et les motivations de cette recherche. Il a été question de la problématique de la toxicité des pesticides, de celle des expérimentations animales et de la solution envisagée.

Dans le premier chapitre, nous proposerons une description synthétique du contexte bibliographique dans lequel s'inscrit ce travail de thèse, en passant en revue dans un premier temps tous les travaux publiés quant à la contamination de notre environnement par les pesticides. Puis dans une seconde partie,une présentation de l'abondante littérature consacrée à l'impact de ces substances sur tous les compartiments de l'environnement sera faite. Ensuite, on clôturera ce chapitre par une rétrospective bibliographique des divers modèles *QSAR* élaborés en vue de la modélisation de la toxicité.

Le second chapitre est une présentation des modèles *QSAR*. On y parlera des éléments nécessaires à l'élaboration de ces modèles, en particulier la base de données, les descripteurs moléculaires ainsi que la sélection des plus pertinents. Dans la seconde partie de ce chapitre, on énoncera les principes et les notions associés aux réseaux de neurones artificiels (*RNA*).

Le sujet de cette thèse a trait aux pesticides. A cet effet, le troisième chapitre leur a été consacré. On y parlera des avantages des pesticides, mais aussi de leurs inconvénients. On mettra en exergue les différentes formes de toxicité de ces substances et on terminera ce chapitre par un paragraphe consacré aux pesticides en Algérie.

Le quatrième chapitre est consacré, quant à lui, à l'élaboration de quatre modèles *QSAR* : un pour les herbicides, un second pour les insecticides, un troisième pour les fongicides et le quatrième est réservé aux pesticides de façon générale. Pour chaque modèle, on discutera du volet matériels et méthodes, de la procédure de développement des modèles et des outils statistiques permettant d'analyser leur performance.

Enfin, la conclusion générale aura pour objet de synthétiser l'ensemble des résultats de ce travail et proposera des perspectives de recherche dans la continuité de cette thèse.

Chapitre 1

REVUE CRITIQUE DE LA LITTERATURE SUR L'ACTIVITE TOXICOLOGIQUE

Sommaire

Introduction 8
1.1 Pesticides et pollution de l'environnement 9
1.2 Impact sanitaires avérés ou suspectes des pesticides 16
1.3 Toxicité et modèles de prediction QSAR 22
1.4 Modèles QSAR et prédiction de la toxicité aigüe des pesticides chez les rats 24
Conclusion 27

Ce chapitre consacré à une revue bibliographique a été scindé en quatre parties. Dans la première partie, on passera en revue la littérature scientifique inhérente à la contamination généralisée de l'environnement par les pesticides. Par cette démarche, il s'agit de rappeler les limites et les dangers de ces substances pour l'environnement, les écosystèmes mais également pour les êtres humains. Les conséquences néfastes et nocives de cette pollution, relatées par une multitude de publications scientifiques, seront abordées dans la seconde partie. La partie suivante, traitera quant à elle, des modèles QSAR de prédiction de la toxicité ayant fait l'objet d'articles scientifiques. Elle mettra en exergue l'importance et l'utilité de cette méthode. La quatrième et dernière partie de cette revue bibliographique a été consacrée à la littérature scientifique parue dans le cas des modèles QSAR de prédiction de la toxicité aigüe (Dose Létale 50 : DL_{50}) des pesticides. Une conclusion clôturera ce chapitre et sera une synthèse par laquelle on justifiera le choix du thème de cette thèse.

INTRODUCTION

Depuis très longtemps, le souci principal de l'humanité était de garantir la sécurité alimentaire compte tenu de la croissance démographique galopante. Pour ce faire, il était impératif de prendre toutes les mesures visant à accroître la production agricole. Cependant, cet objectif ne pouvait être atteint sans une lutte contre tous les organismes responsables des dégâts sur les cultures. Cette lutte a été menée moyennant l'emploi de substances appelées pesticides ou produits phytosanitaires. De succès en succès, le champ d'application de ces substances s'est élargi à d'autres domaines autres que l'agriculture. Des pesticides diversifiés furent produits et ainsi la consommation de pesticides à travers le monde a augmenté de façon spectaculaire de 1960 à 2011 (Figure 1.1).

Environ cinq millions de tonnes de pesticides sont utilisées chaque année dans le monde. Cependant, les utilisations abusives de pesticides deviennent de plus en plus graves, ce qui a entraîné la pollution de l'environnement dans sa totalité. Une partie importante des pesticides appliqués est libérée dans les sols non ciblés, les plans d'eau, dans l'atmosphère et est absorbée par presque tous les organismes vivants. Les pesticides sont également présents dans nos aliments: près de 50% des fruits et des légumes produits par l'agriculture intensive en contiennent. Ils finissent dans nos organismes, apportés là par l'eau, l'air et les aliments consommés. Ces pesticides posent un véritable problème de santé publique, et pas seulement pour les utilisateurs qui sont les plus exposés, mais aussi pour la population générale.

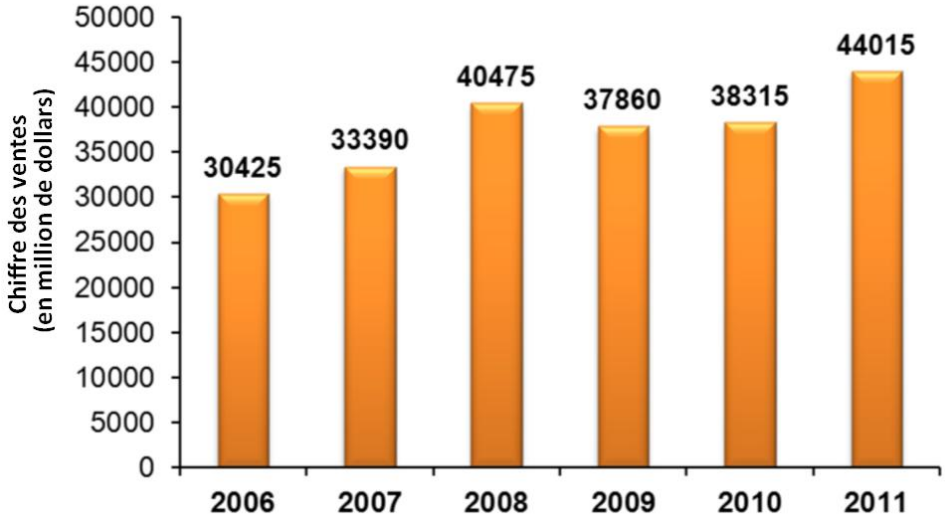


Figure 1.1 : Evolution du chiffre des ventes mondiales entre les années 2006 et 2011 (UIPP, 2012)

1.1 PESTICIDES ET POLLUTION DE L'ENVIRONNEMENT

En fonction des conditions d'utilisation et selon les caractéristiques du milieu, les pesticides sont susceptibles de se retrouver dans les différents compartiments de l'environnement ainsi que dans les denrées alimentaires. Depuis près de cinquante ans, les pesticides ont été mis en évidence 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 (les œufs, le lait, la viande, le poisson...). Ils existent sous leur forme initiale, mais ils peuvent aussi être dégradés, on parle alors de résidus ou de métabolites.

Le potentiel d'un pesticide à se déplacer dépend de ses propriétés chimiques (ionisation, solubilité dans l'eau, volatilité, persistance dans le milieu,…etc.), de sa formulation, des propriétés du sol (taux d'humidité, pH, pourcentage de matière organique, …etc.), du taux et de la méthode d'application, des conditions météorologiques (fréquence et répartition des pluies) et de la profondeur de la nappe (Winter, 1992). D'autres procédés qui influent sur le devenir du produit chimique comprennent l'absorption de la plante, l'adhérence du sol, la lixiviation, et la vaporisation.

De nombreux travaux réalisés au niveau international mettent en évidence la pollution de l'environnement par les pesticides. La revue de la littérature dans ce domaine a porté sur la pollution des divers compartiments environnementaux. On a donc consulté successivement les articles se rapportant à la pollution de l'air, des milieux aquatiques, des sols et de l'alimentation. La base de données consultée est PubMed. A la lecture des résultats de cette recherche, il ressort que 2182 articles scientifiques ont été publiés de 1968 au 18 avril 2014 (date de la consultation du site). Sur la figure 1.2, est reporté le nombre de publications de 2003 à 2013. Pour notre part, on a retenu les publications inhérentes à la pollution de l'environnement un peu partout dans le monde d'une part et celles qui sont les plus récentes d'autre part.

1.1.1 Contamination de l'air

La présence de pesticides est observée dans toutes les phases atmosphériques en concentrations variables dans le temps (avec parfois un caractère saisonnier, en lien avec les périodes d'application) et dans l'espace (selon la proximité des sources). Cette contamination est chronique. La pollution atmosphérique générale par les pesticides est très répandue. Les pesticides dans l'atmosphère sont principalement issus des émissions des usines de pesticides,

de l'évaporation des résidus de pesticides dans les sols et les cours d'eau, de la volatilisation des pesticides pulvérisés, … etc.

Des études émanant de divers groupes de recherche (Briand, 2003; Rousseau *et al*., 2004; Scheyer *et al*., 2005) ont réalisé des mesures de concentrations en pesticides dans l'atmosphère. L'ensemble de ces travaux a révélé la présence de pesticides dans toutes les phases atmosphériques, qu'elles soient gazeuse, liquide ou particulaire dans les aérosols, les gouttelettes de brouillard ou la pluie (Bedos *et al*., 2002). Dans une autre étude, des chercheurs ont pu observer dans les eaux de pluie au Danemark un certain nombre de composés interdits dans ce pays mais autorisés dans d'autres pays européens, indiquant ainsi une contribution significative du transport atmosphérique dans la contamination locale (Alix et al., 2005). Une substance active (le fongicide Epoxiconazole) utilisée sur des plantes a été retrouvée dans l'air (Coscollà *et al*., 2010) et ce malgré sa faible volatilisation. Il en est de même pour l'insecticide chlorpyriphos, qui lui aussi, a été retrouvé dans l'atmosphère (Yao *et al*., 2008 ; Zhou *et al*., 2010).

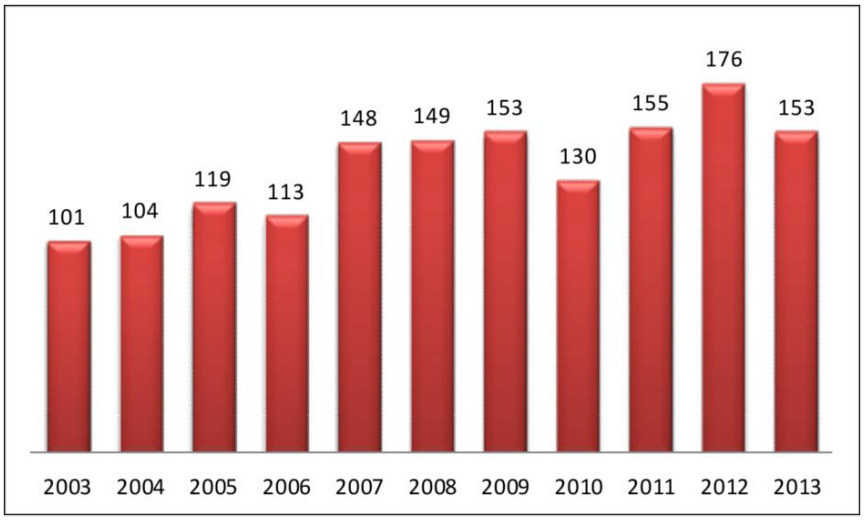


Figure 1.2 : Nombre de publications traitant la contamination de l'environnement par les pesticides (PubMed consulté le 18/04/2014)

Une récente étude (Gunier et al., 2011) montre que les pesticides agricoles utilisés dans un rayon de 1250 mètres autour d'habitations finissent par contaminer l'intérieur de ces maisons. Cette étude montre que l'utilisation de pesticides agricoles, dont certains sont soupçonnés d'être cancérigènes, a un impact sur la contamination des habitations avoisinantes, et donc de l'air que respirent chaque jour leurs habitants. Par ailleurs, l'analyse durant une année complète de l'air de Yangtze (Chine) a révélé la présence de concentrations élevées de pesticides organochlorés. Les auteurs affirment que cette pollution de l'air est

susceptible d'être à l'origine de développement de cancer par inhalation et ce à la suite des calculs du facteur de risque effectués (Zhang et al., 2013). Dans une autre étude, 6 pesticides ont été analysés dans l'atmosphère du nord de l'Algérie par Moussaoui et al. (2012). Les pesticides trouvés dans la phase gazeuse urbaine ont montré des niveaux élevés par rapport à l'échantillonnage d'une une zone rurale. Le malathion et le chloropyrifos étaient présents à des concentrations élevées.

1.1.2 Contamination de l'alimentation

Les pesticides continuent à être utilisés dans la production de denrées alimentaires telles que fruits, légumes, et céréales. Ce sont dès lors des substances dangereuses pour les êtres vivants et donc pour les êtres humains. La pollution des cultures est causée par la pulvérisation des pesticides, le traitement des semences et le traitement des sols avec des pesticides. Des résidus de pesticide sont retrouvés dans les produits végétaux comme les grains, les légumes, les fruits, le thé et les herbes médicinales.

Dans une étude sur la contamination des cultures sur des terres contaminées par le Chlordane, il a été remarqué un transfert de ce pesticide vers les cultures, c'est-à-dire que tous les tubercules des différentes cultures (patate douce, navet, radis, courgette et tomate) ont été contaminés (Cabidoche et Lesueur-Jannoyer, 2012). Nougadère et al., (2012) ont quant à eux recherché la présence de résidus de trois cent vingt-cinq pesticides et de leurs produits de transformation dans des échantillons alimentaires couvrant 90 % du régime alimentaire d'une population. Les résultats ont montré que 37 % des échantillons contenaient un ou plusieurs résidus. Soixante-treize pesticides ont été détectés et quantifiés. Les pesticides les plus fréquemment détectés étaient les insecticides pirimiphos -méthyl et le chlorpyrifos –méthyl. Le pesticide dimethoate et son métabolite ont été détectés dans deux échantillons de cerises à des niveaux supérieurs à la dose journalière admise.

Une étude sur les résidus de pesticides phosphorylés dans l'alimentation totale du Koweït a été entreprise par Talat et al., (2005). Les résultats ont indiqué que 18 % des échantillons contenait des résidus. Le monocrotophos (0,2 mg/kg), le diazinon (0,05 mg/kg), le quinalphos (0,022 mg/kg) le Chlorpyriphos-méthyl (0,01 à 0,33 mg/kg) et le fénitrothion (0,16 à 0,84 mg/kg) étaient les pesticides les plus détectés. L'utilisation des pesticides dans la production maraîchère a fait l'objet d'une étude réalisée sur 108 maraîchers dans la ville rurale de Tori-Bossito dans le sud du Bénin (Ahouangninou et al. 2012). Deux indices de risque ont été calculés pour chaque pesticide: un indice de risque pour l'environnement (*IRE*)

et un indice de risque pour la santé (*IRS*). Les résidus de pesticides ont été trouvés dans 42% des échantillons de feuilles d'aubergine, de concombre, d'amarante et de solanum. Au Ghana, une étude a été menée afin d'évaluer les résidus de pesticides organochlorés, organophosphorés et les résidus de pesticides pyréthroïdes synthétiques dans les fruits et légumes mis en vente dans les marchés (Bempah et al., 2012). 9,8% des 309 échantillons de fruits et légumes présentaient des taux de résidus supérieurs à la limite admise.

Huit pesticides différents ont été détectés lors de l'analyse des résidus de pesticides dans le blé importé par l'Afrique du Sud. Les pesticides les plus fréquemment détectés étaient le mercaptothion, la perméthrine et le chlorpyrifos. Les auteurs soulignent que les aliments à base de ce blé pourraient être une source de contamination aussi bien pour les humains que pour les animaux (Dalvie et London, 2009). Par ailleurs, les pesticides diazinon, chlorpyrifos et quinalphos ont été retrouvés dans des analyses d'une série de fruits (Sanagi et al., 2013).

D'autres produits alimentaires autres que les fruits, les légumes et les céréales ont fait l'objet d'analyses. C'est ainsi que des résidus de pesticides organochlorés et pyréthrinoïdes ont été retrouvés dans des échantillons de thé lors d'une analyse moyennant l'utilisation d'une nouvelle technique chromatographique, sensible et efficace (Dan Liu et Shungeng Min, 2012). La contamination du lait et des produits laitiers a fait l'objet d'une publication scientifique (Fischer et al., 2011). Cet article donne un aperçu général sur la nature, les sources, l'apparition, la détection et le risque potentiel de santé humaine des principaux contaminants chimiques de cet aliment. Des traces de pesticides (DDT, lindane, dieldrine,… etc.) y ont été effectivement retrouvés.

Farajzadeh et al. (2014) viennent de mettre en évidence une méthode chromatographique qui leur a permis de constater la contamination des huiles végétales par les pesticides suivants : fenpropathrine, Sumithrin, cyhalothrine, perméthrine et deltaméthrine. On peut aussi noter l'existence d'une étude dont les résultats ont montré que les concentrations des pesticides organochlorés sont plus élevées dans le miel issu des pays en développement que dans le miel produits dans les pays développés (Wang et al., 2010).

Davodi et al. (2011) ont déterminé les concentrations de pesticides dans 8 espèces de poissons prélevés des marais Shadegan en Iran. Dans tous les échantillons, les concentrations étaient plus élevées que les normes directrices pour la sécurité alimentaire délivrée par l'Union européenne (UE) et la Food and Drug Administration (Etats unis). D'autres chercheurs iraniens (Arzi et al., 2011) ont calculé les concentrations d'aldrine, de dieldrine,

d'heptachlore, d'heptachlore époxyde et de méthoxychlore dans des poissons pêchés dans la province de Khozestan en Iran. Tous les poissons examinés ont été contaminés par les 14 pesticides organochlorés étudiés. Des concentrations élevées ont été trouvées pour certains pesticides. Une enquête a été menée pour évaluer les résidus de pesticides dans les échantillons de poissons de la rivière Densu au Ghana (Fianko et al., 2011). Les données obtenues indiquent que les taux du γ -HCH, d'heptachlore, d'α - endosulfan, d'endosulfan sulfate et de dieldrine ont dépassé la dose de référence, indiquant par là un grand potentiel pour la toxicité systémique des consommateurs. Akan et al. (2013) ont utilisé quatre espèces de poissons de Borno, un des états du Nigéria pour une analyse de résidus. Onze pesticides organochlorés ont été détectés dans tous les échantillons de poissons étudiés. L'endosulfan fut le pesticide le plus abondant dans les tissus étudiés de toutes les espèces de poissons. Cette étude a également révélé que tous les résidus de pesticides dans les échantillons de poissons étudiés étaient au-dessus des limites maximales admises.

1.1.3 Contamination des milieux aquatiques

Cette partie constitue un état des connaissances vis à vis des niveaux de pesticides relevés dans les milieux aquatiques. La contamination des eaux par les pesticides est devenue de plus en plus courante. Compte tenu des risques qu'ils représentent, la présence de pesticides dans les cours d'eau, dans les eaux côtières et dans les eaux souterraines fait l'objet d'un suivi régulier qui n'a cessé de se renforcer lors de la dernière décennie. La concentration de pesticides dans les cours d'eau a été rapportée par plusieurs auteurs. Plusieurs travaux de recherche ont montré qu'en de nombreuses régions du monde, les hydrosystèmes présentent des contaminations significatives par des produits phytosanitaires. Les niveaux de pollution de l'eau par les pesticides sont différents selon qu'il s'agit des eaux des terres cultivées, des eaux de ruissellement, des eaux souterraines, des eaux de rivière, des eaux souterraines profondes et des eaux de mer.

A titre d'exemple, les petites rivières côtières (Louros et l'Arachthos) en Grèce présentent occasionnellement des teneurs en herbicides particulièrement importantes (Munaron, 2004). Steen et al. (2001) ont détecté sur la rivière Scheldt, des pics de résidus de certains pesticides dépassant de loin la norme requise. Dans des compartiments aquatiques, des traces de chlorpyriphos ont aussi été détectés (Coupe et Blomquist, 2004). En Pologne, des études indiquent une contamination des précipitations par des pesticides (Polkowska et al., 2000; Grynkiewicz et al., 2001). Dans des études sur la contamination des eaux de la

région Aquitaine en France conduite par (Barjhoux, 2011), il a été remarqué que 2.6% des contaminants sont des pesticides.

Des échantillons d'eau de surface ont été prélevés dans le Mississippi pour évaluer les niveaux de pesticides dans deux études distinctes. Des traces de pesticides comme le métolachlore et l'acétochlor et certains de leurs métabolites ont été rapportés par Rebich et al. (2004). De leur côté, Target et al. (2014) ont décelé l'hexazinone dans 94 % des échantillons prélevés, suivie par le métolachlore (76%), le tébuthiuron (48%), l'atrazine (47%) et la Métribuzine (6 %). Des ressources hydriques italiennes sont contaminées par le terbuthylazine, pesticide ayant remplacé l'atrazine, et son métabolite (déséthyl terbuthylazine). Ceci a été rapporté par Bottoni et al. (2013) dans une étude sur ce pesticide potentiellement toxique.

En Guadeloupe, les résidus de pesticides organochlorés ont été détectés dans des concentrations très élevées dans les écosystèmes d'eau douce. Ces concentrations dépassent largement la limite résiduelle autorisée du Chlordecone dans les poissons et les crevettes (Coat et al, 2011). Par ailleurs, la majorité des fleuves et sources d'eau potable en Inde est contaminée par les pesticides (Agrawal et al., 2010). Au Pakistan, un certain nombre d'échantillons de poissons et de coquillages ont été analysés pour la détermination de la contamination des pesticides en raison de la pollution de l'environnement marin. Pour les auteurs, les résultats de cette étude ne semblent pas être très alarmants par comparaison aux niveaux de traces trouvées dans les autres parties du monde (Hina et al., 2013).

Une analyse des eaux du réservoir Guanting (Chine) a révélé la présence de pas moins de 18 types de pesticides organochlorés (Wan et al., 2009). En chine toujours, un total de 27 échantillons d'eaux souterraines peu profondes ont été collectées à partir de la région du lac Taihu. Le DDT et l'hexachlorocyclohexanes (HCH) sont les contaminants les plus prépondérants dans ces eaux. Dans une très récente étude portant sur la contamination des eaux souterraines peu profondes, Chunfa et al. (2014) déduit que les valeurs calculées du risque cancérogène des contaminants soulèvent un risque de cancer potentiellement grave pour ceux qui consomment l'eau potable provenant de ces eaux.

1.1.4 Contamination des sols

Selon les différentes études menées un peu partout dans le monde, il a été établi que divers types de sols, y compris les champs cultivés, les champs de légumes et les terres forestières sont contaminées par différents pesticides. La contamination des sols à distance,

c'est-à- dire surtout par la voie aérienne ou éventuellement lors de crues, a parfois été décrite. Les exemples les mieux documentés concernent les pesticides organochlorés (Galiulin *et al.*, 2002). A titre d'exemple, dans un suivi réalisé en 2002, Mast et al., (2003) ont estimé que les apports annuels au niveau du sol atteignaient 45,8 mg/ha pour l'atrazine, 14,2 mg/ha pour le dacthal et 54,8 mg/ha pour le carbaryl.

L'effet de cette contamination est la diminution de la fertilité des sols suite à la destruction des microorganismes et des vers de terre. Dans une étude sur les indicateurs de la contamination des sols par les pesticides, Carine et al. (2011) estiment que les pesticides peuvent affecter indirectement les activités enzymatiques du sol grâce à leur action sur les microorganismes du sol. Par ailleurs, la contamination par des pesticides organochlorés de deux sols situés respectivement à proximité d'une ancienne et nouvelle usine de pesticides a fait l'objet d'une étude en chine (Lifei et al., 2009). Cette étude a révélé que ces sols étaient contaminés différemment par les substances suivantes : HCH, DDT, HCB et Chlordane. Selon Zhao et al. (2009), différents types de sol de la région de Haihe River (Chine) sont contaminés par les pesticides DDT et HCH. Les mêmes pesticides ont été retrouvés également dans les sols de Pearl River Delta en chine (Ma et Ran, 2009).

La dégradation des pesticides selon la nature du sol a fait l'objet d'études. A titre d'exemple, dans une étude portant sur la nature des sols qui subissent le plus de contamination par le pesticide chlorpyriphos, Lian et al. (2013) concluent que la dégradation de cette substance est plus lente pour les sols acides, les sols à forte argile et les sols à basse température. Une étude menée par Oukali-Haouchine et al. (2013) a montré que la Métribuzine est effectivement adsorbée par les sols limono- argileux algériens. Toutefois, les quantités adsorbées restent faibles. D'après les auteurs, environ 3/4 de la Métribuzine introduit ne sont pas conservés par le sol et pourraient être transférés vers les eaux souterraines, ce qui pourrait constituer un risque important de contamination des eaux souterraines.

1.1.5 Contamination des travailleurs agricoles et non agricoles

Ce paragraphe de la revue bibliographique met en exergue les effets des expositions, directe et indirecte, aux pesticides et relate quelques données se rapportant à la contamination par ces derniers. Aussi, dans une étude sur les risques d'exposition des exploitants agricoles aux pesticides, Baldi et al. (2012) ont relevé des contaminations cutanées et respiratoires. Selon leur étude, les phases de préparation et d'application des pesticides sont respectivement

responsables pour 30 et 50% de la contamination. Dupupet et al. (2010) ont étudié l'effet de l'exposition des ouvriers agricoles aux fongicides (dithiocarbamates). L'objectif de cette étude était de mesurer le taux d'éthylénethiourée (ETU), métabolite de ces fongicides, dans les urines. Aussi, il a été observé une augmentation du taux d'ETU après application de ces pesticides. Les formulations de ces fongicides étant en poudre, ils ont conclu à une contamination par voie respiratoire. Une autre étude, portant sur la contamination de personnes (jardiniers, fleuristes et vétérinaires) non exposées professionnellement aux pesticides (Bouvier et al., 2006), a relevé une présence de pesticides sur les mains de ces personnes plus importante que celles ayant contaminé l'air du milieu de leur travail.

Une enquête épidémiologique sur vingt-neuf parmi cinquante décès a été entreprise dans une communauté rurale de Bignona au Sénégal (Touré et al., 2011). Une enquête qualitative a été réalisée sur l'inventaire des différents pesticides utilisés dans les sites d'étude. L'auteur signale que les décès se produisent en septembre – octobre, période d'une grande utilisation de pesticides. L'hypothèse d'empoisonnement par les pesticides (organophosphorés) soit par inhalation ou par une contamination accidentelle de la nourriture a été émise.

1.2 IMPACT SANITAIRES AVERES OU SUSPECTES DES PESTICIDES

Bien que les pesticides aient largement fait bénéficier l'humanité à travers la valorisation des produits agricoles et le contrôle des maladies infectieuses, leur utilisation intensive, à son tour, menace la santé humaine et les composantes environnementales. Le contact à long terme avec les pesticides peut nuire à la vie humaine et peut perturber le fonctionnement de différents organes dans le corps, y compris les systèmes nerveux, endocrinien, immunitaire, reproducteur, rénal, cardio-vasculaire et les systèmes respiratoires. À cet égard, il existe des preuves sur le lien entre l'exposition aux pesticides et l'incidence de maladies chroniques humaines, comme le cancer, la maladie de Parkinson, la maladie d'Alzheimer, le diabète, le vieillissement, les maladies cardiovasculaires et la maladie rénale chronique (Mostafalou et Abdollahi, 2013).

Cette relation significative entre les expositions aux pesticides et certaines pathologies chroniques a fait l'objet de plusieurs publications scientifiques. En effet, on a dénombré pas moins de 22074 articles entre l'année 1980 et le 18 avril 2014 (Date de consultation de la base

de données). A titre d'illustration, est représenté sur la figure 1.3 ci-dessous le nombre de publications scientifiques de l'année 2003 à 2013.

Les pesticides peuvent présenter, en plus de leurs effets intentionnels sur les parasites ou organismes visés, des dangers très variables pour l'homme et les écosystèmes, avec un impact immédiat ou sur le long terme. Des études ont montré que toutes les personnes, notamment les enfants, les femmes enceintes, les agriculteurs, les ouvriers agricoles et les personnes âgées, peuvent ressentir des effets négatifs sur la santé à cause de l'exposition aux pesticides. L'exposition aux pesticides peut causer une intoxication aiguë, le cancer, des dommages neurologiques, des anomalies congénitales,…etc. Beaucoup de preuves ont révélé que de nombreux pesticides couramment utilisés peuvent supprimer la réponse normale du système immunitaire humain, ce qui rend le corps plus vulnérable à l'invasion des virus, des bactéries et des tumeurs (Levin, 2007).

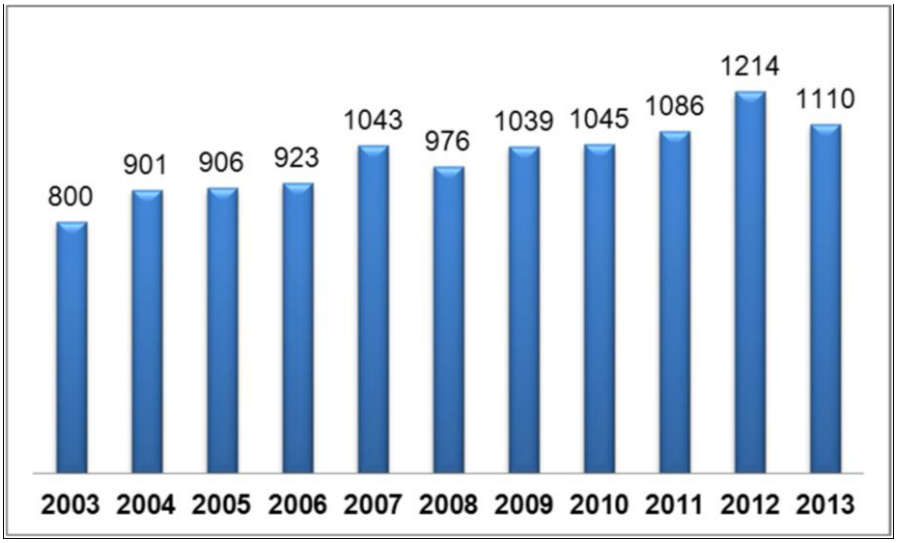


Figure 1.3 : Nombre de publications inhérentes aux effets des pesticides sur la santé (PubMed consulté le 18/04/2014).

1.2.1 Pesticides et troubles de la reproduction

Pour les troubles de la reproduction, des études ont suggéré la possibilité d'un lien entre l'exposition aux pesticides et les risques de stérilité masculine, d'excès d'avortement spontanés, de prématurés, de mort-nés et de certaines malformations fœtales (AIRPARIF, 2007). Des travaux de recherche ont mis en évidence que les fongicides conazoles (par exemple l'Epoxiconazole) agissent comme des perturbateurs endocriniens (Kjaerstad *et al*., 2010). Aussi, il a été observé chez le rat une perturbation du développement reproductif

comme l'allongement de la durée de gestation ou la mort du fœtus (Christiansen *et al*., 2010) et chez la caille, une réduction du nombre de spermatides (Gröte *et al*., 2008).

Une vaste étendue de la littérature a détaillé les effets néfastes des expositions environnementales, notamment aux pesticides, et ce sur les deux systèmes reproducteurs mâles et femelles (Shojaei Saadi et Abdollahi, 2012). A la suite d'une étude sur l'évaluation des troubles de la reproduction chez les travailleurs masculins et féminins exerçant dans des serres, Bretveld et al. (2008) ont affirmé détenir des preuves quant à l'hypothèse que l'exposition aux pesticides affecte la reproduction humaine conduisant ainsi à un avortement spontané et peut être à une grossesse prolongée. La population en général peut être concernée par ces problèmes de fertilité due à une exposition aux pesticides. Ainsi, une autre étude (Orton et al., 2011) révèle que des pesticides auxquels la population européenne est exposée sont en fait des anti-androgènes et menaceraient potentiellement la fertilité masculine.

Une étude menée par Andersen et al. (2008) montre que les garçons dont les mères ont travaillé dans des serres où on a utilisé des pesticides pendant leur grossesse ont un développement des fonctions reproductrices perturbé. L'étude a porté sur le développement des fonctions reproductrices de 110 garçons. Les résultats de l'étude attestent de la prévalence du cryptorchidisme (testicules non descendus), d'un volume testiculaire et d'une concentration sérique en testostérone inférieurs chez les garçons des mères exposées aux pesticides dans les serres que chez les garçons des mères non exposées. Pour les auteurs, ces résultats suggèrent un effet négatif de l'usage professionnel des pesticides par les mères pendant la grossesse sur le développement des fonctions de reproduction chez leurs garçons, malgré les précautions d'utilisation prises.

1.2.2 Malformations

Une étude récente (Chevrier et al., 2011) renforce l'hypothèse selon laquelle l'exposition environnementale des femmes enceintes au pesticide atrazine augmente le risque d'effets indésirables sur le fœtus. Les scientifiques ont suivi une cohorte de femmes enceintes de 2002 à 2006 en Bretagne. Les auteurs ont noté des risques accrus de faible poids et de faible circonférence crânienne à la naissance pour les bébés des mamans exposées à l'atrazine par leur environnement

1.2.3 Affection du système immunitaire

Chez les mammifères, il a été observé que les insecticides affectent le système immunitaire (Nandi et al., 2011).

1.2.4 Effet sur les organismes aquatiques

Les organismes aquatiques se révèlent particulièrement sensibles à l'action des pesticides. En effet, des études ont montré que le chlorpyriphos (insecticide) est toxique vis-à-vis du zooplancton dulçaquicole (Pablo et al., 2008 ; Colville et al., 2008). Il en est de même avec des poissons tels qu'Oreochomis *niloticus* (Oruç, 2010)

1.2.5 Pesticides et cancer

BALDI et ses collaborateurs, dans des travaux datant de 1998, ont fait le point sur l'état des connaissances épidémiologiques sur les effets des pesticides sur la santé en retraçant l'évolution des connaissances et des hypothèses qui concourent à établir un lien entre pesticides et cancer. Selon ces auteurs, les premiers travaux sont issus du constat d'une différence de mortalité par cancer entre les agriculteurs (personnes exposées directement aux pesticides) et les autres catégories professionnelles et ce pour un certain nombre de localisations tumorales. Par ailleurs, dans une étude comparative sur l'apparition de cancer chez les populations agricoles et non agricoles, il a été constaté que des cancers bien spécifiques (leucémies, myélomes et lymphomes) apparaissent plus fréquemment chez les premières citées.

Les pesticides DDT, lindane, simazine et le risque de survenue du cancer de la prostate a été évoqué lors d'une étude sur l'exposition des agriculteurs de la Colombie britannique à ces substances (Band et al., 2011).A la suite des études sur l'impact des pesticides sur la santé des ouvriers agricoles, Mills et Yang (2007) sont arrivés à une conclusion similaire à celle d'autres études sur le cancer gastrique menées en Europe. Ils suggèrent que le risque élevé de cancer de l'estomac vécu par les travailleurs agricoles peut être attribué à l'exposition en milieu de travail. Chrisman et al. (2009) ont évalué statistiquement le degré de corrélation entre les ventes de pesticides en 1985 et le taux de mortalité par cancer entre 1996-1998 et ce dans 11 états du Brésil. Les résultats suggèrent une corrélation significative avec un coefficient variant de 0.61 (cancer de l'œsophage) à 0.73 (cancer de la lèvre). En outre, ces auteurs suggèrent que l'exposition de la population aux pesticides dans les années 1980 dans certains États du Brésil peut être associée à l'apparition de cancers une décennie plus tard.

Provost et al. (2007) ont montré un risque accru de développer certains cancers du cerveau pour les personnes exposées aux pesticides par leur activité professionnelle ou à la maison. Selon cette étude, l'augmentation du risque est statistiquement significative pour de forts niveaux d'exposition aux pesticides. Pour les agriculteurs exposés aux niveaux les plus

élevés, le risque est ainsi plus que doublé. Les femmes travaillant à la ferme sont plus susceptibles de développer un cancer du sein que les autres, selon une étude Canadienne (Brophy et al., 2002). Les chercheurs qui se sont penchés sur les antécédents professionnels de 564 femmes atteintes du cancer du sein, dans la région de Windsor (centre du Canada), ont noté que le risque était multiplié par 2,8 pour celles qui ont travaillé dans une ferme à un moment de leur vie.

Le risque de leucémie chez l'enfant est associé à l'exposition de la mère aux pesticides pendant la grossesse (Wigle et al., 2009). Les chercheurs ont étudié les résultats de 31 études épidémiologiques publiées entre 1950 et 2009 étudiant le lien entre les leucémies chez l'enfant et l'exposition des parents aux pesticides. Le résultat stipule que le risque de leucémie était doublé chez les enfants dont les mères ont été exposées professionnellement à des pesticides pendant la grossesse, par rapport à des enfants de femmes non exposées. Pour les agricultrices, le risque était augmenté de 40%. De plus, le risque de leucémie est lié aussi à la nature du pesticide.

1.2.6 Pesticides et diabète

Les utilisateurs professionnels de pesticides qui ont employé des pesticides chlorés pendant plus de 100 jours durant leur vie ont un risque accru de diabète, d'après une étude de chercheurs de l'Institut National pour la Santé des Etats Unis (Montgomery et al., 2008). Selon la nature des pesticides, le risque peut être accru de 20 à 200%. Cette étude qui a été conduite sur plus de 30 000 agriculteurs, montre que parmi les 50 pesticides différents auxquels les chercheurs se sont intéressés, 7 produits en particulier ont retenu leur attention (aldrine, chlordane, heptachlor, dichlorvos, trichlorfon, alachlore et cynazine).

1.2.7 Pesticides et pathologies neurologiques

Parmi les effets neurotoxiques retardés qui pourraient être liés à l'utilisation de pesticides, BALDI et al. (2012) retiennent les troubles suivants : les polyneuropathies, les troubles neuropsychologiques, la Maladie de Parkinson. Dans une autre étude sur les effets retardés des pesticides sur la santé humaine, il a été conclu que l'exposition professionnelle aux pesticides organochlorés et organophosphorés est associée à l'apparition de troubles neuropsychologiques et neurocomportementaux (Multigner, 2005).Un certain nombre d'études a révélé que les personnes exposées aux pesticides (insecticides et herbicides) sont celles qui présentent un risque élevé de contraction de la maladie de Parkinson (Freire et Koifman, 2012; Van Maele-Fabry et al., 2012 ; Van der Mark et al., 2012).

Lors des tests touchant les pesticides, certains auteurs ont trouvé un lien direct et significatif entre l'exposition professionnelle aux pesticides organophosphorés et le développement de la maladie d'Alzheimer au cours de leur vie (Hayden et al., 2010). De plus, Parron et al. (2011) ont montré que les personnes vivant dans les zones à haut niveau d'utilisation des pesticides ont un risque élevé de contracter la maladie d'Alzheimer. L'exposition à des pesticides semble également liée à un risque plus grand de développer les maladies de Parkinson et d'Alzheimer. Ainsi une étude montre que, chez des agriculteurs hommes, le risque de développer la maladie de parkinson était multiplié par 5.6 et celui de développer la maladie d'Alzheimer multiplié par 2.4 par rapport à des groupes non exposés (Baldi et al., 2003). Une autre étude plus récente (Costello et al., 2009) montre que l'exposition aux pesticides Maneb et/ou paraquat augmente en moyenne de 75% le risque de développer la maladie de Parkinson chez les personnes exposées. Le risque est maximum chez les jeunes sujets exposés chez qui le risque est multiplié par 2.27 suite à l'exposition à un de ces deux pesticides ou multiplié par 4.17 en cas d'exposition aux deux pesticides.

1.2.8 Autisme

Des scientifiques américains (Roberts et al., 2007) ont cherché à évaluer le risque de développer des pathologies comme l'autisme pour les enfants à naître de femmes enceintes exposées à des pesticides utilisés dans des zones agricoles proches (moins de 500 m) et ce pendant les premières semaines de grossesse. Les résultats furent éloquents. En effet, pour les femmes les plus exposées par leur environnement à des pesticides organochlorés comme le dicofol et l'endosulfan pendant le début de leur grossesse, les scientifiques ont mis en évidence un risque de donner naissance à des enfants autistes 6 fois plus important que pour les femmes ne vivant pas près des zones agricoles. Par ailleurs, le risque augmente en fonction des quantités de pesticides utilisées et de la proximité de la zone d'habitation des zones d'utilisation des pesticides.

1.2.9 Affections respiratoires et pesticides

Une analyse (Hoppin et al., 2007) a mis en évidence un excès de risque pour la survenue de bronchites chroniques avec l'emploi de deux insecticides : le diazinon et le malathion. Par ailleurs, dans une étude consacrée à des fermiers employant une quarantaine de pesticides, des sifflements respiratoires ont été signalés par 19% de paysans (Hoppin et al., 2002).

1.3 TOXICITE ET MODELES DE PREDICTION QSAR

Dans cette partie de la revue bibliographique, on parlera de la littérature scientifique consacrée à l'utilisation des relations quantitatives structure-activité (*QSAR*) en vue de la prédiction de la toxicité aigüe des substances chimiques et ce de façon générale. Par cette revue bibliographique, nous aspirons à montrer tout l'intérêt grandissant et l'utilité que revêt cette notion de QSAR. Aussi, à titre d'illustration, nous avons reporté sur la figure 1.4 ci-contre, le nombre de publications dédiées à l'étude de la toxicité par la méthode QSAR et ce de l'année 2003 à l'année 2013.Sur la base de l'examen de la littérature, plus d'une centaine d'articles traitant de ce sujet voient le jour chaque année. Nous relaterons ci-après, quelques publications intéressantes qui ont retenu notre attention.

Pour prédire les DL_{50} orales de 95 alcools chez les rats et les souris, Wang et Bai (1998) ont proposé une échelle de classification basée sur des règles simples. Les données de toxicité ont été converties en quatre classes. Quatre descripteurs décrivant la présence (codée par un "oui") ou l'absence (codée par un "non") d'un groupe alkyle, alcényle, alcynyle ou d'un alcool aromatique et deux descripteurs correspondant au nombre de groupes hydroxyle et au nombre d'atomes de carbone ont été utilisés. Les résultats de la classification coïncident correctement avec les données expérimentales de 80 alcools. Il est à noter que seules les valeurs calculées de 9% des alcools utilisés n'étaient pas corrélées avec les valeurs expérimentales correspondantes. Par ailleurs, pour tester le pouvoir prédictif du système, un ensemble de test externe de 25 alcools a été utilisé. Les valeurs prédites de 22 alcools étaient en bon accord avec les données expérimentales alors que pour 3 alcools, la prédiction était incorrecte. Les auteurs ont conclu que les alcools constitués de 4 à 10 atomes de carbone étaient les plus toxiques.

Jäckel et Klein (1991) ont utilisé DL_{50} rat par voie orale de la base de données "Register of Toxic Effects of Chemical Substances" (RTECS) pour prédire la toxicité aiguë des anilines substituées. Ils ont montré que la toxicité a été prédite mieux grâce à une combinaison de paramètres électroniques, stériques et hydrophobes. La relation de la toxicité aiguë des alkyl benzènes au nombre de 9 a été examinée chez des souris par Tanii et al. (1995). Deux types de modèles simples reliant le $log\ (1/DL_{50})$ au coefficient de partition octanol/eau (*log P*) ont été obtenus. Les mêmes auteurs ont établi d'autres modèles de relation structure activité avec l'éthylène glycol éthers (Tanii et al., 1992).

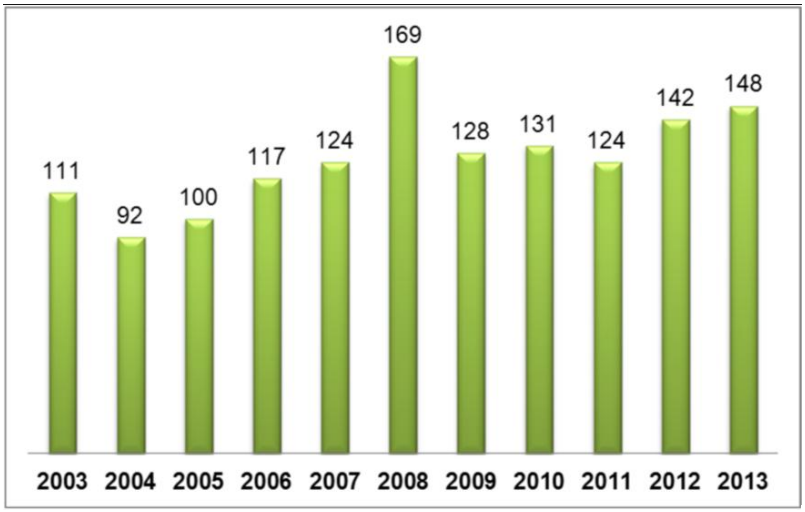


Figure 1.4 : Nombre de publications inhérentes aux études QSAR de la toxicité des substances Chimiques (PubMed consulté le 18/04/2014).

Pour leur part, Johnson et Jurs (1997) ont utilisé la régression et des réseaux de neurones artificiels à 3 couches réseau neural (RNA) pour prédire la toxicité aiguë d'anilines substituées sur les souris. Une base de données de 115 valeurs de DL_{50} orale de souris a été divisée aléatoirement en deux sous-ensembles : de formation (103 valeurs) et de validation externe (12 valeurs). 166 descripteurs (topologiques, électroniques et géométriques) ont été calculés pour chaque composé. Un modèle de régression linéaire utilisant cinq descripteurs a été trouvé avec une erreur quadratique moyenne (*EQM*) de 0,27. Avec le *RNA*, la valeur de l'*EQM* est de 0,25. Toutefois, le fait d'exprimer la DL_{50} en mg / kg au lieu de mmole/kg (ou mole/kg) a été le seul inconvénient pour ces modèles.

Zhu et al. (2009a) ont récemment rapporté une nouvelle approche de modélisation pour prédire la toxicité aiguë pour rongeur. Dans cette étude, les auteurs ont divisé l'ensemble de données en deux groupes, à savoir les composés avec une bonne ou une mauvaise corrélation entre l'IC_{50} et la LD_{50}. Ensuite, à l'aide de descripteurs exclusivement chimiques, les auteurs ont développé des modèles spécifiques de *QSAR*. En raison de leur complexité, il est difficile de juger si les modèles sont reproductibles, car ni l'algorithme utilisé est transparent, ni les compositions exactes des ensembles de formation et de test sont disponibles.

Un certain nombre de modèles *QSAR* pour la toxicité aiguë par voie orale chez le rat utilisant de grands ensembles de données ont été développés et rapportés par Zhu et al., (2009b). Ces modèles ont été construits en utilisant plusieurs ensembles de descripteurs et

employant plusieurs méthodes de modélisation statistique. Une meilleure précision dans la prédiction de la LD_{50} était perceptible par rapport aux modèles individuels. Toutefois, en raison de l'utilisation d'un grand nombre de descripteurs et de la grande complexité de la procédure de modélisation, ces modèles sont difficiles à reproduire, donc ils ne sont pas facilement transférables et utiles dans la pratique. Néanmoins, la procédure de modélisation élaborée dans cette étude est intéressante pour servir de base à d'autres recherches sur les méthodes de calcul permettant de prédire la toxicité aiguë par voie orale.

Des modèles appelés *AMT* ont été proposés par Raevsky et al. (2010). Les auteurs ont montré que les valeurs de DL_{50} ont pu être prédites avec des valeurs R^2 allant jusqu'à 0,78. L'erreur de prévision était de l'ordre de 0,30 à 0,52 unités logarithmiques, ce qui est comparable à l'erreur expérimentale associé à la détermination de la DL_{50}. Dans un autre travail, dix-neuf fluoroacetamides ont été testés pour leur toxicité aiguë sur des rats adultes Wistar (Juranic et al., 2006). Les DL_{50} converties en mol/kg ont été corrélées avec différents descripteurs physico-chimiques. Neuf équations ont été établies avec quatre descripteurs. De leur étude, ils ont déduit que les facteurs qui influent sur la toxicité aiguë de fluoroacetamides sont essentiellement les propriétés électroniques du groupement amide.

1.4 MODELES QSAR ET PREDICTION DE LA TOXICITE AIGÜE DES PESTICIDES CHEZ LES RATS

Il existe un nombre limité de modèles QSAR, disponibles dans la littérature, pour la prédiction de la toxicité aiguë des pesticides chez les rats. Le tableau 1.1 énumère les seuls exemples notables d'études QSAR de la DL_{50} publiées au fil des années.

Adamson et al. (1984) ont examiné la relation entre la DL_{50} chez le rat par voie orale de 129 herbicides benzimidazoles et leur structure chimique à l'aide de descripteurs structurels générés par ordinateur. Les auteurs suggèrent que cette approche a permis une estimation approximative de la toxicité des composés, mais des prévisions fiables de haute précision n'était pas possible. Nendza et al. (1991) ont établi un modèle de prédiction de la toxicité aigüe d'herbicides du type phénylurée. Les DL_{50} par voie orale chez le rat (mmol/kg) obtenues à partir de la base de données RTECS et les paramètres hydrophobes et électroniques ont été utilisés. Les auteurs ont souligné que les deux modèles quoiqu'ils étaient intéressants mais restent insuffisants car des écarts entre les valeurs expérimentales et prédites ont été relevées. La toxicité aiguë par voie orale de 50 herbicides amides chez le rat (DL_{50} en

mmol/kg) a été prédite moyennant l'utilisation d'indices electrotopologiques (Gough et Hall, 1999). Un modèle à cinq variables produit des statistiques fiables. L'équation du modèle obtenu a été utilisée pour prédire la toxicité de neuf amides, non compris dans l'ensemble d'apprentissage. L'erreur moyenne absolue du jeu de test était égale à 0,27.

Zakarya et al. (1996) ont voulu comparer les performances de la régression linéaire et celle d'un perceptron à trois couches (RNA) lors de la prédiction de la toxicité aigüe (DL_{50}) d'un ensemble de 44 herbicides amides. Il a été démontré que l'outil statistique qu'est le réseau de neurone artificiel est plus performant que la régression linéaire. De leur côté, Eldred et Jurs (1999) ont tenté de modéliser la DL_{50} orale chez le rat et ce pour une série de pesticides organophosphorés. Pour ce faire, ils ont utilisé 212 descripteurs dont 119 indices topologiques, 29 descripteurs géométriques, 18 descripteurs électroniques et 46 descripteurs de charges. La procédure de recherche des descripteurs pertinents a abouti à la sélection de 29 descripteurs. Deux modèles, l'un de régression linéaire avec sept descripteurs et l'autre obtenu par réseaux de neurones (RNA) ont été proposés. Le modèle avec les RNA a été plus fiable que celui avec régression et ce par comparaison des EQM qui sont respectivement de 0.9 et 0.25.

Tableau 1.1 : Vue d'ensemble des travaux antérieurs sur les modèles QSAR de prédiction de la toxicité orale aigüe pour le rat (Sazonovas et al., 2010).

Composés étudiés	Nombre de composés	Méthode QSAR	Références
Herbicides	27	AR	Can et al. (2013)
Variés	253	ANN	Zhu *et al*. (2009)
Organophosphorés	62	AR	Garcia-Domenech *et al*. (2007)
Organophosphorés	30	CoMFA	Guo *et al*. (2006)
Organophosphorés	60	RNA	Devillers (2004)
Organophosphorés	67	AR	Zahouily (2002)
Amide herbicides	59	AR	Gough and Hall (1999)
Organophosphorés	54	RA / RNA	Eldred and Jurs (1999)
Herbicides amides	44	AR	Zakarya *et al*. (1996)
Herbicides phenylurée	12	AR	Nendza (1991)
Variés	525	AR	Enslein, (1978)

AR : Analyse de régression
RNA : Réseau de Neurones Artificiels
CoMFA: Comparative Molecular Field Analysis

Un autre modèle QSAR pour le même type de pesticides a été déduit de l'étude de prédiction de la DL_{50} orale chez le rat par Zahouily et al. (2002). Les données de toxicité aiguë de 47 pesticides, provenant de différentes sources, ont été converties en mol/kg. Trois descripteurs décrivant les propriétés des substituants ont été utilisés et un modèle de régression en a été le résultat. Par ailleurs, ces trois descripteurs ont été utilisés comme entrées d'un RNA avec perceptron à trois couches et utilisation d'un algorithme de rétro-propagation. Le meilleur modèle a été celui généré par le RNA dont la configuration neuronale optimale est 3/5/1. Les résultats statistiques des modèles de régression et de RNA sont respectivement de 0.77 et 0.93 pour les coefficients de détermination R^2et de 0.42 et 0.2 0 pour l'EQM.

Parmi le modèle QSAR pour la toxicité aiguë par voie orale pour rongeur (rat), les pesticides organophosphorés ont reçu une attention particulière (Devillers, 2004). Cet auteur a utilisé la méthode de régression des moindres carrés partiels (PLS : Partial least squares) et les réseaux de neurones artificiels (ANN) pour prédire les valeurs de la DL_{50} pour les rats et ce pour 51 pesticides organophosphorés .Les toxicités aigües ont été converties en mmol/kg et utilisées en 1/log(DL_{50}) et une série de descripteurs a été utilisée. Les meilleurs résultats PLS ont été obtenus avec un modèle avec quatre variables importantes expliquant 64 % de la variabilité de la variable dépendante. L'obtention d'un modèle QSAR non linéaire à partir d'un perceptron à trois couches a été le second but des auteurs. Le meilleur modèle RNA a été obtenu avec la configuration suivante : 8/4/1, alors que l'EQM est de 0,26. Cette étude a mis en évidence l'utilité de descripteurs tels que le caractère lipophile et la réfractivité molaire. Différents descripteurs topologiques ont été utilisés par Garcia- Domenech et al. (2007) pour prédire la DL_{50} orale pour les rats. Les DL_{50} ont été exprimées en mmol/kg avec une transformation logarithmique avant leur utilisation. Un modèle a été généré consistant en une équation avec huit paramètres pour un ensemble de 62 pesticides. Les valeurs du coefficient de détermination R^2 et de L'EQM calculée sont respectivement égales à 0.82 et 0.44.

Dans un travail très récent, Can et al. (2013) ont proposé un modèle quantitatif de relation structure- toxicité (*QSTR*) pour l'estimation de la toxicité aiguë par voie orale d'herbicides à des rats mâles. La LD_{50} (en mmol/kg) de 20 herbicides de l'ensemble de formation et sept autres réservés pour la validation externe a été décrite au moyen de 4 descripteurs (caractère lipophile, polarité, géométrie moléculaire et un descripteur de chimie quantique). L'élaboration du modèle pour prédire la toxicité des herbicides sulfonylurées et phénylurée a été effectuée en utilisant une régression linéaire multiple. Les paramètres statistiques du modèle comme l'EQM et le coefficient R^2 sont respectivement égaux à 0.041 et

0.93. Les auteurs ont estimé que le caractère lipophile, le moment dipolaire, la réfractivité molaire et la masse moléculaire sont des paramètres effectifs qui décrivent la toxicité de ces substances. Aussi, pour que tout nouvel herbicide de cette famille soit le moins toxique possible, il doit avoir les caractéristiques suivantes : être fortement polaire, soluble dans l'eau, avoir une réfractivité faible ainsi qu'une masse molaire également faible.

CONCLUSION

A la lumière de la littérature scientifique consultée, il ressort que les pesticides sont un sujet de préoccupation majeure pour toute l'humanité. Indépendamment de leurs avantages, ils constituent une menace sérieuse aussi bien pour les écosystèmes que pour la santé humaine.

La première partie de cette revue a été consacrée au devenir des pesticides dans l'environnement. L'abondante littérature scientifique qui y est consacrée est une preuve indiscutable quant au fait que toutes les composantes de l'environnement sont contaminées par les pesticides. Ni l'air, ni les eaux aquatiques et ni les sols ne sont épargnés. De plus, l'alimentation est aussi infectée par les résidus de pesticides. Autre fait saillant, tous les pays de la planète sont touchés par un tel fléau.

La seconde partie, dédiée à l'impact de ces pesticides, est très édifiante quant aux effets néfastes et nocifs de ces substances aussi bien sur l'être humain ou sur la faune et la flore. Toutes les enquêtes épidémiologiques et les recherches effectuées sur les animaux convergent vers un seul point : le développement de certaines maladies (cancers, maladie d'Alzheimer, maladie de Parkinson, malformation, …etc.) aurait un lien direct et significatif avec l'exposition aux pesticides. Par ailleurs, les pesticides sont tenus responsables de la disparition, ou de la réduction du nombre, de certaines espèces d'animaux. Cette partie nous renseigne aussi sur les efforts inlassables qui restent à fournir pour élucider le rôle exact de ces substances dans la contraction de ces maladies.

Avec l'importance croissante des problèmes environnementaux, un intérêt continu est accordé aux modèles capables d'évaluer les risques de ces substances polluantes. Dans ce domaine, les modèles QSAR (relations quantitatives structure-activité) semblent d'un grand intérêt pour prédire la toxicité des pesticides, directement à partir de leur structure moléculaire. Aussi, le nombre de publications consacrées aux modèles QSAR pour la prédiction de la toxicité des pesticides, troisième partie de cette revue bibliographique, est révélateur du développement et de l'intérêt de cette méthode. En d'autres termes, il existe un

besoin et un intérêt considérable de développer des outils prédictifs capables de produire des données toxicologiques pour le très grand nombre de produits chimiques, en particulier les pesticides, fabriqués chaque jour. Ces outils seront un moyen de remplacement et de réduction de l'expérimentation animale notamment. A ce titre, une foule de modèles basés sur les relations quantitatives structure-activité (QSAR) ont été proposés pour prédire la DL_{50}.

La quatrième et dernière partie de cette revue bibliographique est inhérente aux publications scientifiques traitant la prédiction de la toxicité aigüe des pesticides pour les rongeurs. Etant l'objectif de la présente thèse, une attention particulière a été accordée à cette quatrième partie de la revue de la littérature scientifique. Tout d'abord, notons que très peu d'articles ont été publiés sur les modèles QSAR de prédiction de la toxicité aigüe (***DL_{50}***) des pesticides pour rongeurs. D'autre part, certains de ces modèles, quoique proposés depuis longtemps et de surcroit intéressants, souffrent d'un certain nombre de lacunes. Dans ce contexte, une analyse critique des modèles existants a été effectuée.

Une analyse détaillée des articles publiés sur les modèles QSAR de toxicité aiguë des pesticides pour les rongeurs montre clairement que ces modèles ont été tirés à partir d'ensembles de données limitées de pesticides (ne dépassant pas la cinquantaine pour la plupart) structurellement similaires telles que les pesticides organophosphorés ou les herbicides amides. Aussi, malgré l'importance des paramètres statistiques obtenus dans les travaux récents, la plupart de ces modèles ont une utilité limitée dans la prédiction. Par ailleurs, le genre est rarement spécifié, notamment pour les anciennes publications. Par conséquent, cela peut induire des erreurs lorsqu'on sait que des différences importantes peuvent exister entre les rats mâles et femelles (souris) en ce qui concerne leur sensibilité aux pesticides. De plus, la souche du rongeur n'est pas généralement indiquée. Là aussi, la fiabilité du modèle est réduite sachant que des différences de sensibilité aux produits chimiques existent entre les différentes souches disponibles pour une espèce unique (Kacew, 1996). A cet égard, une grande attention doit être portée au choix des DL_{50} pour la conception de modèles de corrélation QSAR. Idéalement, les modèles doivent dériver des valeurs de DL_{50} obtenues sur les organismes du même sexe, appartenant à une même souche, mais aussi selon la même voie d'exposition.

Les descripteurs moléculaires sont le second paramètre essentiel à l'établissement d'un modèle QSAR. L'analyse de la littérature scientifique disponible nous amène à conclure que parmi les descripteurs moléculaires utilisés pour dériver les modèles QSAR, le coefficient de partage octanol/eau (*log P*) est le plus largement cité. De plus, l'utilisation de quelques

descripteurs non traditionnels est à signaler. Cependant, la taille de la base de descripteurs de départ et la méthode statistique utilisée pour la sélection des descripteurs les plus pertinents ne sont pas généralement précisées, surtout dans les anciens articles. Notons qu'actuellement, des centaines de descripteurs voir des milliers sont obtenus moyennant l'utilisation de logiciels et que plusieurs techniques statistiques de sélection des descripteurs les plus efficaces sont disponibles.

Certains auteurs proposent un modèle linéaire exprimant la toxicité aigüe en fonction des descripteurs utilisés. Par contre, d'autres préconisent un modèle non linéaire obtenu par utilisation des réseaux de neurones artificiels (RNA). Même si les résultats de prédiction quantitatifs de la toxicité sont satisfaisants pour les deux types de modèles, il est mentionné que la capacité de prévision de celui non linéaire est de loin la meilleure. Dans les articles passés en revue, l'établissement du modèle non linéaire s'est effectué par utilisation d'un seul algorithme d'apprentissage et d'une seule fonction de transfert, alors que le mieux serait de faire une optimisation quant au choix de l'algorithme d'apprentissage et des fonctions de transfert à utiliser. D'un autre côté, pour évaluer correctement la puissance prédictive d'un modèle QSAR, on utilise un ensemble de test externe. Notons que les détails sur la validation statistique et les domaines d'application du modèle ne sont pas précisés dans la plupart des modèles publiés. Par conséquent, ces modèles ne peuvent satisfaire pleinement les principes énoncés par l'OCDE pour leur validation.

En guise de conclusion, cette revue bibliographique nous a servi à fixer l'objectif de ce travail et qui consiste en la prédiction de la toxicité orale aigüe des pesticides pour les rats par utilisation des réseaux de neurones artificiels et de la régression linéaire. Toutefois, nous avons fait en sorte de combler toutes les lacunes relevées dans les publications antérieures et ce en :

- ✓ choisissant une base de pesticides la plus large possible
- ✓ établissant une base de données des DL_{50} obtenues dans les mêmes conditions expérimentales (espèce, souche, voie d'exposition).
- ✓ utilisant une large base de descripteurs et en déterminer les plus pertinents.
- ✓ procédant à une optimisation pour le choix de l'algorithme d'apprentissage optimal et pour le choix des fonctions de transfert.
- ✓ tenant compte des principes énoncés (qui seront donnés dans le chapitre 2 de cette thèse) par l'Organisation de Coopération et de Développement Économiques (*OCDE*).

Chapitre 2

RELATIONS QUANTITATIVES STRUCTURE – ACTIVITE

Sommaire

2.1 Utilités et applications des méthodes QSAR ……. 31
2.2 Outils de QSAR ……..……... 32
2.3 Formulation mathématique des QSAR …….. 34
2.4 Methodes d'apprentissage statistiques …….. 35
2.5 La regression linéaire multiple (*RLM*) …….. 35
2.6 Les réseaux de neurones artificiels …… 36

La chémoinformatique est la discipline dont le but est l'étude et la résolution des problèmes relatifs à la chimie en appliquant des méthodes informatiques. Cette discipline est devenue une véritable interface entre la chimie et l'informatique (Willett, 2011). En d'autres termes, la grande quantité de données issues des expériences ne peut être traitée et analysée que par des méthodes informatiques. Ces dernières désignées sous le terme de "techniques in silico" sont nombreuses et variées. Notre intérêt se portera sur l'application de cette technique inhérente à la prédiction de la toxicité à partir de la structure moléculaire.

Les relations quantitatives structure-activité (en anglais QSAR : Quantitative Structure Activity Relationship) sont un exemple typique de cette catégorie. Les relations entre les structures des molécules et leurs activités sont généralement établies à l'aide de méthodes de modélisation par apprentissage statistique. Les techniques usuelles reposent sur la caractérisation des molécules par un ensemble de descripteurs, nombres réels mesurés ou

calculés à partir des structures moléculaires. Il est alors possible d'établir une relation entre ces descripteurs et la grandeur modélisée. Parmi les méthodes d'apprentissage qui sont nombreuses, notre choix s'est porté sur celle relative aux réseaux de neurones artificiels (*RNA* ou *ANN* en anglais : Artificial Neural Networks).

2.1 UTILITES ET APPLICATIONS DES METHODES QSAR

Initiées sur la base de relations empiriques durant la seconde partie du 19ème siècle, les QSAR ont été mathématiquement formalisées dans les années 1930 et leur acceptation, en tant que discipline à part entière, a été définitivement acquise au début des années 1960 suite aux travaux de Hansch et Fujita (Devillers, 2013) qui ont proposé un modèle mathématique pour corréler l'activité biologique et la structure chimique. Depuis lors, un grand nombre de modèles a été conçu pour la prédiction de diverses activités d'intérêt biologique.

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

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

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

- ✓ l'optimisation de l'activité pharmacologique, biocide ou pesticide.
- ✓ la conception rationnelle de nombreux autres produits tels que des agents tensio-actifs, des parfums, des colorants ou des produits chimiques.
- ✓ l'identification des composés dangereux à des stades précoces de développement.
- ✓ la conception de la toxicité et des effets secondaires pour les nouveaux composés.
- ✓ la prédiction de la toxicité pour les humains pour les multiples expositions (délibérée, occasionnelle ou professionnelle) et de l'écotoxicité.

- ✓ la prédiction d'une variété de propriétés physico- chimiques des molécules (qu'il s'agisse de produits pharmaceutiques, de pesticides, de produits d'hygiène personnelle, la chimie fine, etc.).

2.2 OUTILS DE QSAR

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

2.2.1 Données biologiques

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

2.2.2 Les descripteurs moléculaires

De nombreuses recherches ont été menées, au cours des dernières décennies, pour trouver la meilleure façon de représenter l'information contenue dans la structure des molécules. Ainsi, les recherches ont abouti à la représentation de la structure moléculaire par un ensemble de valeurs numériques appelées descripteurs moléculaires. On en dénombre aujourd'hui plus de 3000 types, qui quantifient des caractéristiques physico-chimiques ou structurelles de molécules. Ils peuvent être obtenus de manière expérimentale, mais les descripteurs calculés, et non mesurés, sont à privilégier : ils permettent en effet d'effectuer des prédictions sans avoir à synthétiser les molécules, ce qui est un des objectifs de la modélisation. Les descripteurs moléculaires peuvent être divisés en deux grandes catégories (Poezevara, 2011) : les descripteurs expérimentaux (par exemple : la lipophilie [*logP*], la polarisabilité, la densité électronique, le potentiel d'ionisation) et les descripteurs moléculaires théoriques. Ces derniers sont à nouveaux classés en trois catégories en fonction

de leur dimension (Hong, 20 08).

2.2.2.1 *Les descripteurs 1D*

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

2.2.2.2 *Les descripteurs 2D*

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

- ✓ les indices constitutionnels caractérisent les différents composants de la molécule. Il s'agit par exemple du nombre de liaisons simples ou multiples, du nombre de cycles…
- ✓ les indices topologiques peuvent être obtenus à partir de la structure 2D de la molécule, et donnent des informations sur sa taille, sa forme globale et ses ramifications. L'indice de Wiener par exemple, permet de caractériser le volume moléculaire et la ramification d'une molécule. Un autre indice, nommé "indice de Randic", est l'un des descripteurs les plus utilisés. Il peut être interprété comme une mesure de l'aire de la molécule accessible au solvant.

2.2.2.3 *Les descripteurs 3D*

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

- ✓ **Les descripteurs géométriques** les plus importants sont le volume moléculaire, la surface accessible au solvant, le moment principal d'inertie.
- ✓ **Les descripteurs électroniques** permettent de quantifier différents types d'interactions inter et intramoléculaires, de grande influence sur l'activité biologique de molécules. Par exemple, les énergies de la plus haute orbitale moléculaire occupée

et de la plus basse vacante sont des descripteurs fréquemment sélectionnés. Par ailleurs, le moment dipolaire, le potentiel d'ionisation, et différentes énergies relatives à la molécule sont d'autres paramètres importants.

- ✓ **Les descripteurs spectroscopiques**: les molécules peuvent être caractérisées par des mesures spectroscopiques, par exemples par leurs fonctions d'onde vibrationnelles. En effet, les vibrations d'une molécule dépendent de la masse des atomes et des forces d'interaction entre ceux-ci ; ces vibrations fournissent donc des informations sur la structure de la molécule et sur sa conformation. Les spectres infrarouges peuvent être obtenus soit de manière expérimentale, soit par calcul théorique, après recherche de la géométrie optimale de la molécule. Par exemple, les descripteurs de type *MoRSE* sont calculés à partir d'une simulation du spectre infrarouge.

Par exemple, le programme appelé DRAGON permet de calculer différents groupes de descripteurs : autocorrelations 2D, descripteurs géométriques, descripteurs RDF, descripteurs 3D-MoRSE, descripteurs WHIM, descripteurs GETAWAY, …etc.

2.3 FORMULATION MATHEMATIQUE DES QSAR

Les équations *QSAR* prennent la forme générale suivante :

Activité biologique = fonction (Descripteurs)

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

$$\log\left(\frac{1}{C}\right) = b_0 + \sum_{i=1}^{j}(b_i D_i) \qquad (2.1)$$

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

Les relations *QSARs* les plus connues ont été introduites par C. Hansh et T. Fujita (Poezevara, 2011). Une des équations les plus utilisées présume que les propriétés physiques responsable de l'activité biologique peuvent être séparées en trois catégories : (i) une

composante hydrophobique (π), (ii) une composante électronique (σ) et (iii) une composante stérique (E_S, RM) :

$$\log\left(\frac{1}{C}\right) = a + b\pi + c\pi^2 + dE_s + eRM + f\sigma \qquad (2.2)$$

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

2.4 METHODES D'APPRENTISSAGE STATISTIQUES

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

2.5 LA REGRESSION LINEAIRE MULTIPLE (*RLM*)

La régression linéaire multiple est la plus simple des méthodes de modélisation. Son but est de rechercher une relation linéaire mettant en jeu la variable Y (par exemple la toxicité) à modéliser et "p" variables x_j (les descripteurs par exemple). Le nombre d'observations étant "n". L'équation linéaire de régression recherchée s'écrit :

$$y_i = a_o + a_1 x_{i,1} + \cdots + a_p x_{i,p} + \varepsilon_i \qquad (2.3)$$

où y_i est la $i^{ème}$ observation de la variable Y ; $x_{i,j}$ est la $i^{ème}$ observation de la $j^{ème}$ variable ; a_i représentent les coefficients du modèle et ε_i est l'erreur du modèle. ε_i est défini comme la différence (figure 2.1) entre la valeur observée de la grandeur à modéliser pour cet élément i (y_i) et l'estimation du modèle ($\hat{y}_i$).

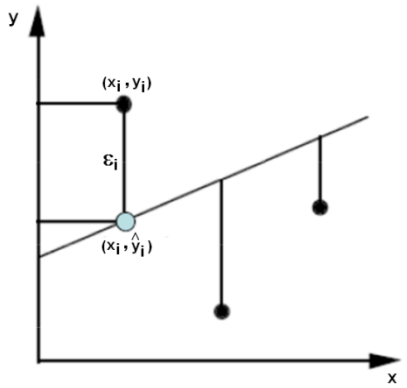


Figure 2.1 : Méthodologie du calcul de ε_i

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

2.6 LES RESEAUX DE NEURONES ARTIFICIELS

L'objectif de cette partie est de donner un aperçu sur les réseaux de neurones artificiels (*RNA*). En effet, ces réseaux nécessitent la sélection des variables d'entrée x_i (descripteurs par exemple) ayant une influence sur la donnée "*y*" à modéliser (toxicité par exemple). Ensuite, il s'agit de développer une fonction algébrique *f*, reliant l'évolution de la toxicité à ces variables. Ceci sera développé à l'aide d'une modélisation, par régression non linéaire vis-à-vis des paramètres de la fonction estimée, utilisant les réseaux de neurones artificiels (*RNA*) (figure 2.2).

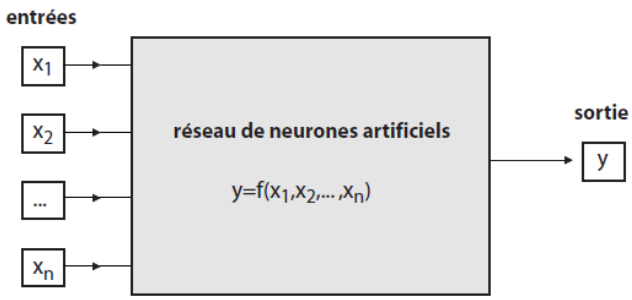


Figure 2.2 : Réseau de neurones artificiel à n variables d'entrée et une sortie

Les réseaux de neurones tirent leur puissance de modélisation de leur capacité à capter les dépendances complexes qui impliquent plusieurs variables simultanément. Les *RNA*

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

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

2.6.1 Le neurone formel

Le neurone formel est l'élément essentiel d'un réseau de neurones. C'est un opérateur mathématique très simple, dont on peut facilement calculer la valeur numérique. Son fonctionnement est schématisé sur la figure 2.3. Un neurone formel est une fonction algébrique paramétrée, non linéaire en ses paramètres, et à valeurs bornées. Ses entrées peuvent être les sorties d'autres neurones ou des entrées de signaux extérieurs. Sa sortie est une fonction non linéaire f d'une combinaison linéaire v des entrées (x_i). Le potentiel v le plus fréquemment utilisé est la somme pondérée des entrées x_i pondérées par les coefficients (w_i) également appelés poids de connexions :

$$v = w_0 + \sum_{i=1}^{n} (w_i\, x_i) \qquad (2.4)$$

Précisons que le poids w_0 est affecté à une entrée constante appelé "biais". Une fonction f, appelée fonction d'activation, est appliquée à ce potentiel v. Cette fonction est la plupart du

temps une fonction en "*S*", bornée en ses extrema, continue et dérivable. Le choix d'une fonction d'activation se révèle dans certains cas être un élément constitutif important des réseaux de neurones. Il en existe plusieurs types, les plus utilisés étant la fonction sigmoïde, la fonction arctangente et la fonction tangente hyperbolique. Un neurone formel, réalise donc l'opération suivante :

$$y = f(v) = f\left(w_0 + \sum_{i=1}^{n}(w_i\, x_i)\right) \tag{2.5}$$

La fonction *f* peut être définie de la manière suivante :

$$f(x) = th(x) = \frac{(e^x - e^{-x})}{(e^x + e^{-x})} \tag{2.6}$$

Un neurone formel réalise donc une somme pondérée suivie d'une non-linéarité. Sa sortie est par conséquent une fonction non linéaire d'une combinaison linéaire des entrées pondérées par les poids.

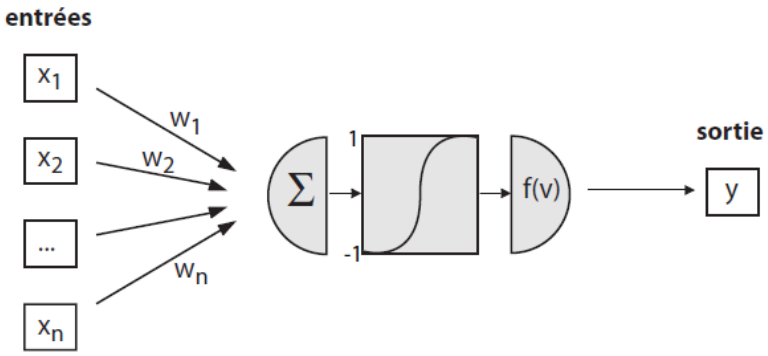


Figure 2.3: Schématisation d'un neurone formel

2.6.2 Architecture d'un réseau de neurones

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

- ✓ les réseaux de neurones non bouclés. Les réseaux de neurones non bouclés réalisent une (ou plusieurs) fonction(s) algébrique(s) des entrées, par composition des fonctions réalisées par chaque neurone. Les réseaux de neurones non bouclés sont des

outils statiques, utilisés principalement pour effectuer des tâches d'approximation de fonctions non linéaires, de modélisation de processus statiques non linéaires.

- ✓ Les réseaux de neurones bouclés. Les réseaux de neurones bouclés constituent quant à eux un système dynamique "à temps discret", régi par une (ou plusieurs) équation(s) aux différences non linéaires, résultant de la composition des fonctions réalisées par chacun des neurones et des retards associés à chacune des connexions. Ils sont utilisés pour effectuer des tâches de modélisation et d'adaptation de systèmes dynamiques, de commande de processus, ou de filtrage (Dreyfus, 1998).

Dans notre étude, nous avons adopté le modèle le plus couramment utilisé, le Perceptron Multi- couches(*PMC*), illustré par la figure 2.4 et dont les principes sont exposés dans ce qui suit.

2.6.3 Perceptron multicouche

Le PMC est un réseau de neurones non bouclé caractérisé par une structure bien particulière (Hornik et al., 1989,1994). La figure2.4 illustre la structure de ce type de réseau. Ses neurones sont organisés en couches successives où les informations circulent dans un seul sens, de la couche d'entrée vers la couche de sortie. Les neurones d'une même couche ne sont pas interconnectés. Un neurone ne peut envoyer son résultat qu'à un neurone situé dans une couche postérieure à la sienne. Le réseau est formé de plusieurs couches, qui s'appellent, par convention :

- ✓ couche d'entrée : cette couche représente toujours une couche associée aux entrées du système, composée des cellules d'entrée qui correspondent aux *k* variables d'entrée.
- ✓ couche(s) cachée(s) : chaque couche est composée de p neurones, dont la fonction d'activation est de la famille des courbes en "S" bornées. Ces neurones n'ont aucun lien avec l'extérieur et sont appelés neurones cachés (Rivals, 1995).
- ✓ couche de sortie : elle est constituée par la valeur de sortie de la combinaison linéaire des fonctions sigmoïdes de la couche cachée.

Les neurones sont reliés entre eux par des connexions pondérées. Ce sont les poids de ces connexions qui gouvernent le fonctionnement du réseau et programment une application de l'espace des entrées vers l'espace des sorties, à l'aide de la transformation non linéaire (Parizeau, 2004)

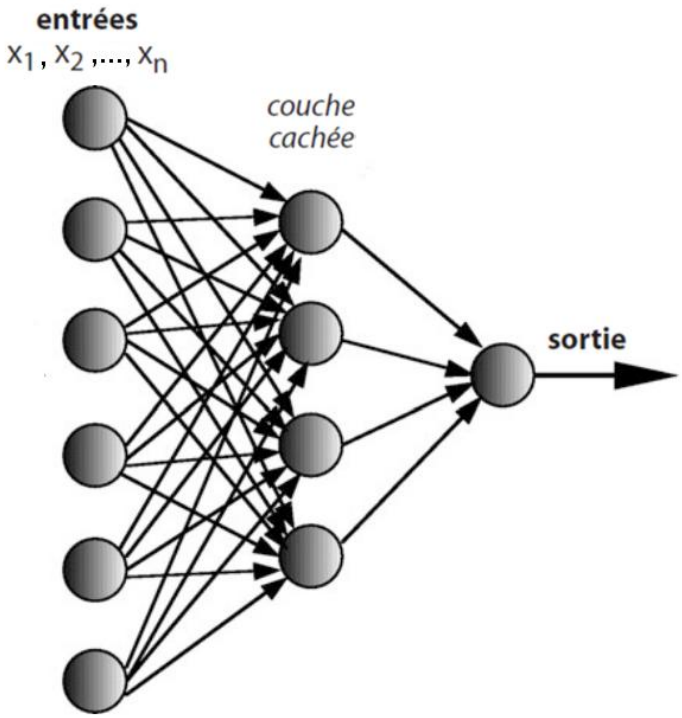


Figure 2.4 : Exemple d'architecture d'un *PMC* à n variables d'entrée, une couche cachée avec quatre neurones cachés et une couche de sortie

2.6.4 Expression algébrique de la régression neuronale

Au final, l'expression algébrique du réseau de neurones, de la fonction de régression non linéaire neuronale, aura pour expression :

$$y = w_{0,s} + \sum_{j=1}^{N} \left(w_j * f \left(w_{0,j} + \sum_{i=1}^{n} (w_i x_i) \right) \right) \qquad (2.7)$$

Avec :

n : le nombre de variables d'entrées

N : le nombre de neurones dans la couche cachée

w_i: les poids relatifs à la connexion entre variables d'entrée et neurone

w_j: les poids relatifs à la connexion allant du neurone j vers la sortie

$w_{O,j}$: les poids relatifs à la connexion entre biais et neurone

$w_{O,S}$: le poids relatif à la connexion entre biais et sortie

2.6.5 Optimisation de la structure du réseau

Les performances du réseau de neurones sont très souvent liées à son architecture. Or, le choix de l'architecture est aussi un problème difficile à résoudre. On ne peut pas choisir au hasard le nombre de neurones de la couche d'entrée ou de la couche intermédiaire. Avec un nombre limité de neurones, le réseau ne sera pas performant sur l'apprentissage. Avec un nombre trop important, il sera surentraîné et s'adaptera au bruit généré par les variables qui constituent la base de données. On juge la performance d'un réseau sur sa capacité à généraliser les résultats et non pas sur sa capacité à reproduire uniquement les données connues de l'apprentissage. Ainsi, il n'est pas toujours profitable d'utiliser un réseau avec un nombre élevé de neurones. Il existe des algorithmes d'adaptation, qui en partant d'un réseau avec un nombre élevé de neurones, enlèvent des neurones jusqu'à ce que les performances du réseau en pâtissent de manière trop conséquente. Il existe aussi des algorithmes qui partent d'un réseau ayant peu de neurones et qui ajoutent progressivement des neurones afin d'en améliorer les performances.

La détermination de l'architecture optimale est basée sur diverses méthodes empiriques mettant en jeu plusieurs tests statistiques da validation. Comme nous l'avons mentionné, nous avons retenu le Perceptron Multicouches comme base du modèle. Nous structurerons ce réseau en précisant le nombre de couches et de neurones cachés pour que le réseau soit en mesure de reproduire ce qui est déterministe dans les données.

- ✓ **Nombre de couches cachées** : Mis à part les couches d'entrée et de sortie, il faut décider du nombre de couches intermédiaires ou cachées. Sans couche cachée, le réseau n'offre que de faibles possibilités d'adaptation. Néanmoins, il a été démontré qu'un Perceptron Multicouches avec une seule couche cachée pourvue d'un nombre suffisant de neurones, peut approximer n'importe quelle fonction continue et bornée avec la précision souhaitée (Hornik, 1991).
- ✓ **Nombre de neurones cachés** : Chaque neurone peut prendre en compte des profils spécifiques de neurones d'entrée. Un nombre plus important permet donc de mieux coller aux données présentées mais diminue la capacité de généralisation du réseau. Il faut alors trouver le nombre adéquat de neurones cachés nécessaire pour obtenir une approximation satisfaisante. Le nombre de neurones caches dépend du type de données dont on dispose. Suivant les cas, on utilisera différentes méthodes d'optimisation de ce dernier. La détermination de ce nombre est généralement le fruit d'essais empiriques. La technique d'usage consiste à augmenter progressivement la

complexité du modèle, en ajoutant un neurone. A chaque nouvelle addition d'un neurone, on évalue l'efficacité de généralisation. Quant au nombre de couches cachées, rappelons que nous avons choisi la structure la plus simple existante, comportant une seule couche cachée, une couche d'entrée et une couche de sortie.

2.6.6 Mise en œuvre de la régression neuronale

La construction d'un modèle implique dans un premier temps le choix des échantillons des données d'apprentissage, de test et de validation. Le choix du type de réseau intervient dans un second temps. Les quatre grandes étapes de la création d'un réseau de neurones seront détaillées par la suite.

2.6.7 Construction de la base de données et choix des échantillons

Le processus d'élaboration d'un réseau de neurones commence par la construction d'une base de données et le choix des échantillons. Il faut donc déterminer les grandeurs qui ont une influence significative sur le phénomène que l'on cherche à modéliser. Cette étape est cruciale et aide à déterminer le type de réseau le plus approprié pour résoudre un problème donné. En particulier, la façon dont se présente l'échantillon conditionne le type de réseau, le nombre de cellules d'entrée et de sortie et la façon dont il faudra mener l'apprentissage, le test et la validation (Bishop, 1995). Afin de développer une application à base de réseaux de neurones, il est nécessaire de disposer de deux bases de données, une pour effectuer l'apprentissage et l'autre pour tester le réseau obtenu et déterminer ses performances. Pour mieux contrôler la phase d'apprentissage, il est souvent préférable de posséder une troisième base de données appelée base de validation croisée (Stone et al. 1977). Notons qu'il n'y pas de règle pour déterminer ce partage d'une manière quantitative, néanmoins chaque base doit satisfaire aux contraintes de représentativité de chaque classe de données et doit généralement refléter la distribution réelle, c'est-à-dire la probabilité d'occurrence des diverses classes (Nascimento et al., 2000). Généralement, les bases de données subissent un prétraitement qui consiste à effectuer une normalisation appropriée tenant compte de l'amplitude des valeurs acceptées par le réseau. Les valeurs d'entrée et de sortie sont normalisées dans un intervalle spécifique afin de donner à chaque paramètre la même influence statistique. Nous détaillerons dans le chapitre IV l'ensemble des données

utilisées dans cette étude et la méthodologie employée concernant ce découpage entre base d'apprentissage et base de validation.

2.6.8 Conception du modèle non linéaire

La conception d'un modèle non linéaire en ses paramètres comprend plusieurs étapes dont la sélection des entrées du modèle. Ces entrées doivent être dans notre cas, physiquement et chimiquement représentatives de la toxicité. Autrement dit, il est nécessaire de vérifier que l'information contenue dans les variables d'entrée a bien une influence sur la variable de sortie que l'on tente de modéliser. Si ce n'est pas le cas, il est indispensable de les supprimer. Cette réduction dimensionnelle (sans perte d'information) est effectuée par l'utilisation de méthodes statistiques. Cette sélection doit évidemment tendre vers l'exhaustivité, sans introduire de redondance. Cette sélection, à partir de la somme des variables candidates à l'élaboration de la fonction de régression, doit nous permettre d'éliminer les moins pertinentes d'entre elles.

Plusieurs techniques statistiques d'élagage des variables d'entrée pourraient être utilisées. L'une d'entre elles consiste à maximiser un critère statistique en retirant pas à pas les poids w_i des connexions neuronales qui n'apportent pas d'amélioration de l'erreur quadratique moyenne (EQM). Cette élimination conduit à ne garder que les variables d'entrée pertinentes. Le critère utilisé, appelé BIC (Bayesian Information Criterium), représente le logarithme népérien de l'erreur moyenne quadratique pénalisée par une fonction entre le nombre de poids et de données d'apprentissage :

$$BIC = ln\left(\frac{EQM}{N}\right) + W\frac{\ln(N)}{N} \tag{2.8}$$

où:

N est la taille des données d'apprentissage

W est le nombre de paramètres ou poids ajustables w_i des différentes connexions

EQM est l'erreur quadratique moyenne, définie par:

$$EQM = \frac{1}{n}\sum_{i=1}^{n}\left(y_{prédite} - y_{observée}\right)^2 \tag{2.9}$$

avec $y_{observé}$ et $y_{prédite}$ les valeurs respectivement mesurées et estimées par le modèle.

2.6.9 Apprentissage

Une fois l'architecture choisie, elle doit subir une phase d'apprentissage. Cette dernière consiste à calculer les poids w_i de telle manière que les sorties du réseau de neurones soient, pour les exemples utilisés lors de l'apprentissage, aussi proches que possibles des sorties désirées. Celles-ci correspondent à la valeur de la fonction que l'on veut approcher ou à la sortie du processus que l'on veut modéliser. L'apprentissage correspond donc à la phase du développement du réseau durant la quelle les coefficients de la régression sont modifiés jusqu'à l'obtention du comportement désiré. Il existe deux types d'apprentissage, supervisé ou non supervisé.

- ✓ L'apprentissage est dit supervisé lors que les exemples sont constitués des couples de valeurs du type:(valeur d'entrée, valeur de sortie désirée). Tout le problème de l'apprentissage supervisé consiste, étant donné un ensemble d'apprentissage de n couples (entrée-sortie associée), à déterminer le vecteur des poids w_i d'un réseau capable de mettre ces informations en correspondance.
- ✓ L'apprentissage est qualifié de non supervisé lors que 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 des réponses à des valeurs d'entrée similaires.

Dans notre étude, comme nous disposons des valeurs de sortie, nous avons choisi l'apprentissage supervisé, mettant en jeu un algorithme d'optimisation (figure 2.5).Ce dernier cherche à minimiser, par des méthodes d'optimisation non linéaire, une fonction de coût qui constitue une mesure de l'écart entre les réponses réelles du réseau et ses réponses désirées (Dreyfus et al., 2002).

On distingue deux étapes dans la phase d'apprentissage :

- ✓ une phase de propagation, qui consiste à présenter une configuration d'entrée au réseau, puis à la propager à celle de sortie en passant par les couches cachées.
- ✓ Une phase de rétro propagation, qui consiste, après le processus de propagation, à minimiser l'erreur commise sur l'ensemble des exemples présentés, erreur considérée comme une fonction des poids synaptiques (Parizeau, 2004). Cette erreur représente la somme des différences au carré entre les réponses calculées et celles désirées pour tous les exemples contenus dans l'ensemble de l'apprentissage.

Le fait que les réseaux de neurones soient non linéaires par rapport aux poids leur confère l'avantage de la parcimonie décrit précédemment, mais, par contre, rend l'apprentissage plus compliqué que la méthode des moindres carrés utilisé en régression classique. En effet, si le modèle est linéaire par rapport à ses poids alors la fonction d'erreur des moindres carrés est quadratique par rapport à ceux-ci: elle ne présente qu'un minimum. Quand le modèle est non linéaire par rapport aux poids, la fonction d'erreur peut présenter plusieurs minima: on parle de "minima locaux", parmi les quels il convient de trouver le "minimum vrai" c'est-à-dire le plus petit. La figure 2.5 illustre ce problème. C'est pour quoi la méthode d'estimation des poids w_i procède par itérations afin de minimiser au mieux la fonction d'erreur.

2.6.10 Algorithme de rétro-propagation

Nous avons vu que le travail d'optimisation de l'architecture du réseau de neurones consiste en résumé à retenir la meilleure fonction d'activation pour un nombre de neurones optimal à l'obtention de la meilleure performance possible compte tenu des données disponibles. Cette optimisation structurelle nous amènera, préalablement à tout calcul, à sélectionner la fonction d'activation parmi les fonctions existantes. Lors de l'apprentissage, un algorithme de rétro-propagation permet de minimiser la fonction d'erreur entre valeurs mesurées et valeurs calculées. Cette minimisation est effectuée en modifiant les poids des connexions de la sortie jusqu'aux entrées. C'est une des règles les plus utilisées pour l'apprentissage supervisé de réseaux de neurones. Nous avons donc utilisé une méthode du second ordre comme algorithme de minimisation itérative des poids du modèle : celle de Levenberg-Marquard. Cette technique permet en effet de converger vers le minimum de la fonction d'erreur plus rapidement du fait qu'elle considère la dérivée seconde de la fonction d'erreur.

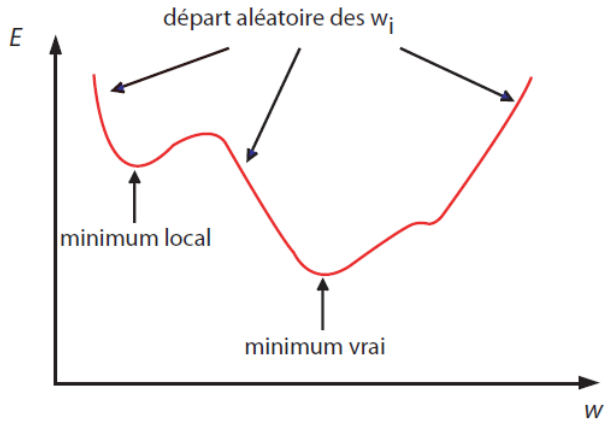


Figure 2.5 : Algorithme de rétropropagation du gradient d'erreur

2.6.11 Apprentissage et sur-apprentissage

D'une manière générale, plus on augmente le nombre de neurones dans la couche cachée, meilleur sera l'apprentissage. Mais choisir le modèle en se basant uniquement sur la qualité de ses performances sur la base d'apprentissage pose un problème. En effet, il y'a un fort risque de retenir une structure soumise au phénomène de sur-apprentissage. Pour éviter cela, il faut veiller à ce que la *REQM* sur les sets d'apprentissage ($REQM_{appr}$) et de validation ($REQM_{valid}$) (équation 2.10 et 2.11) soit du même ordre de grandeur et aussi faibles que possible. La figure2.6 illustre ce compromis entre sur apprentissage, sous apprentissage et généralisation optimale.

$$REQM_{appr} = \sqrt{\frac{1}{n_{appr}} \sum_{i=1}^{n_{appr}} (y_{pred} - y_{observée})^2} \qquad (2.10)$$

$$REQM_{valid} = \sqrt{\frac{1}{n_{valid}} \sum_{i=1}^{n_{valid}} (y_{pred} - y_{observée})^2} \qquad (2.11)$$

Avec n_{appr} et n_{valid} les nombres de données constituant respectivement les bases d'apprentissage et de validation du modèle.

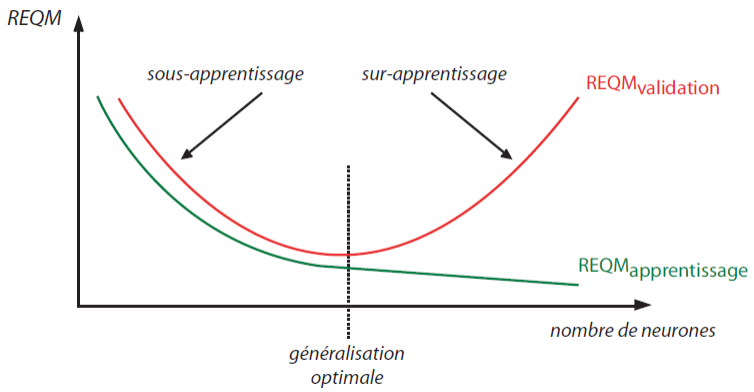


Figure 2.6 : Généralisation optimale

2.6.12 Validation du modèle

Une fois les poids de chaque connexion du réseau optimisés, il faut évaluer les performances du réseau, en d'autres termes la qualité de la régression. Cette qualité garantit les capacités de généralisation du modèle. Cette estimation doit être réalisée sur un ensemble composé de données composant des couples entrée-sortie n'appartenant pas à la base d'apprentissage. On procède à ce que l'on appelle une validation en aveugle. Elle permet à la fois d'apprécier les performances du système neuronal et de détecter le type de

données qui pose problème. Si les performances ne sont pas satisfaisantes, il faudra, soit modifier l'architecture du réseau, soit modifier la base d'apprentissage. C'est ainsi que l'on pourra évaluer la qualité de la généralisation offerte par le modèle établi, à l'aide de divers indicateurs statistiques présentés plus loin dans ce chapitre. Une validation rigoureuse du modèle produit se traduit donc par une proportion importante de prédictions exactes. Cette évaluation ne pourra être faite que sur cet ensemble de validation en aveugle uniquement.

2.6.13 Détermination des performances du modèle

Nous avons vu précédemment que pour déterminer quelle était l'architecture optimale du PMC, nous avions recours à un indicateur statistique, la *REQM*. Une fois le modèle optimal choisi, nous présenterons ses performances dans le Chapitre IV. Ses performances seront estimées à l'aide d'un certain nombre d'indicateurs. On trouve généralement dans la littérature deux groupes d'indicateurs de performance pouvant être utilisés pour évaluer la qualité des modèles. L'un de ces groupes représente l'accord global entre valeurs observées et estimées. Les indicateurs appartenant à ce groupe seront étudiés en détail au chapitre suivant.

En conclusion, nous récapitulons, sur la figure 2.7 toutes les étapes nécessaires à la construction d'un modèle neuronal.

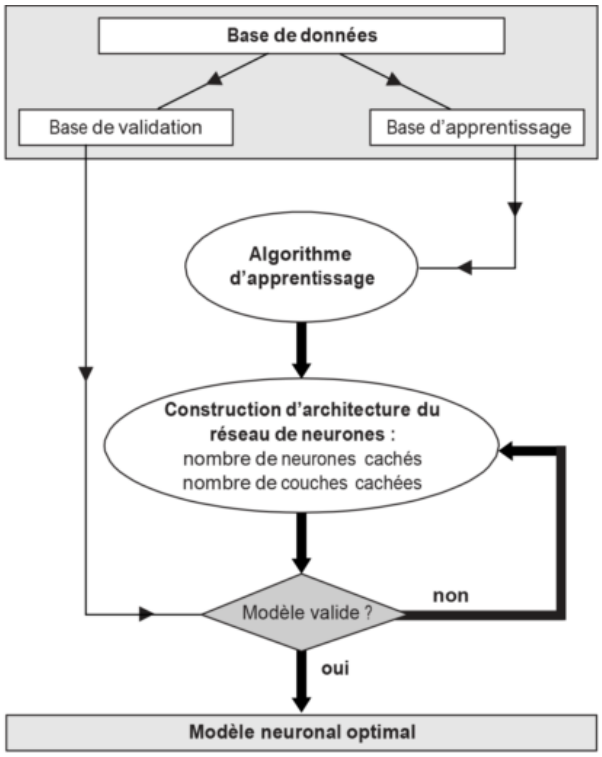


Figure 2.7 : Schéma d'élaboration d'un modèle neuronal

Chapitre 3

LES PESTICIDES

Sommaire

3.1 Historique 49
3.2 Classification des pesticides 50
3.3 Résistance aux pesticides 51
3.4 Toxicologie des pesticides 52
3.5 Utilisation des pesticides en Algérie 54

Le terme pesticide désigne l'ensemble des produits chimiques, naturels ou de synthèse, ayant pour but de repousser ou d'éliminer les organismes jugés indésirables, qu'il s'agisse de plantes, d'animaux, de champignons ou de bactéries et ce durant la production, le stockage ou la commercialisation de produits agricoles, de denrées alimentaires, ou de bois. Ils servent également à combattre les différents vecteurs de maladies humaines ou animales. Les pesticides étant majoritairement utilisés en agriculture, on parle alors de produits phytosanitaires ou phytopharmaceutiques. L'appellation "pesticides" comprend les insecticides, les herbicides, rodenticides, fongicides, nématicides et acaricides ainsi que des désinfectants, des fumigènes et les régulateurs de croissance des plantes.

3.1 HISTORIQUE

Depuis très longtemps, les pesticides ont été utilisés pour protéger les cultures et la santé publique. L'agriculture repose sur les pesticides pour produire des aliments de haute qualité et assurer des approvisionnements réguliers. Dans certains cas, ces substances peuvent faire la différence entre le succès et l'échec d'une culture. Les pesticides sont devenus une partie vitale de l'agriculture moderne et de la protection des aliments des dommages causés par les insectes, les mauvaises herbes, les maladies et les rongeurs.

Les propriétés toxiques de nombreuses plantes sont progressivement découvertes. Maïmonide rédige ainsi, en 1135, un *Traité des poisons*. Leur utilisation en tant que pesticides se répand. L'*aconit* est par exemple employé au Moyen âge contre les rongeurs. La *roténone* est utilisée en Inde comme insecticide dès la fin du *XVI*^{ème} siècle. Un siècle plus tard, on découvre les propriétés insecticides de la *nicotine*, extraite du tabac. Cependant, l'ère des pesticides a pris une ampleur au XIX^{ème} siècle eu égard à l'essor de la chimie minérale.L'utilisation du *soufre* a débuté en 1843 contre les pucerons, ensuite les composés du soufre ont été développés en tant que fongicides. À la fin du XIX^{ème} siècle, les composés de *l'arsenic* (en 1872) et du *sulfate de cuivre* (en 1885) ont été introduits pour lutter contre les insectes qui attaquent les cultures de fruits et de légumes (Levin, 2007).

Faisant suite au développement de la chimie organique de synthèse, les pesticides connurent un nouvel essor. Dans les années 1940, les pesticides organochlorés, notamment le Dichloro Diphényl Trichloroéthane (*DDT*), ont été introduits. Ils ont été largement utilisés dans l'agriculture et dans la lutte contre le paludisme et d'autres maladies transmises par les insectes. Parce qu'ils avaient peu ou pas de toxicité immédiate, leur utilisation a été largement adoptée.

L'utilisation généralisée des pesticides de synthèse a commencé après la Seconde Guerre mondiale. Cette utilisation, qui englobait les insecticides, les rodenticides, les fongicides et les herbicides, a augmenté de plus de trente-trois fois durant la seconde moitié du XX^{ème} siècle (Levin, 2007). Elle a été créditée de l'augmentation spectaculaire des rendements des cultures en réduisant les menaces naturelles et est devenue une partie intégrante des pratiques agricoles.Les prestations de santé, telles que celles liées à l'éradication des moustiques porteurs de maladie comme la malaria, ont été également accomplies avec succès.

La gamme des pesticides disponibles s'est alors rapidement diversifiée et leur usage a connu un très fort développement au cours des décennies suivantes. Petit à petit, ils se sont

imposés dans les usages domestiques (anti-moustiques par exemple), dans le traitement de la voierie publique, des voies ferrés, des aérodromes, des routes, des terrains de golf, des stades, des espaces verts et des jardins.Les pesticides ayant des propriétés antiseptiques et désinfectantes ont eu leur part d'utilisation en médecine. Cependant, cette utilisation massive de centaines de pesticides a aussi engendré divers problèmes, entre autres la disparition d'insectes bénéfiques. Par ailleurs, le développement de résistances vis-à-vis des pesticides (Holmes, 2010 ; Walker *et al*., 2011) a fait que beaucoup d'entre eux ne sont plus efficaces. Cet état de fait a poussé les agriculteurs à utiliser des doses plus élevées, des applications plus fréquentes et des combinaisons de pesticides.

3.2 CLASSIFICATION DES PESTICIDES

Les pesticides sont classés par grandes familles selon un triple classement.

3.2.1 Classement par cible

On distingue quatre grandes familles :

Les insecticides : destinés à lutter contre les insectes. Ils interviennent en tuant ou en empêchant la reproduction des insectes, ce sont souvent les plus toxiques. En voici quelques exemples :

- ✓ L'arsenic, très utilisé avant la seconde guerre mondiale.
- ✓ Le DDT (Dichloro Diphényl Trichloroéthane), insecticide très puissant très utilisé jusqu'à son interdiction, très persistant, très mobile et très soluble.
- ✓ Le carbaryl tristement célèbre car il fut la cause de la catastrophe de Bhôpal (en décembre 1984), due à l'échappement de l'isocyanate de méthyle de l'usine où il était fabriqué.

Les fongicides : destinés à éliminer les moisissures et parasites (champignons...) des plantes. Les fongicides les plus anciens sont le soufre, le cuivre et ses dérivés organiques

Les herbicides : destinés à lutter contre certains végétaux (les " mauvaises herbes "), qui entrent en concurrence avec les plantes à protéger en ralentissant leur croissance. Leur mode d'épandage est différent puisqu'ils sont déposés directement au sol, par opposition aux autres produits, plutôt pulvérisés sur la plante en croissance. Les herbicides constituent aujourd'hui le groupe le plus important, le plus utilisé.

Les pesticides spéciaux : cette famille comprend :

- ✓ les nématicides : ont la propriété d'éradiquer les nématodes (ou vers ronds).
- ✓ les acaricides : ont la propriété d'éradiquer les acariens.
- ✓ les rodenticides : ont la propriété d'éradiquer les rongeurs.
- ✓ les molluscides : ont la propriété d'éradiquer les mollusques (limaces, escargots, …etc.).
- ✓ les algicides : destinée à tuer ou lutter contre le développement des algues.

3.2.2 Classement par groupe chimique

Il s'agit d'un classement technique à partir de la molécule principale utilisée. On distingue :

- ✓ Les organochlorés, pesticides de synthèse à base de chlore. Ils sont parmi les plus anciens et les plus persistants. Ils sont surtout utilisés comme insecticides en agriculture.
- ✓ Les organophosphorés, eux aussi utilisés comme insecticides. Ils ont été développés à partir des années 1970. Ces produits présentent une toxicité aiguë bien plus forte que les organochlorés.
- ✓ Les carbamates, utilisés comme fongicides et insecticides. Tout comme les organophosphorés, les carbamates sont extrêmement toxiques et donnent lieu à de nombreuses intoxications.
- ✓ Les phénox utilisés comme herbicides.
- ✓ Les organo-azotés, repérables par le suffixe "zine", principalement utilisés comme herbicides.
- ✓ Les urées, repérables par le suffixe "uron", utilisés comme herbicides et fongicides.

3.2.3 Classement par mode d'action

Ce troisième classement a été institué en tenant compte du mode d'action des pesticides. Sur le tableau 1.1, sont donnés à titre d'exemple, quelques modes d'action ainsi que leur signification.

3.3 RESISTANCE AUX PESTICIDES

En plus de l'empoisonnement de notre environnement et de notre nourriture, les pesticides constituent une grave menace pour le système de production alimentaire. En dépit du fait que

ces substances, produits chimiques miracles, aient réussi à augmenter la production alimentaire en réduisant les dommages des ravageurs, ils ont également favorisé l'apparition de souches d'insectes, de souches d'agents pathogènes de plantes, de souches de nématodes (vers parasites), ainsi que des souches de mauvaises herbes résistantes. Cette résistance est due au fait que les pesticides ne tuent jamais 100% d'une population d'organismes nuisibles. Comme conséquence, les survivants ont tendance à avoir une faible sensibilité à ce type de produit. Par ailleurs, la résistance à un pesticide confère souvent une résistance ou un développement plus rapide de résistance à toute une famille de pesticides apparentés (Levin, 2007).

Tableau 3.1 : Mode d'action de certains pesticides.

Pesticides	Mode d'action	Signification
Insecticides		
	Contact	agit lorsque l'insecte entre en contact avec le produit
	Ingestion	agit lorsque l'insecte se nourrit du produit
	Inhalation	agit lorsque l'insecte respire le produit.
Herbicides		
	Contact	agit sur les parties de la plante avec lesquelles il entre en contact.
	Systémique	absorbé par la plante, se déplace à l'intérieur de celle-ci.
	Résiduaire	se dégrade lentement et contrôle les plantes sur une longue période
	Sélectif	ne contrôle que certaines plantes traitées.
	Non-sélectif	contrôle toutes les plantes traitées
Fongicides		
	Curatif	réprime une maladie qui est déjà développée
	Préventif	Protège la plante en empêchant que la maladie ne se développe.

3.4 TOXICOLOGIE DES PESTICIDES

La toxicologie est la science qui s'intéresse à la toxicité des substances et à leurs effets sur l'organisme. Par ailleurs, l'écotoxicologie évalue l'impact d'une substance sur l'environnement. Du fait de leurs éventuels effets néfastes, la nouvelle réglementation

REACH (Registration, Evaluation, Authorisation and restriction of Chemicals) oblige les entreprises industrielles de fournir des informations toxicologiques et écotoxicologique de toute substance chimique, les pesticides y compris, produites ou importées.Les informations toxicologiques ont trait entre autres, aux propriétés de la substance en matière d'irritation, de toxicité aigüe, de toxicité chronique, de toxicité pour la reproduction et de carcinogénicité. Les informations écotoxicologiques, quant à elles, se focalisent sur la toxicité aquatique, sur le devenir et le comportement dans l'environnement, sur les effets vis-à-vis des organismes terrestre et sur les oiseaux, …etc.

La toxicité d'une substance se distingue par deux formes. La première est appelée toxicité aigüe, alors que la seconde forme est désignée sous le vocable de toxicité chronique.

3.4.1 Toxicité aigüe

Elle correspond à la toxicité induite suite à l'administration d'une dose unique et massive d'un produit toxique. La principale conséquence de cette toxicité est la mort de l'organisme contaminé.

3.4.2 Toxicité chronique

Elle survient à la suite d'une administration de doses répétées d'une substance, ou d'expositions à celle-ci, au cours d'une période relativement longue allant de plusieurs jours à quelques années. Elle est le résultat du cumul et des sommations des effets d'une substance. Cette dernière, une fois ingérée, n'est pas éliminée. Elle s'accumule dans l'organisme jusqu'à atteindre une dose, à partir de laquelle les effets vont apparaitre. Ce genre de toxicité est souvent en lien avec la présence de résidus de pesticides dans l'alimentation, dans les eaux, …etc.

3.4.3 Evaluation expérimentale de la toxicité aigüe des pesticides

Pour évaluer la toxicité aigüe d'un pesticide, on utilise des indicateurs quantitatifs. Le plus utilisé est la Dose Létale 50 ou DL_{50} (LD_{50} en anglais pour Lethal Dose 50). Il a été inventé par J.W. Trevan en 1927 et permet de classifier tous les produits selon leur nocivité pour l'homme à court et à moyen termes (Poezevara, 2011). La DL_{50} correspond à la dose d'une substance pouvant causer la mort de 50 % d'une population animale dans des conditions d'expérimentation précises. On administre généralement le produit toxique à des rats ou à des souris répartis en plusieurs groupes. Cet indicateur est mesuré en milligrammes

de matière active par kilogramme d'animal. Plus ce chiffre est petit, plus la substance est toxique. Notons que chaque mesure repose sur un contexte expérimental bien précis, défini pour l'espèce étudiée, son sexe, le mode d'introduction dans l'organisme ou encore la durée d'exposition.

Lorsque la toxicité mesurée est semblable chez toutes les espèces d'animaux testées, l'hypothèse est émise qu'elle sera probablement semblable chez les humains. Lorsque les mesures sont différentes chez diverses espèces animales, des approximations et diverses hypothèses permettent d'estimer la dose mortelle probable chez l'homme (Poezevara, 2011).

3.5 UTILISATION DES PESTICIDES EN ALGERIE

3.5.1 Usage et importation de pesticides

L'Algérie est un pays à vocation agricole avec une superficie totale qui s'élève à 238 millions d'hectares dont 191 millions sont improductifs. La superficie agricole globale exploitée est estimée à 47 millions d'hectares, dont 32 millions d'hectares de parcours, 7 millions de forêts et de maquis et 8500 km^2 de terres agricoles utiles (FAO, 2014).

Dans notre pays, l'usage des produits phytosanitaires, des insecticides, des fertilisants, et autres se répand de plus en plus avec le développement de systèmes de culture «intensifs», fondés sur le recours aux pesticides et la recherche de hauts rendements. L'utilisation des pesticides et des engrais est relativement importante pour les cultures menées sous serre. Par ailleurs, la prolifération des moustiques dans toutes les agglomérations du pays oblige aussi les ménages à utiliser en abondance divers types d'insecticides. Le nombre de produits phytosanitaires commercialisés est de l'ordre de 400 produits dont une quarantaine de variétés sont largement utilisées par les agriculteurs (Bouziani, 2007).

La lutte antiacridienne a entraîné par ailleurs le déversement de milliers de tonnes d'insecticides. Par exemple, au cours de la campagne de lutte antiacridienne de 2004-2005, plusieurs tonnes de pesticides (des organophosphorés, des carbamates et la Deltaméthrine) ont été utilisées par épandage ou par pulvérisation dans les régions du Sud et du Tell infestées par le criquet pèlerin. Récemment, une autre forme d'utilisation intensive de pesticides se généralise dans de nombreuses wilayates du pays, dans le cadre du Programme national de lutte contre les zoonoses et plus particulièrement dans la lutte contre une pathologie

dénommée "leishmaniose", une maladie qui a pris de l'ampleur dans toutes les régions du pays (Bouziani, 2007).

Si l'on se réfère aux statistiques de l'organisation des nations unies pour l'alimentation et l'agriculture (FAO, 2014) en ce qui concerne les importations de pesticides entre les années 2000-2012 (Figure 3.1), on remarque que notre pays se situe en position médiane vis-à-vis de la Tunisie et du Maroc. Les importations de ce dernier pays sont de loin les plus importantes. D'autre part, les importations de l'Algérie suivent une courbe ascendante à partir de l'année 2010.

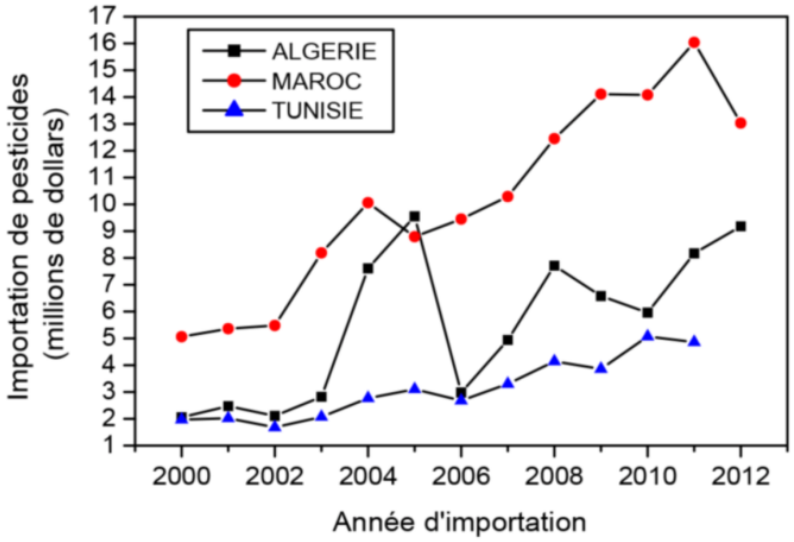


Figure 3.1 : Evolution des importations de pesticides pour les trois pays du Maghreb (FAO, 2014).

Par ailleurs, en regardant de près la part de chaque type de pesticides dans les importations Algériennes, on s'aperçoit que le gros de ces importations est à l'actif des insecticides. La part réservée aux herbicides et aux fongicides est presque la même avec quelquefois un léger avantage pour les fongicides (Figure 3.2).

3.5.2 Contamination par les pesticides

L'utilisation excessive des pesticides est synonyme d'une pollution massive des milieux naturels (sols, eaux superficielles, nappes souterraines, …etc.) par les pesticides et leurs résidus et ce dans toutes les régions du pays. Compte tenu du manque de données et/ou de publications se rapportant aux impacts des pesticides d'une manière générale et vue que les analyses des résidus de pesticides pour évaluer le degré de contamination des cultures et des

eaux superficielles ne sont pas faites systématiquement, on résumera les quelques informations qu'on a pu trouver.

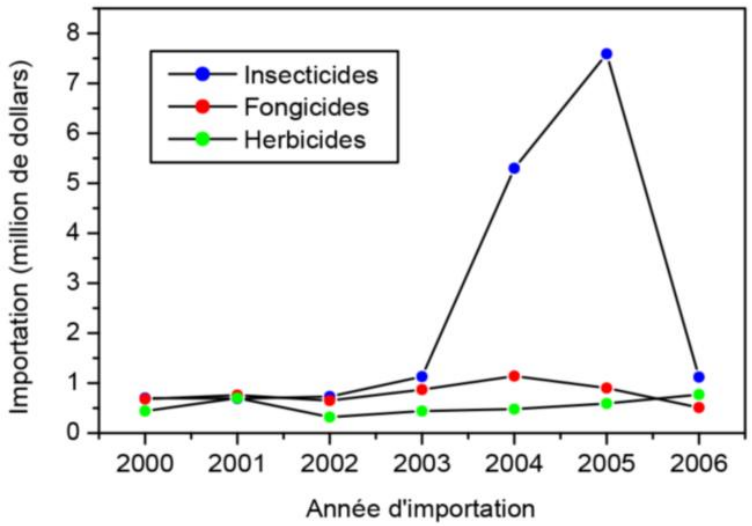


Figure 3.2 : Evolution des importations de fongicides, d'herbicides et d'insecticides en Algérie (FAO, 2014).

En premier lieu, notons que quelques dizaines de cas de contaminations accidentelles des manipulateurs par les pesticides au cours des opérations de lutte antiacridienne ont été enregistrés en mai 2004 (Bouziani, 2007). Des analyses effectuées sur des échantillons d'eau prélevés dans la région de Staoueli (Alger) ont montré que dans plus de 30% des échantillons, la concentration de certaines molécules organophosphorés (diazinon, parathion) et organochlorées (lindane, DDT, DDE) dépasse les valeurs limites préconisées par l'OMS (Moussaoui et al. 2001). Ces mêmes auteurs ont relevé la présence de résidus de pesticides dans des échantillons de laitue et de tomates.

En second lieu, certains faits inédits ont été rapportés (Bouziani, 2007). Des corbeaux ont attaqué à Ain Defla un champ de melons, nourriture qui ne fait pas partie de leur alimentation traditionnelle. Les fourmis dévastaient des plantations de tomates et d'abricots. Dans la même région, un fellah raconte que les escargots ont détruit un champ de pomme de terre. Ce comportement inhabituel de certains animaux, oiseaux et insectes trouve son explication, selon les spécialistes, dans le bouleversement du milieu naturel de ces espèces suite à l'utilisation immodérée de toutes sortes de pesticides. En conséquence, cet état de fait les pousse à chercher de nouvelles sources d'alimentation suite à la disparition de leur propre système alimentaire.

Chapitre 4

MODELISATION DE LA PREDICTION DE L'ACTIVITE TOXICOLOGIQUE DES PESTICIDES

Sommaire

4.1 Matériel et méthodes 58
4.2 Prediction de la toxicité des herbicides 69
4.3 Prediction de la toxicité des Insecticides 87
4.4 Prediction de la toxicité des Fongicides 96
4.5 Prediction de la toxicité des Pesticides 109
4.6 Comparaison entre modèles 126

Ce chapitre traite de la prédiction de l'activité toxicologique des pesticides. Pour ce faire, on commencera par une première partie intitulée "Matériels et méthodes». Cette dernière comportera trois axes. Dans le premier axe, on relatera les diverses étapes de l'élaboration de la base de données nécessaire au fonctionnement des méthodes d'apprentissage statistique qu'on utilisera. Cette base de données doit rassembler à la fois les valeurs de la sortie escomptée ainsi que les variables pertinentes (correspondantes) pouvant corréler avec elle. Dans notre cas, le paramètre à modéliser (c'est-à-dire la sortie) est inhérent à la toxicité orale aigüe pour un ensemble de pesticides. Pour les variables nécessaires à la description de cette toxicité, nous avons opté pour l'utilisation des descripteurs obtenus par calcul moyennant l'utilisation du logiciel Dragon 1.0. La méthode permettant de sélectionner

les descripteurs pertinents qui constitueront la base de données revêt une importance capitale car elle conditionne la qualité du modèle qui sera obtenu.

Le second axe sera consacré aux techniques les plus courantes pour établir des modèles QSAR : la régression linéaire multiple (RLM ou MLR en anglais) et les réseaux de neurones artificiels (RNA). Le troisième et dernier axe traitera du développement des modèles QSAR. Il sera question de l'ensemble des étapes permettant l'élaboration de deux modèles linéaire (par RLM) et non linéaire (par RNA) de prédiction de la toxicité orale aigüe. Pour estimer les performances des deux modèles, nous utiliserons un certain nombre d'indicateurs statistiques. Pour estimer les performances d'interpolation des modèles, on utilisera un ensemble de pesticides qui n'ont pas été inclus dans la base de données ayant servi à l'apprentissage et à la validation.

La seconde partie de ce chapitre se veut une application à l'élaboration de modèles QSAR pour la prédiction de la toxicité aigüe de pesticides. On y discutera des résultats obtenus et on les comparera aux quelques résultats disponibles dans la littérature.

4.1 MATERIEL ET METHODES

Deux étapes sont les plus couramment pratiquées dans le développement d'un modèle QSAR. La première a trait à la préparation des données, qui comprend la collecte et le nettoyage de données, le calcul de descripteurs moléculaires pour les produits chimiques ciblés et le prétraitement de la base de données. La deuxième étape est celle de la génération du modèle, qui permet l'établissement de relations statistiquement significatives entre la propriété ciblée (toxicité par exemple) et les valeurs des descripteurs.

4.1.1 Collecte des données de la toxicité orale aigüe

La DL_{50} (Dose Létale) est une façon de mesurer le potentiel toxique à court terme (toxicité aiguë) d'une matière. La DL_{50} est définie comme la quantité d'une matière (pesticide par exemple) administrée en une seule fois, qui cause la mort de 50 % (la moitié) d'un groupe d'animaux d'essai.

La toxicité orale aiguë chez le rat de 416 pesticides (répartis en 82 fongicides, 126 herbicides, 200 insecticides, 5 rodenticides, 2 acaricides et 1 bactéricide) appartenant à des classes chimiques différentes (Annexe A) exprimées en dose létale (DL_{50} - mmol / kg) ont été recueillies à partir des données sélectionnées à partir des sources suivantes : the Pesticide

Properties Data Base,PPDB–University of Hertfordshire (http:// www.herts.ac.uk/aeru/projects/ppdb/index.htm*)*; International Programme of Chemical Safety, IPCS (http:// *www.inchem.org/*); the Pesticide Action Network, PAN (*http:// www.pesticideinfo.org/*) and the European Centre for Ecotoxicology and Toxicology of Chemicals, ECETOC (http:// *www.ecetoc.org/chemical-databases*). Sur les figures 4.1 et 4.2 ci-dessous, nous donnons la capture d'écran de deux des sites utilisés.

Lors de la collecte de ces toxicités, il est extrêmement important de veiller à ne sélectionner que les données expérimentales exactes, précises et cohérentes. Aussi, nous avons tenu à ce qu'ils s'agissent de celles concernant les rats appartenant à la même souche et obtenues selon le même protocole expérimental.

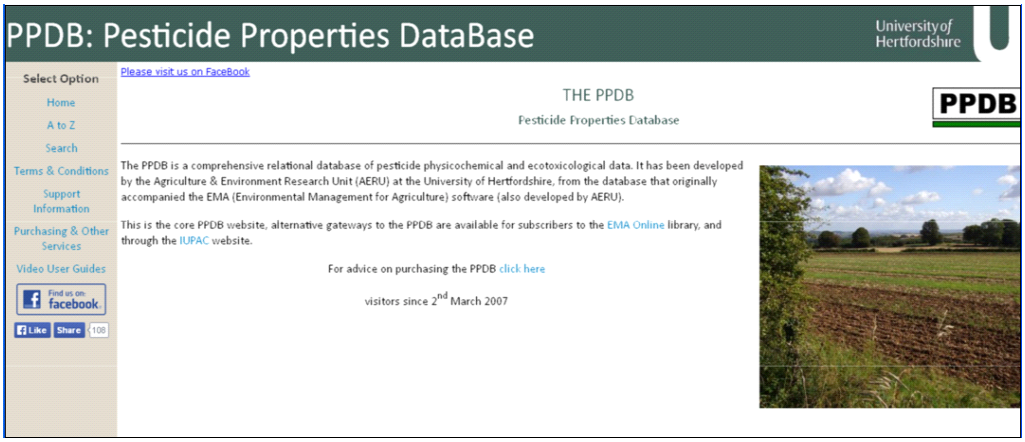


Figure 4.1 : Capture d'écran du site intitulé PPDB (*sitem.herts.ac.uk/aeru/projects/**ppdb**/)*



Figure 4.2 : Capture d'écran du site intitulé PAN (*www.**pesticide**info.org/)*

4.1.2 Obtention des descripteurs

La représentation numérique de la structure moléculaire a été évaluée en termes de descripteurs moléculaires. L'écriture SMILES nécessaire au calcul des descripteurs a été tirée à partir du site sur les propriétés des pesticides PPDB – University of Hertfordshire (http://*www.herts.ac.uk/aeru/projects/ppdb/index.htm*). Ces écritures ont été utilisées pour générer 1664 descripteurs répartis en vingt catégories: (i) constitutional descriptors; (ii) topological descriptors; (iii) walk and path counts; (iv) connectivity indices; (v) information indices; (vi) 2D autocorrelations; (vii) edge adjacency indices; (viii) Burden eigenvalue descriptors; (ix) topological charge indices; (x) eigenvalue-based indices; (xi) Randic molecular profiles; (xii) geometrical descriptors; (xiii) RDF descriptors; (xiv) 3D-MoRSE descriptors; (xv) WHIM descriptors; (xvi) GETAWAY descriptors; (xvii) functional group counts; (xviii) atom-centered fragments; (xix) charge descriptors; and (xx) molecular properties. Tous les descripteurs ont été obtenus par le biais des programmes en ligne E-Dragon 1,0 (http://www.vcclab.org). Sur le tableau 4.1 ci-après, est répertorié le nombre de descripteurs obtenus par catégorie.

4.1.3 Prétraitement de la base de données

Le prétraitement de la base de données a pour finalité d'éliminer les descripteurs qui ne sont pas pertinents afin d'éviter le phénomène du sur-apprentissage. Il s'agit alors de la réduction de la dimension de l'espace des variables (descripteurs) du modèle avec l'élimination des entrées qui n'ont pas ou possèdent peu d'influence sur la sortie (toxicité). Plusieurs méthodes pour simplifier la base de données sont disponibles dans la littérature. Par exemple, l'analyse en composantes principales ou curviligne (PCA / CCA) ou la méthode d'orthogonalization de Graam-schmidt sont utilisées. Au cours de notre étude, nous avons utilisé une méthode décrite par certains auteurs (Konovalov *et al.*, 2008; Srinivas Reddy *et al.*, 2010) tout en apportant des modifications au processus de sélection des descripteurs importants. Elle a lieu en plusieurs étapes, décrites ci-dessous :

1. On calcule le minimum et le maximum pour chaque descripteur à l'aide d'un programme MATLAB, puis on enlève les descripteurs qui ont le maximum et le minimum égaux.
2. On élimine le descripteur qui possède la même valeur pour plus de 75% des échantillons.

3. On calcule la déviation standard relative (RSD) pour chaque descripteur et on élimine ceux ayant des valeurs de RSD inférieures à 0.05.
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.
5. On utilise un programme basé sur la méthode stepwise pour la sélection des descripteurs les plus pertinents parmi ceux obtenus à l'étape 4.

Tableau 4.1 : Nombre de descripteurs calculés (http://www.vcclab.org)

Catégorie de descripteurs	Nombre par catégorie
Constitutional descriptors	48
Topological descriptors	119
Walk and path counts	47
Connectivity indices	33
Information indices	47
2D autocorrelations	96
Edge adjacency indices	107
Burden eigenvalue descriptors	64
Topological charge indices	21
Eigenvalue-based indices	44
Randic molecular profiles	41
Geometrical descriptors	74
RDF descriptors	150
3D-morse descriptors	160
WHIM descriptors	99
GETAWAY descriptors	197
Functional group counts	154
Atom-centred fragments	120
Charge descriptors	14
Molecular properties	29
Total	**1664**

Par ailleurs, les descripteurs sélectionnés ne doivent pas être multicolinéaires. Pour détecter les multicolinéarités et identifier les descripteurs qui y sont impliqués, on détermine soit leur coefficient de corrélation R^2 ou le paramètre VIF (Variance Inflation Factor) qu'on calcule selon l'équation 4.1 suivante :

$$VIF = \frac{1}{1 - R^2} \qquad (4.1)$$

La valeur de ce paramètre doit se situer dans la gamme de valeur de 1 à 5 (Confais et Le Guen, 2006). La valeur du facteur d'inflation de la variance (VIF) est calculée par le logiciel XLSTAT utilisé à cet effet.

4.1.4 Base de données finale

La structure de la base de données finale est schématisée ci-après sur la figure 4.3.

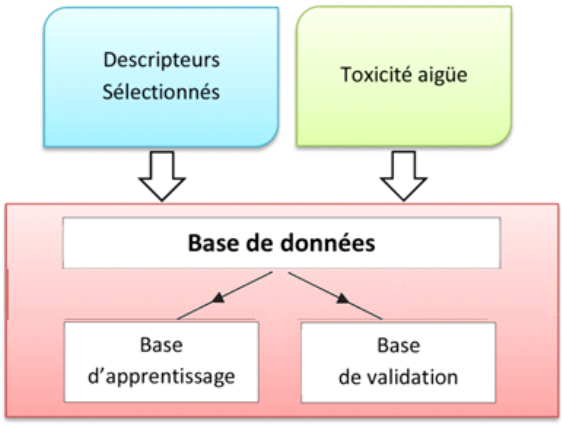


Figure 4.3 : Schématisation générale et répartition du contenu de la base de données

4.1.5 Développement des modèles QSAR

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

4.1.5.1 La régression linéaire multiple

La régression linéaire multiple est un outil statistique dont l'objectif est d'établir une relation linéaire entre une propriété d'intérêt (toxicité par exemple) et un ensemble de variables (descripteurs dans notre cas) et ce suivant l'équation 4.2 suivante :

$$\text{Toxicité} = a_0 + \sum_{i=1}^{n} a_i d_i \qquad (4.2)$$

où d_i représentent les descripteurs moléculaires, a_i leur coefficient et a_0 est une constante (appelé intercept).

4.1.5.2 Les réseaux de neurones artificiels

Les Réseaux de Neurones Artificiels (RNA) sont particulièrement bien adaptés pour les modèles QSAR / QSPR en raison de leur capacité à extraire des informations non-linéaires présentes dans la matrice de données (Yangjeh et Jenagharad, 2009). Pour trouver un modèle non linéaire entre les descripteurs moléculaires sélectionnés selon la méthode pas à pas et la toxicité aiguë, l'utilisation de la méthode RNA est nécessaire et est devenue une technique de modélisation importante. Elle est largement utilisée pour les études de QSAR non linéaires et est un outil puissant pour construire des modèles prédictifs. Le type de réseaux de neurones retenu pour ce travail est le perceptron multicouche (PMC). Le perceptron multicouche est un des réseaux de neurones les plus utilisés pour des problèmes d'approximation, de classification et de prédiction.

4.1.6 Estimation de la performance d'un modèle

L'objectif principal de toute modélisation QSAR est que le modèle développé doit être suffisamment robuste pour être capable de faire des prévisions fiables sur les activités biologiques (toxicité par exemple) de nouveaux composés qui n'ont pas été testés biologiquement. Ainsi, il est important de déterminer la capacité du modèle QSAR développé pour prédire les activités des composés qui ne sont pas présents dans l'ensemble d'apprentissage. Tropsha et al. (2003) soulignent d'ailleurs l'importance de la validation rigoureuse, une partie intégrante essentielle de l'élaboration de modèles QSAR. Pour une

évaluation réaliste du véritable pouvoir prédictif des modèles établis lors de ce travail, trois axes de travail ont été adoptés :

1. la stabilité et la robustesse ont été validées quantitativement en utilisant la validation croisée avec un certain nombre de paramètres statistiques.
2. Le domaine d'applicabilité de ce modèle a été défini.
3. Le pouvoir prédictif de ce modèle a été testé par utilisation du modèle mathématique obtenu.

4.1.6.1 Validation croisée

L'une des méthodes couramment utilisée pour déterminer la capacité prédictive d'un modèle QSAR (Wold et Eriksson, 1995 ; Pigram et McDonald 2001) est une validation interne appelée validation croisée. Cette méthode a pour but de tester les performances de généralisation et ce à partir des produits qui ont été utilisés pour élaborer le modèle. Pour ce faire, une certaine proportion d'observations est affectée à l'ensemble d'apprentissage et les observations restantes sont affectées à l'ensemble de validation. L'ensemble d'apprentissage a pour but de minimiser l'erreur et ensuite vérifier avec l'ensemble de validation après chaque itération afin d'éviter un sur-apprentissage et une perte de la capacité de généralisation.

4.1.6.2 Paramètres statistiques

En plus de cette validation croisée, un certain nombre d'outils statistiques (paramètres) sont requis pour évaluer la performance du modèle élaboré.

- **Le coefficient de determination multiple ajusté** ($R^2_{adjusté}$) :

$$R^2_{adj} = 1 - \frac{(1 - R^2)(n - 1)}{n - k - 1} \qquad (4.3)$$

plus ($R^2_{adjusté}$) est proche de 1**,** plus la corrélation entre les valeurs expérimentales et prédites est meilleure.

- **Le coefficient de prédiction** (Q^2_{LOO}) :

Afin d'estimer la véritable puissance prédictive d'un modèle QSAR, on doit comparer les activités prédites et observées (expérimentales) d'un assez grand ensemble de test externe de

composés qui n'ont pas été utilisés dans l'élaboration du modèle (Golbraikh et Tropsha., 2002). C'est la précision de la performance sur cet ensemble de test qui détermine le réel pouvoir de prédiction d'un modèle QSAR. Le pouvoir de prédiction d'un tel modèle peut être estimé par le coefficient Q^2_{LOO} calculé selon l'équation suivante :

$$Q^2_{LOO} = 1 - \frac{\sum_{i=1}^{n}(y_{exp} - y_{pred})^2}{\sum_{i=1}^{n}(y_{exp} - \bar{y})^2} \qquad (4.4)$$

Une valeur de Q^2_{LOO} > 0.5 est généralement considérée comme satisfaisante, une valeur supérieure à 0.9 est excellente.

En plus des 2 paramètres cités ci-dessus, il est vivement recommandé pour juger de la qualité de prédiction du modèle, de considérer la racine de l'écart quadratique moyenne (voir équations 4.5) calculée sur l'ensemble de validation. Ces valeurs de REQM sont mieux adaptées pour juger de la qualité d'un modèle que les valeurs de R^2_{adj} et Q^2_{LOO} seules.

- **Ensemble de validation croisée (appelée SDEP)**.

$$SDEP = \sqrt{\frac{\sum_{i=1}^{n}(y_{exp} - y_{pred})^2}{n}} \qquad (4.5)$$

Pour les équations 4.3 à 4.5

n est le nombre d'observations (ou de molécules utilisées) de chaque ensemble.

k est le nombre de descripteurs sélectionnés.

$y_{exp}, y_{pred}, \bar{y}$ sont respectivement la valeur expérimentale, la valeur estimée ou prédite et la valeur moyenne des valeurs expérimentales du paramètre à prédire (la toxicité dans notre cas).

4.1.6.3 Domaine d'applicabilité

Même les modèles les plus exhaustifs et validés, ne peuvent prédire des propriétés de manière fiable pour l'intégralité des composés existants. Aussi, les domaines d'applicabilité (AD) des modèles doivent-ils être définis et seules les prédictions concernant des molécules dans ce domaine peuvent être considérées comme recevables. En effet, un modèle QSPR n'est pas destiné à être employé en dehors de son domaine d'applicabilité, c'est-à-dire en dehors de

l'espace chimique couvert par son jeu d'entraînement. La détermination des AD est donc d'une grande importance (OCDE, 2009).

Le domaine d'applicabilité d'un modèle a été récemment (Netzeva et al, 2005) défini comme suit: "Le domaine d'applicabilité d'un modèle QSAR est la portion de l'espace dans laquelle le modèle fait des prédictions avec une fiabilité donnée. Ainsi, le domaine d'applicabilité est une région théorique dans l'espace défini par la réponse modélisée et les descripteurs du modèle, pour laquelle un modèle QSAR donné devrait faire des prévisions fiables. Cette région est définie par la nature des produits chimiques dans l'ensemble d'apprentissage et de validation. Dans ce domaine, les lignes droites horizontales et verticales indiquent les limites des valeurs normales: la première pour les valeurs aberrantes et la deuxième pour les produits chimiques influents.

Le domaine d'applicabilité du modèle QSAR est défini à partir du diagramme de Williams (plot Williams) reliant les leviers "h_{ii}" (leverages) des produits (calculés par le logiciel STATISTICA) en fonction des résidus normalisés "d_i" (équation 4.8). La valeur du levier critique h* est donnée par l'équation 4.9 ci-dessous.

$$d_i = \frac{y_{exp} - y_{pred}}{s} = \frac{y_{exp} - y_{pred}}{\sqrt{\frac{\sum_{i=1}^{n}(y_{exp}-y_{pred})^2}{n-k-1}}} \qquad (4.8)$$

$$h^* = \frac{3(k+1)}{n} \qquad (4.9)$$

où *n*, *k* et *s* sont respectivement le nombre d'observation (ou produits) de l'ensemble d'apprentissage ou de validation, le nombre de descripteurs et l'écart type.

4.1.6.4 Utilisation de l'expression mathématique du modèle

Dans ce cadre, il s'agit de tester la performance du modèle neuronal optimisé et ce par l'application de l'équation mathématique décrivant ce modèle. Autrement dit, il s'agit de tester les capacités de généralisation du modèle obtenu en l'appliquant à des situations où la valeur des différentes variables est bien comprise entre les bornes minimales et maximales des exemples ayant servi à l'apprentissage du modèle.

4.1.7 Analyse de la contribution des descripteurs

L'importance de chaque variable doit être déterminée de manière à confirmer l'influence des paramètres sélectionnés en entrée sur la prédiction de la toxicité des pesticides. Cependant, la fonction implémentée par le réseau étant complexe, l'analyse de la contribution de chaque variable est donc extrêmement difficile. Ainsi, diverses méthodes ont été mises en œuvre pour caractériser la "boîte noire" de type RNA (Gevrey et al., 2003). Ces méthodes, de nature quantitative, permettent d'analyser la contribution des variables d'entrée sur la prédiction des réseaux neuronaux. De leur côté, Olden et al. (2004) ont évalué la pertinence de chacune de ces méthodes pour estimer la contribution des variables. Ils montrèrent alors que l'une d'entre elles, appelée "approche des poids de connexions" est le plus souvent la plus précise et la plus fiable.

L'approche des poids de connexions (Connection Weight Approach) utilise les poids de connexions bruts des connexions entrée-cachée (Input-Hidden : IH) et cachée-sortie (Hidden-output : HO) d'un RNA. Selon (Olden et al., 2004), elle fournit la meilleure méthodologie pour une identification précise de l'importance des paramètres du modèle. Cette approche permet donc d'identifier avec succès l'importance réelle de toutes les variables du réseau de neurones. Pour ce faire, on calcule le résultat des connexions brutes des neurones entrée-cachée et cachée-sortie pour chaque neurone d'entrée et le neurone de sortie [P (i, j)] et ensuite on détermine l'importance de chaque variable [I(i)] par la sommation des résultats à travers tous les neurones cachés (Olden et Jackson, 2002) et ce suivant les équations 4.6 et 4.7 suivantes :

$$P(i,j) = IH(i,j) * HO(j) \tag{4.6}$$

$$\mathrm{I(i)} = \sum_{j=1}^{m} P(i,j) \tag{4.7}$$

avec i : nombre de variables d'entrée

j : nombre de neurones cachés

Une variable d'entrée avec des poids de connexions forts est significative d'une forte contribution sur la prédiction du réseau par rapport aux autres variables caractérisées par des poids de connexions plus petits.

4.1.8 Outils informatiques utilises

Au cours de ce travail, nous avons eu à utiliser plusieurs logiciels (voir capture d'écran sur la figure 4.4 ci-dessous) :

- STATISTICA (STATISTICA 7.0, Tulsa, StatSoft Inc., OK, États-Unis)
- MATLAB R2010b
- XLSTAT 2012



Figure 4.4 : Capture d'écran des logiciels STATISTICA et MATLAB

4.2 PREDICTION DE LA TOXICITE DES HERBICIDES

Cette première partie de notre travail a fait l'objet de la publication scientifique suivante:

Mabrouk Hamadache, Latifa Khaouane, Othmane Benkortbi, Cherif Si Moussa, Salah Hanini, and Abdeltif Amrane (2014). ***Prediction of Acute Herbicide Toxicity in Rats from Quantitative Structure–Activity Relationship Modeling.*** *ENVIRONMENTAL ENGINEERING SCIENCE. Volume 31, Number 5,243-252.*

Elle a été consacrée à la prédiction de la toxicité orale aigüe chez le rat d'une série d'herbicides. La première étape a été la sélection des descripteurs pertinents. On présentera l'ensemble des étapes permettant l'élaboration de deux modèles, linéaire (par MLR) et non linéaire (par RNA) de prédiction de la toxicité orale aigüe. Pour le modèle MLR, on discutera de son développement et on donnera l'expression mathématique obtenue. En ce qui concerne le modèle relatif au RNA, on traitera tour à tour le découpage de la base de données, l'apprentissage et l'optimisation du réseau de neurones, la validation du modèle neuronal, l'expression mathématique ainsi que la contribution de chaque descripteur. Pour estimer les performances des deux modèles, nous utiliserons un certain nombre d'indicateurs statistiques. Pour estimer les performances d'interpolation des modèles, on utilisera un ensemble de pesticides qui n'ont pas été inclus dans la base de données ayant servi à l'apprentissage et à la validation. Une comparaison avec les quelques modèles rapportés dans la littérature clôturera ce chapitre.

4.2.1 Résultats et discussion

Nous avons sélectionné 62 composés de la famille des herbicides. Leur toxicité orale aigüe sur les rats (en mg/Kg), leur masse molaire et leur écriture SMILES ont été tirées des sites internet indiqués au paragraphe 4.1 de ce présent chapitre. L'unité de la toxicité a ensuite été convertie en mmol/Kg.

4.2.1.1 Sélection des descripteurs

Les descripteurs sélectionnés par application de la méthode donnée au paragraphe 4.1.3 sont répertoriés sur le tableau 4.2 ci-dessous. Finalement 160 descripteurs ont été obtenus. Pour la sélection des descripteurs pertinents, la régression multiple pas à pas

(Stepwise MLR) a ensuite été utilisée (Xu et Zhang, 2001). Dans cette procédure une variable qui est utilisée dans le modèle dans les premières étapes de sélection peut être supprimée dans la suite des étapes. L'addition graduelle d'autres descripteurs se poursuit pour trouver les meilleurs modèles de régression multi-paramètres avec les valeurs optimales des critères statistiques (valeurs les plus élevées du coefficient de détermination). Au final, 18 descripteurs ont été sélectionnés. Comme le nombre d'herbicides utilisé est de 62, il est important de réduire le nombre de descripteurs jusqu'à ce que le rapport "nombre d'herbicides / prédicteurs" soit égal ou supérieur à 5 (Huynh, 2007).

Tableau 4.2 : Etapes de sélection des descripteurs pertinents.

Descripteurs	N_{avp}	N_{app}	$N_{obt.}$
Constitutional Descriptors	48	8	1
Topological Descriptors	119	10	0
Walk and Path counts Descriptors	47	1	0
Connectivity Indices	33	5	0
Information Indices	47	8	0
2D Autocorrelations	96	19	3
Edge Adjacency Indices	107	2	0
Burden Eigen values	64	5	0
Topological Charge Indices	21	3	1
Eigenvalue based Indices	44	3	0
Randic Molecular Profiles	41	1	0
Geometrical Descriptors	74	7	1
RDF Descriptors	150	14	2
3D MoRSE Descriptors	160	16	1
WHIM Descriptors	99	10	4
Getaway Descriptors	197	32	3
Functional Group Counts	154	4	0
Atom centered fragments	120	7	2
Charge descriptor	14	0	0
Molecular Properties	31	5	0
Total	**1664**	**160**	**18**

N_{avp} : Nombre de descripteurs avant prétraitement
N_{app} : Nombre de descripteurs après prétraitement
$N_{obt.}$: Nombre de descripteurs obtenu après stepwise MLR

Pour la sélection des descripteurs les plus importants et pour sélectionner le nombre optimal, les influences du nombre de descripteurs sur les coefficients de corrélation (R^2, $R^2_{ajusté}$) et sur la racine de l'erreur quadratique moyenne (REQM) ont été étudiées pour les dix-huit descripteurs. Les résultats sont reproduits sur la figure 4.5. On remarque que les valeurs de R^2 et $R^2_{ajusté}$ augmentent avec l'augmentation du nombre de descripteurs. Cependant, les valeurs de la REQM diminuent avec l'augmentation du nombre de descripteurs. Aussi, au-delà de 12 descripteurs, on n'observe aucune amélioration significative de ces paramètres statistiques. Pour ces raisons, le nombre de descripteurs retenus pour développer le modèle est de douze.

Finalement, les 12 descripteurs retenus sont : MATS2p et MATS6m (2D Autocorrelations indices), C-013 et N-072 (Atom- centred fragments), DISPe (Geometrical Descriptors), HATSe et HATS0m (Getaway Descriptors) et Du, G1s, E1u, G3p (Whim Descriptors) et nS (Constitutional descriptors).

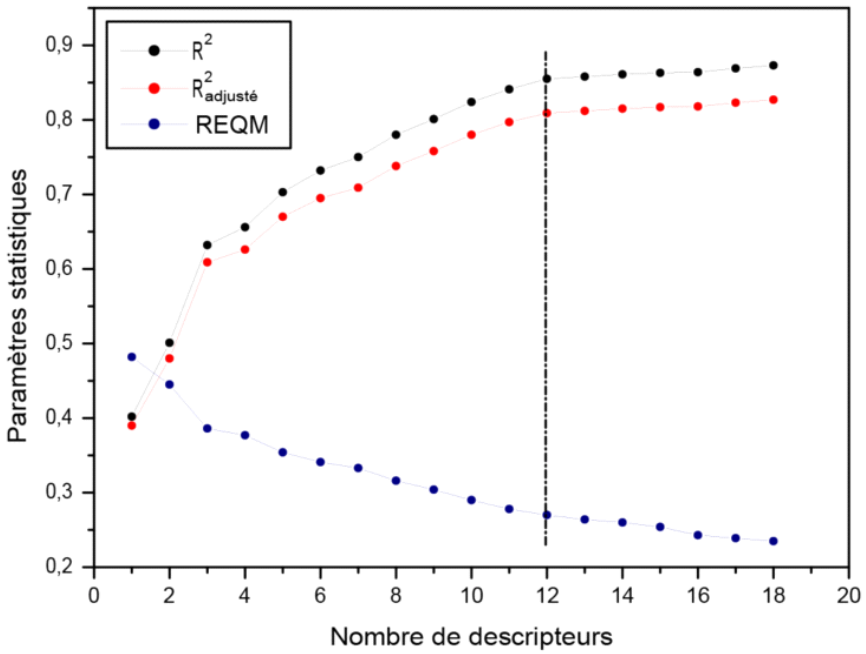


Figure 4.5 : Influence du nombre de descripteurs sur les paramètres statistiques R^2, $R^2_{ajusté}$ et REQM.

Pour statuer sur la relation entre les descripteurs, la matrice de corrélation de ces derniers a été établie et reproduite sur le tableau 4.3. Sur la base des résultats obtenus, on peut voir que la valeur du coefficient de corrélation la plus élevée de chaque paire de descripteurs choisis est de 0,619, ce qui signifie que les descripteurs retenus sont indépendants.

Tableau 4.3: Matrice des coefficients de corrélation des descripteurs sélectionnés.

	HATS0m	HATSe	nS	DISPe	N-072	MATS2p	C-013	Du	MATS6m	G1s	E1u	G3p
HATS0m	1.000	0.242	-0.236	0.084	-0.145	0.045	-0.026	-0.297	0.071	0.021	-0.152	-0.034
HATSe		1.000	-0.050	0.400	0.028	0.516	0.440	-0.441	-0.161	-0.396	-0.272	-0.273
nS			1.000	-0.001	0.095	-0.169	0.109	0.147	0.326	-0.163	0.032	-0.093
DISPe				1.000	0.213	0.224	**0.619**	-0.097	-0.088	-0.188	-0.079	-0.065
N-072					1.000	0.370	0.186	0.186	-0.092	-0.199	0.099	-0.235
MATS2p						1.000	0.202	-0.357	-0.265	-0.335	-0.162	-0.277
C-013							1.000	0.099	0.121	-0.079	0.091	-0.046
Du								1.000	0.152	0.392	0.582	-0.145
MATS6m									1.000	-0.068	0.107	-0.025
G1s										1.000	0.549	0.524
E1u											1.000	0.309
G3p												1.000

D'autre part, la multicolinéarité entre les douze descripteurs a été vérifiée par le calcul du paramètre VIF. Si le VIF se trouve dans la gamme de valeurs de 1 à 5, le modèle est acceptable. Comme on peut le voir sur le tableau 4.4, toutes les variables ont des valeurs du VIF inférieures à 3.956, ce qui indique que le modèle obtenu a une signification statistique (Confais et Le Guen, 2006). La valeur du facteur d'inflation de la variance est calculée par le logiciel XLSTAT.

4.2.1.2 Développement des modèles QSAR

Les 12 descripteurs sélectionnés ont été utilisés pour développer des modèles prédictifs. Deux approches différentes pour élaborer des modèles de prévision QSAR ont été utilisées. Le modèle linéaire a été développé par l'application de la régression linéaire multiple (RLM). La RLM est l'une des méthodes les plus couramment utilisées (Sarkhosh et al., 2012). Elle consiste en une relation quantitative entre un groupe de variables prédictives (X) et une réponse Y comme décrit dans le chapitre 2.

Par ailleurs, très souvent, les données biologiques (la toxicité dans notre cas) doivent être transformées avant d'être utilisées pour dériver un modèle QSAR. Ainsi, par exemple, les valeurs de la DL_{50} doivent être exprimées sur une base molaire. En outre, ces données doivent être converties en $1/DL_{50}$ ou en une échelle logarithmique (Devillers., 2013).

a) *Modèle RLM*

Le modèle obtenu pour la prédiction de la toxicité aiguë par voie orale des herbicides sur des rats, en utilisant 62 composés, est le modèle linéaire suivant (équation 4.8) avec les paramètres statistiques y afférant :

$$\frac{1}{DL_{50}} = 4.755 + 0.308[HATS0m] - 0.421[HATSe] + 0.420[nS] + 0.162[DISPe] - 0.248[N072] + 0.148[MATS2p] + 1.033[C013] - 4.090[Du] - 0.642[MATS6m] - 0.752[G1s] + 2.159[E1u] - 0.375[G3p] \quad (4.8)$$

N=62, R^2 =0.855, $R^2_{adj.}$ =0.809, $R^2_{prediction}$ = 0.880, RMSE=0.270, F=18.664, p < 0.0001

Les modèles QSAR sont généralement des fonctions de la structure d'une molécule, des propriétés électroniques et d'hydrophobicité (Sanderson et al., 2004). Dans notre modèle MLR, MATS6m, C-013, N-072, HATS0m, Du, G1, E1U implique la structure de la molécule. MATS2p, DISPe, HATSe, G3p représentent les propriétés électroniques (polarisabilité et électronégativité). Le descripteur nS implique le nombre d'atomes de soufre.

Afin de distinguer l'importance de chaque descripteur sur la toxicité des herbicides, des valeurs de t-test des descripteurs impliqués sont également répertoriés dans le tableau 4.4. Comme remarqué sur ce tableau, le descripteur Getaway (HATS0m) possède la plus grande valeur du t-test (11,394). Ceci indique que, dans ce modèle QSAR, l'influence de HATS0m sur la toxicité est la plus forte.

Tableau 4.4 : Valeurs du VIF, t-test et catégories des descripteurs sélectionnés.

Descriptor	**Category**	**VIF**	**t-test**
MATS2p	2D Autocorrelations indices	2.159	0.746
MATS6m	2D Autocorrelations indices	1.436	-3.238
C-013	Atom- centred fragments	2.391	4.166
N-072	Atom- centred fragments	1.548	-4.319
DISPe	Geometrical Descriptors	1.850	0.714
HATSe	Getaway Descriptors	2.616	-8.245
HATS0m	Getaway Descriptors	1.382	11.394
Du	Whim Descriptors	3.956	-3.202
G1s	Whim Descriptors	2.616	-1.938
E1u	Whim Descriptors	2.372	2.725
G3p	Whim Descriptors	2.626	-1.879
nS	Constitutional descriptors	1.303	6.130

Les signes des coefficients de régression indiquent la direction de l'influence des descripteurs dans un modèle donné. Selon les coefficients de régression de l'équation (4.8), il est à noter que les descripteurs WHIM (E1U, Du) sont ceux qui contribuent le plus à la toxicité. Les coefficients de régression des descripteurs E1u, C-013, nS et HATS0m présentent des signes positifs. Ces descripteurs ont fourni une influence positive et les valeurs les plus élevées sont favorables à l'amélioration de la toxicité des herbicides. Par ailleurs, les coefficients de régression des descripteurs Du, G1S, MATS6m, N-072 et HATSe présentent des signes négatifs, donnant ainsi un impact négatif sur la toxicité. Pour un herbicide donné, les petites valeurs de ces descripteurs aideraient à améliorer sa toxicité. Ainsi, une conclusion peut être émise de la façon suivante : la toxicité orale aiguë (DL_{50}) pour les rats est influencée principalement par trois facteurs importants : la structure de la molécule, les propriétés électroniques et le nombre d'atomes de soufre.

Les descripteurs utilisés dans notre modèle ont été utilisés dans des modèles QSAR rapportés dans la littérature. Dans une modélisation QSAR de toxicité aiguë DL_{50} pour les rats causée par des composés aromatiques, Rasulev (Bakhtiyor et al., 2010) a constaté que le descripteur MATS2p contribue de manière significative à la toxicité de ces composés. Dans une autre étude sur la pénétration de la barrière hématoencéphalique, l'absorption intestinale humaine et l' hydrophobicité, Soto (Soto et al., 2009) a proposé des modèles QSAR linéaire et QSPR non linéaires qui comprennent le descripteur MATS2p. Des modèles de haute qualité statistique interne et externe pour prédire la toxicité d'un ensemble de 28 composés ont été développés par Borges (Borges, 2012) avec le descripteur MATS2p. Le descripteur E1u a été employé dans un autre modèle QSAR sur la prédiction de la DL_{50} orale chez le rat de 58 produits chimiques polyfluorés développé par Bhhatarai et Gramatica (2011). Les auteurs ont conclu qu'E1u est l'un des descripteurs les plus importants.

Par ailleurs, certains auteurs (Xu et al., 2012 ; Tugcu et al., 2012 ; Duchowicz et al , 2009 ; . Hongying et al, 2008 ; Ning. Xin et al., 2010) ont constaté que parmi les descripteurs qui affectent la toxicité des composés étudiés, une grande partie appartient à la catégorie des descripteurs WHIM, descripteurs Getaway, descripteurs 2D autocorrélations et descripteurs Atome centered fragments. Pour notre étude, un grand nombre de descripteurs (67%) impliqués dans le modèle appartiennent également à ces 4 catégories. Il est évident que les descripteurs de ce type ont une importance majeure dans la toxicité des herbicides.

La toxicité aiguë par voie orale d'herbicides prédite selon le modèle RLM est reportée dans le tableau 4.5 ci-contre et sur la figure 4.6 ci-dessous.

Tableau 4.5 : Valeurs des toxicités aigües expérimentales, prédites par RLM et par RNA

Herbicide	CAS number	1/DL50 (mmol/Kg)$^{-1}$		
		Expérimentale	Prédite par RLM	Prédite par RNA
2,4,5-trichlorophenol*	95-95-4	0,24	0,32	0.30
2,4,5-trichlorophenoxyacetic acid	93-76-5	0,51	0,36	0,51
2,4-DB*	94-82-6	0,28	0.15	0.28
2,4-dichlorophenoxyacetic acid	94-75-7	0,47	0,45	0,42
2-naphthyloxyacetic acid	120-23-0	0,04	-0,03	-0,14
4-chlorophenoxy)acetic acid	122-88-3	0,22	0,33	0,30
Acetochlor	34256-82-1	0,14	0,12	0,23
Acifluorfen	50594-66-6	0,26	-0,22	0,25
Acrolein	107-02-8	1,93	1,80	1,64
Alachlor	15972-60-8	0,29	0,28	0,26
Allidochlor	93-71-0	0,60	0,32	0,51
Ametryn	834-12-8	2,07	1,00	2,09
Amicarbazone	129909-90-6	0,24	-0,06	0,06
Amiprofos-methyl	36001-88-4	0,98	0,72	0,83
Anilofos	64249-01-0	0,78	0,82	0,84
Barban*	101-27-9	0,49	0.36	0,47
Benazolin ethyl	25059-80-7	0,07	0,34	0,12
Bensulide	741-58-2	1,47	1,19	1,43
Bentazone	25057-89-0	0,48	0,50	0,34
Benzthiazuron	1929-88-0	0,16	0,50	0,34
Bromacil	314-40-9	0,20	0,34	0,34
Bromofenoxim*	13181-17-4	0,42	0.40	0,70
Bromoxynil	1689-84-5	3,42	2,86	3,42
Bromoxynil heptanoate	56634-95-8	1,34	1,78	1,31
Bromoxynil octanoate*	1689-99-2	1,69	1,25	1.78
Butachlor	23184-66-9	0,16	0,08	0,12
Butamifos	36335-67-8	0,53	0,78	0,68
Buthidazole*	55511-98-3	0,17	0.01	0,30
Butralin*	33629-47-9	0,28	0.36	0,16
Butroxydim	138164-12-2	0,24	0,20	0,30
Buturon	3766-60-7	0,13	0,07	0,09
Butylate	2008-41-5	0,06	0,22	0,04
Carbetamide	16118-49-3	0,14	0,13	0,15
Chlorbromuron	13360-45-7	0,14	0,14	0,23
Chlorbufam	1967-16-4	0,09	0,32	0,15
Chlorfenac	85-34-7	0,13	0,28	0,18
Ethanedial	557-30-2	0,49	0,46	0,49
Ethoxysulfuron	126801-58-9	0,12	0,21	0,12
Ethyl dipropylthiocarbamate	759-94-4	0,21	0,17	0,16
Glufosinate	51276-47-2	0,11	0,01	0,14

Tableau 4.5 (suite)

Herbicide	CAS number	1/DL50 (mmol/Kg)$^{-1}$		
		Expérimentale	Prédite par RLM	Prédite par RNA
Methanecarboxylic acid*	758-12-3	0,02	-0,14	-0,03
Propachlor	1918-16-7	0,38	0,23	0,34
Propanil	709-98-8	0,23	0,45	0,25
Propazine	139-40-2	0,06	0,13	0,08
Propisochlor*	86763-47-5	0,12	0.16	0,11
Prosulfocarb	52888-80-9	0,14	0,36	0,10
Prosulfuron	94125-34-5	0,43	0,41	0,46
Prynachlor	21267-72-1	0,19	0,32	0,22
Pyrazoxyfen	71561-11-0	0,25	0,11	0,12
Quinclorac	84087-01-4	0,09	0,10	0,10
Quizalofop-ethyl	76578-14-8	0,26	0,27	0,25
Quizalofop-P-ethyl	100646-51-3	0,32	0,14	0,13
Quizalofop-P-tefuryl*	119738-06-6	0,42	0.14	0,24
Sethoxydim	74051-80-2	0,12	0,73	0,25
Simetryn	1014-70-6	0,28	0,76	0,29
S-metolachlor	87392-12-9	0,11	-0,01	0,08
Tebutam	35256-85-0	0,04	0,04	0,13
Tebuthiuron	34014-18-1	0,35	0,20	0,37
Tepraloxydim	149979-41-9	0,07	0,15	0,18
Thiazafluron	25366-23-8	0,86	0,53	0,87
Tralkoxydim	87820-88-0	0,35	0,28	0,35
Triclopyr butotyl*	64700-56-7	0,62	0,35	0,59

** pesticides de l'ensemble de validation*

Comme on peut le constater (tableau 4.5 et figure 4.6), les valeurs prédites pour $1/DL_{50}$ par la RLM sont en bon accord avec les valeurs expérimentales. Plus que cela, la grande valeur du coefficient F, qui est de 18,664 signifie que l'équation (4.9) est significative pour prédire les valeurs de $1/DL_{50}$. Par ailleurs, le modèle présente une valeur du coefficient $R^2_{ajusté}$ de 0,809, ce qui implique une bonne concordance entre les valeurs prédite et expérimentales. D'autre part, les erreurs relatives pour les herbicides étudiés ont été examinées à l'aide des résidus normalisés par rapport aux valeurs prédites et observées. On peut noter que les résidus normalisés sont répartis de manière symétrique autour du zéro. Les caractéristiques statistiques obtenues dans cette étude, indique que le modèle RLM à un bon pouvoir prédictif de la toxicité orale aigüe des herbicides.

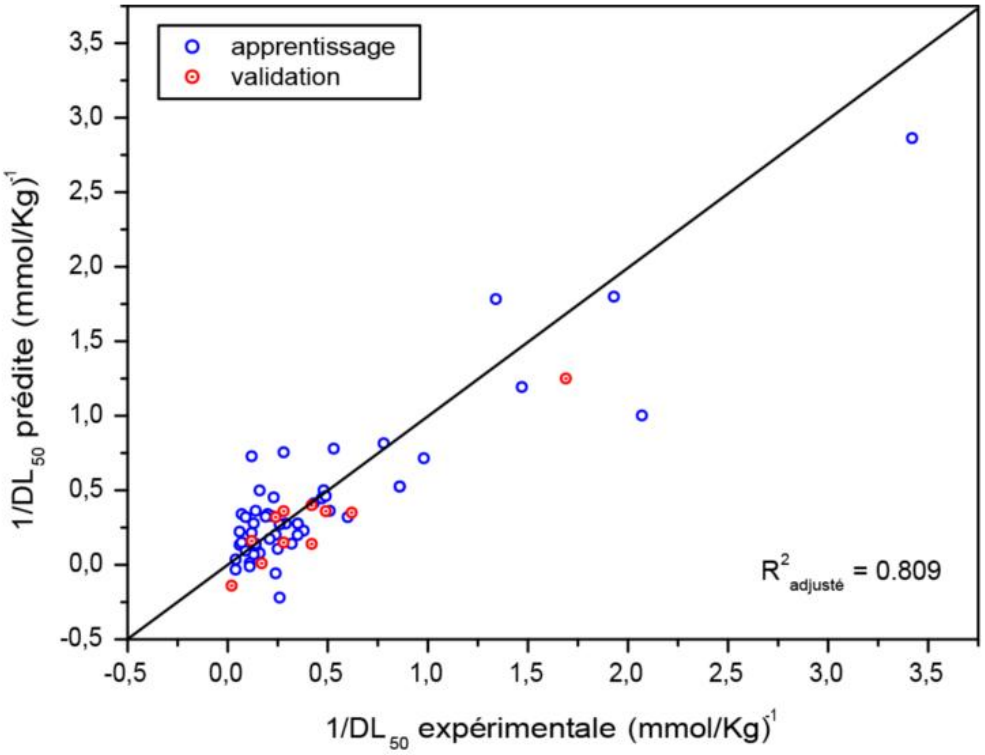


Figure 4.6 : Représentation graphique du modèle RLM de prédiction de la toxicité orale aigüe ($1/DL_{50}$) chez le rat d'une série d'herbicides

b) Modèle RNA

Les Réseaux de Neurones Artificiels (RNA) sont particulièrement bien adaptés pour les modèles QSAR / QSPR en raison de leur capacité à extraire des informations non-linéaires présentes dans la matrice de données (Yangjeh et Jenagharad, 2009). Pour trouver un modèle non linéaire entre les descripteurs moléculaires sélectionnés selon la méthode pas à pas et la toxicité aiguë pour les rats, l'utilisation de la méthode RNA est nécessaire et est devenue une technique de modélisation importante. Elle est largement utilisée pour les études de QSAR non linéaires et est un outil puissant pour construire des modèles prédictifs. Le perceptron multicouche adopté comporte trois couches : une pour l'entrée, une couche cachée et une troisième comme couche de sortie pour la toxicité. Le nombre de neurones dans la couche d'entrée est égal au nombre de descripteurs utilisés dans le modèle MLR.

L'utilisation d'une régression neuronale passe par le choix des paramètres d'entrée mais aussi par l'optimisation de l'architecture du réseau de neurones lui-même. L'optimisation à la fois du type de fonction non linéaire et de la géométrie du réseau de neurones (nombre de neurones dans la couche cachée) a été faite après de nombreux essais empiriques.

Pour optimiser le nombre de neurones dans la couche cachée, plusieurs calculs ont été effectués avec différents nombres (de 1 à 30). Chaque réseau a ainsi été "entraîné" avec la fonction tangente hyperbolique comme fonction d'activation et la quasi-Newton BFGS comme algorithme d'apprentissage. Il convient de noter que la formation du réseau a été arrêtée une fois que l'erreur quadratique moyenne de validation (EQMV) commence à augmenter ou que le surentraînement a commencé.

Pour surveiller le surentraînement au cours de la phase d'apprentissage, les valeurs d'erreur quadratique moyenne ont été calculées. En fin de compte, c'est le modèle qui présente le meilleur résultat (la plus petite EQM) qui est sélectionné. Le jeu de données a été divisé en deux ensembles : un ensemble d'apprentissage et un ensemble de validation composé respectivement de 51 et 11 herbicides. L'ensemble d'apprentissage est utilisé pour générer le modèle et l'ensemble de validation (ou de prédiction), formé d'herbicides qui n'ont pas été utilisés dans la procédure de modélisation a été utilisé pour évaluer le modèle obtenu. Ce réseau est composé de douze entrées, les mêmes descripteurs que dans le modèle de MLR et d'une sortie pour $1/DL_{50}$. Aussi, un modèle RNA a été généré avec l'architecture figurant sur le tableau 4.6

Tableau 4.6 : Architecture du réseau de neurones optimisé

	Critère retenu pour cette étude
Architecture du Perceptron Multi-Couche (PMC)	Neurones d'entrée : 12 Neurone de sortie : 01 Neurones cachés : 14 (le plus petit nombre produisant la plus faible erreur de prédiction sur l'ensemble de validation
Fonction d'activation	Tangente hyperbolique (pour la couche cachée) Purelin (pour la couche de sortie)
Algorithme d'apprentissage	B.F.G.S
Critère d'arrêt de l'apprentissage	Lorsque la REQM sur l'ensemble de validation atteint un minimum
Nombre d'itérations dans la phase de rétropropagation	100
Découpages des ensembles de données	Apprentissage : pour optimiser les poids des connexions Validation : pour évaluer les performances du modèle

Les valeurs prédites de la toxicité pour les herbicides dans les ensembles d'apprentissage et de validation à l'aide du modèle RNA ont été reportées sur le tableau 4.5 et tracées par rapport à leurs valeurs observées (Figure 4.7). Une corrélation étroite entre les valeurs prédites par le modèle PMC- RNA et les valeurs observées de toxicité a été constatée. L'étude des résidus pour les valeurs observées dans les ensembles d'apprentissage et de validation et celles prédites ont été déterminés et étudiés. On peut remarquer que le modèle n'a pas montré d'erreur proportionnelle et systématique, parce que la distribution des résidus des deux côtés de la ligne médiane est aléatoire.

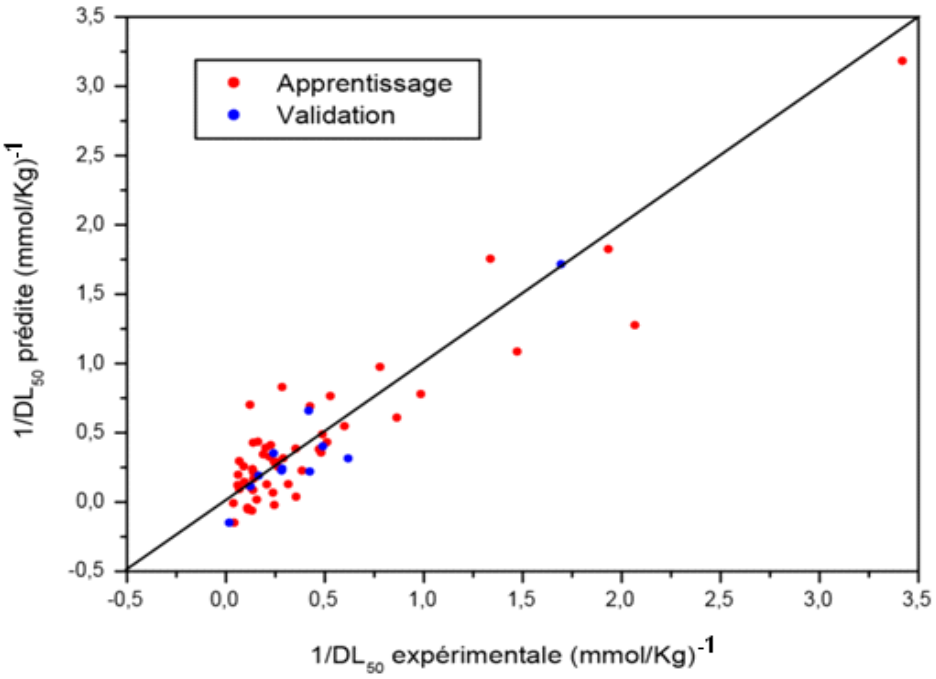


Figure 4.7 : Représentation graphique du modèle PMC-RNA de prédiction de la toxicité orale aigüe ($1/DL_{50}$) chez le rat d'une série d'herbicides

4.2.1.3 Contribution des descripteurs

Pour voir l'importance de chaque variable sur la prédiction de la toxicité aiguë, une analyse de sensibilité à l'aide du logiciel STATISTICA a été menée. Cette méthode proposée par Garson (1991) puis reprise par Goh (1995) fournit une quantification de l'importance relative des différentes entrées (variables) sur la sortie du réseau de neurones. L'importance des descripteurs du modèle RNA est indiquée sur la figure 4.8. On notera que quatre descripteurs (HATS0m, HATSe, G3p et nS) capitalisent 71 % de la contribution à la toxicité

des herbicides étudiés. D'autre part, les deux descripteurs les plus significatifs dans les deux modèles RLM et RNA sont les mêmes : HATS0m et HATSe. La contribution relative des autres descripteurs est différente dans les deux modèles.

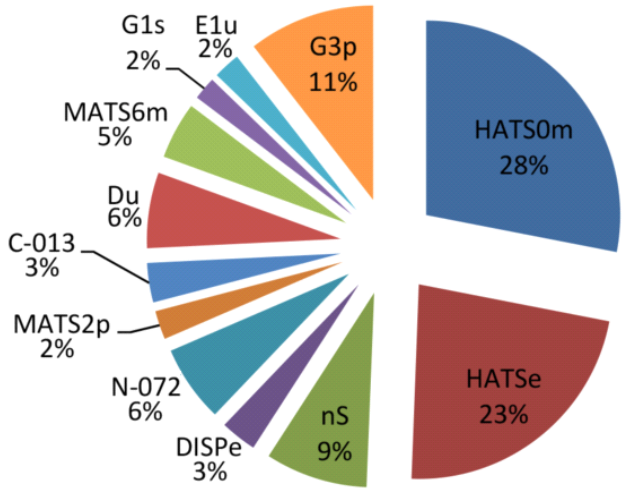


Figure 4.8 : Importance des descripteurs dans la prédiction de la toxicité orale aigüe des herbicides.

4.2.2 Comparaison entre les modèles RLM et RNA

Pour comparer à la fois les performances et la qualité de la prédiction des deux modèles obtenus dans ce travail (*RLM et PMC-RNA*), une évaluation statistique des deux modèles pour l'ensemble de validation est donnée dans le tableau 4.7. Les coefficients de corrélation (R) ayant une valeur supérieure à 0,8 indiquent que la valeur de la toxicité prédite est acceptable. Il en est de même pour la racine carrée de l'erreur quadratique moyenne (*SDEP*) dont les valeurs sont inférieures à 0,3. Cependant, la prédiction déterminée par le modèle *PMC-RNA* est meilleure que celle fournie par le modèle *RLM*. Une amélioration substantielle du paramètre statistique (Q^2_{ext}) pour le modèle *PMC-RNA* peut donc être notée.

Tableau 4.7: Comparaison des paramètres statistiques des deux modèles RLM et PMC-RNA

Paramètres statistiques	Ensemble de validation	
	RLM	PMC-RNA
n	11	11
R	0.878	0.945
Q^2_{ext}	0.782	0.878
$SDEP_{ext}$	0.201	0.150

Par rapport aux résidus du modèle RLM, ceux donnés par le modèle PMC-RNA sont plus près de la ligne médiane représentant l'égalité entre la toxicité observée et celle prédite (figure 4.9). Aussi, on peut donc conclure que le modèle PMC-RNA a un meilleur pouvoir prédictif que le modèle RLM. Cela signifie que le modèle obtenu avec le réseau de neurones permet, dans une large mesure, de confirmer qu'une relation non linéaire relie la toxicité aiguë des herbicides à leur structure moléculaire.

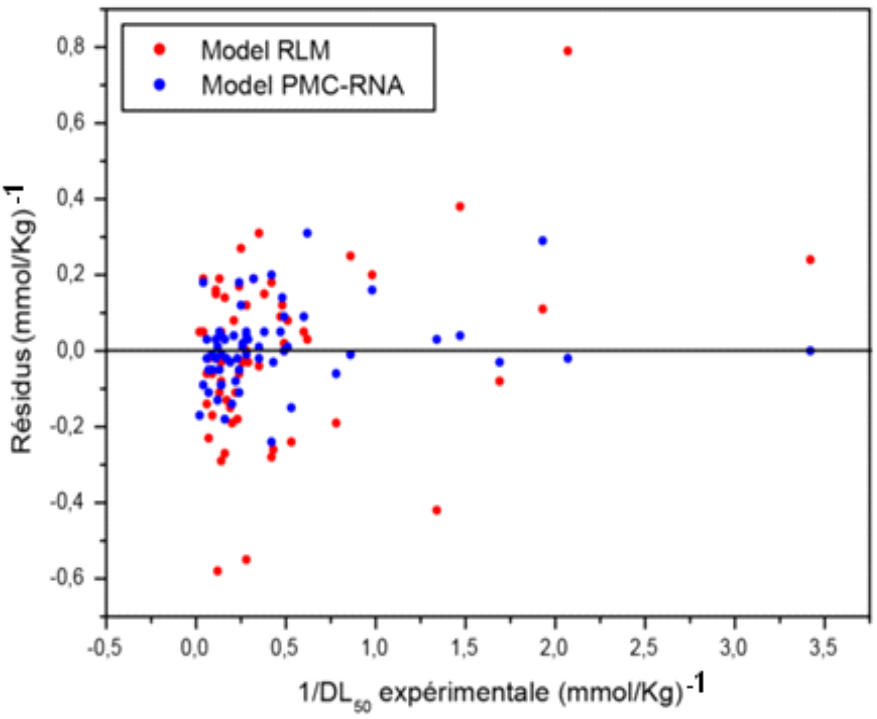


Figure 4.9 : représentation graphique des résidus en fonction de la toxicité ($1/DL_{50}$) expérimentale.

4.2.3 Domaine d'applicabilité du modèle

Le domaine d'application (Eriksson et al., 2003) d'un modèle QSAR doit être défini si le modèle est utilisé pour le criblage de nouveaux composés. Aussi, les prédictions pour les composés qui font partie de ce domaine peuvent être considérées comme fiables. Le domaine d'applicabilité est une surface sur le diagramme de Williams dont la ligne de référence verticale est la valeur de l'effet de levier critique (h*) et les lignes de référence horizontales représentent les limites des valeurs normales (fixées à ± 3 fois l'écart type [± 3s]) pour les résidus standardisés.

Les valeurs de levier calculées pour chaque composé et tracées par rapport aux résidus standardisés (diagramme de Williams) permettent une détection graphique à la fois des valeurs aberrantes et des produits chimiques influents dans un modèle. Le diagramme de Williams pour le modèle de prédiction de la toxicité orale aigüe des herbicides est représenté sur la figure 4.10 ci-après. Les valeurs de leviers (hii) de tous les composés dans les ensembles d'apprentissage et de validation sont inférieures à la valeur de levier critique (h * = 0,63).

Par ailleurs, tous les composés des deux ensembles ont des résidus normalisés plus petits que trois unités d'écart type (3s). Cela signifie que toutes les valeurs prédites sont acceptables. Néanmoins, un seul composé (Bromoxynil octanoate) présente un résidu normalisé > à +3s. Ces résultats montrent que le modèle PMC-RNA obtenu prédit de façon acceptable la toxicité orale aigüe pour plus de 98% des herbicides utilisés.

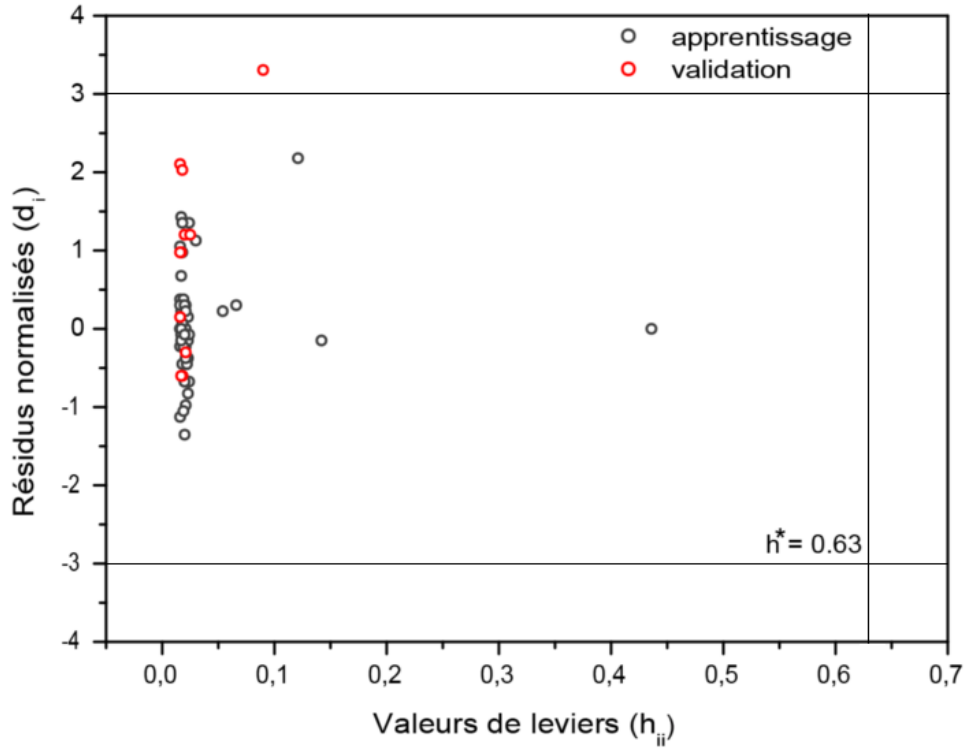


Figure 4.10 : Diagramme de Williams pour le modèle PMC-RNA de prédiction de la toxicité ($1/DL_{50}$) des herbicides**.**

4.2.4 Expression mathématique du modèle PMC-RNA obtenu

L'architecture du réseau PMC-RNA optimisé est représentée schématiquement sur la figure 4.11 ci-dessous.

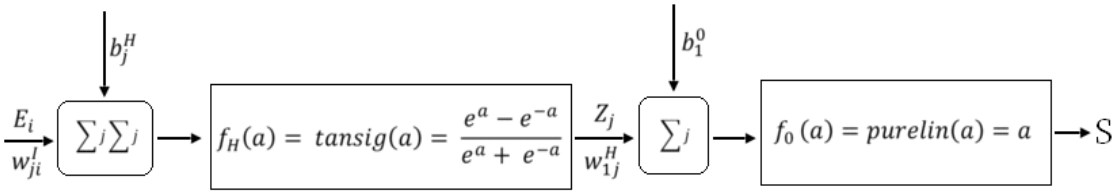


Figure 4.11 : représentation schématique du réseau PMC-RNA optimisé

E_i représentent les entrées (i allant de 1 à 12)

w^I_{ij} représentent les poids qui relient les entrées aux "j" neurones de la couche cachée (j allant de 1 à 14)

b^H_j sont les biais des poids reliant les entrées aux "j" neurones de la couche cachée

Z_j sont les sorties des "j" neurones de la couche cachée

w^H_{1j} représentent les poids des "j" neurones de la couche cachée allant vers la sortie

b^0_1 est le biais des poids de la couche cachée allant vers la sortie

S est la sortie désirée (la toxicité dans le cas de notre étude).

Les expressions mathématiques des sorties Z_j de la couche cachée et S de la couche de sortie sont données ci-après :

$$Z_j = f\left[\sum_{i=1}^{12} w_{ij}^I E_i + b_j^H\right] = \frac{exp\left(\sum_{i=1}^{12} w_{ij}^I E_i + b_j^H\right) - exp\left(-\sum_{i=1}^{12} w_{ij}^I E_i + b_j^H\right)}{exp\left(\sum_{i=1}^{12} w_{ij}^I E_i + b_j^H\right) + exp\left(-\sum_{i=1}^{12} w_{ij}^I E_i + b_j^H\right)} \qquad (4.9)$$

$$S = f_0\left[\sum_{j=1}^{14} w_{1j}^H Z_j + b_1^0\right] = \sum_{j=1}^{14} w_{1j}^H Z_j + b_1^0 \qquad (4.10)$$

4.2.5 Prediction de la toxicité à partir de l'expression mathématique

Dans cette partie, nous avons voulu tester, une fois de plus, la performance du modèle neuronal optimisé par l'application de l'équation mathématique (4.10) décrivant ce modèle et ce à des fins de prédiction de la toxicité ($1/LD_{50}$) de 29 d'herbicides non utilisés lors de la phase d'élaboration du modèle. La matrice des poids et biais du réseau optimisé est donnée sur le tableau 4.8 ci-dessous.

Tableau 4.8: Poids et biais du réseau de neurone optimal

Connexions Couche d'entrée-Couche cachée							
Poids							
w_{j1}^{I}	w_{j2}^{I}	w_{j3}^{I}	w_{j4}^{I}	w_{j5}^{I}	w_{j6}^{I}	w_{j7}^{I}	w_{j8}^{I}
0,60591	-1,05836	-0,10059	-0,05995	0,01360	-0,02813	-0,43880	0,65274
-0,00316	1,42744	-0,52033	0,22750	0,04993	-0,48253	0,46492	0,08951
0,28997	0,78470	-0,35022	0,14398	0,10318	0,27727	-0,42367	0,05550
0,15350	-0,92046	-0,08590	0,10285	-0,17198	0,59450	-0,40146	-0,20080
-0,48627	-0,43540	0,08636	0,16533	-0,14738	0,07093	0,09440	0,16843
-0,11762	1,49614	0,05101	0,72118	0,06545	0,35783	0,65477	-0,58021
0,08501	1,02338	0,02471	-0,39639	-0,35269	-0,07145	0,21246	-0,13498
-0,39455	0,06816	-1,01318	0,00387	-0,07955	-0,14344	-0,30634	0,01778
-0,12128	-0,53164	-0,01800	-0,33468	-0,08479	-0,18086	0,01052	-0,39286
-0,07839	0,70440	-0,12781	-0,05386	0,14043	-0,84867	-0,04714	0,32299
-0,16325	0,32990	0,45316	0,14823	0,05450	-0,36061	0,04407	-0,27263
-0,06904	-0,21066	-0,00170	0,08236	-0,13389	0,19963	-1,10270	0,20519
-0,80177	-0,11688	0,08651	-0,01055	-0,04166	-0,52795	0,88192	-2,06527
0,73759	0,09206	-0,11941	0,16084	-0,05471	1,50403	-0,41491	0,70547

Tableau 4.8: Poids et biais du réseau de neurone optimal (suite)

Connexions Couche d'entrée-Couche cachée					**Connexions Couche cachée-Couche de sortie**	
Poids				**Biais**	**Poids**	**Biais**
w_{j9}^{I}	w_{j10}^{I}	w_{j11}^{I}	w_{j12}^{I}	b_{j}^{H}	w_{1j}^{H}	b_{1}^{0}
0,04365	0,01492	-0,00241	0,00526	-0,11161	0,28520	
0,64040	0,05059	-0,00611	0,11636	0,39164	-1,02934	
-0,00257	0,01037	0,05699	0,07016	-0,41192	-0,54159	
-0,49667	0,03040	-0,02354	0,00531	0,24971	0,49967	
-0,10932	0,00090	-0,21451	-0,00871	0,07236	-0,10223	
-0,05159	0,14090	-0,05090	0,02457	-0,16539	-0,17922	
0,07980	-0,20630	-0,06457	0,01152	-0,55568	-0,72228	0,40979
-0,30142	-0,02629	0,00571	0,00461	-0,17370	-0,36769	
-0,01851	-0,11618	0,11391	-0,00435	-0,05373	0,86459	
-0,02224	-0,06178	0,00377	-0,02493	0,33629	0,50555	
0,12556	0,09859	0,03359	0,02113	0,06205	0,06336	
0,06294	0,01636	0,02591	-0,07600	-0,08291	-0,02115	
0,05775	0,01727	0,00287	0,02011	-0,04103	0,01080	
-0,06085	0,04629	0,01822	-0,04924	-0,01164	-0,05970	

Les valeurs prédites de la toxicité sont répertoriées dans le tableau 4.9 et représentées graphiquement sur la figure 4.12 ci-après.

Tableau 4.9: Valeurs de $1/DL_{50}$ des herbicides prédites à partir du modèle mathématique

Herbicide	$1/DL_{50}$ Expérimentale $(mmol/Kg)^{-1}$	$1/DL_{50}$ Prédite $(mmol/Kg)^{-1}$
2,4-D	0,47	0,39
Acide 4 chlorophenoxyacetique	0,22	0,34
Chloridazone	0,10	0,17
Chlorprophame	0,05	0,11
Clodinafop-propargyl	0,25	0,21
Clomazone	0,18	0,10
Cyanazine	0,84	0,81
Diallate	0,68	0,51
Dichlorprop	0,28	0,31
Dichlorprop-p	0,41	0,31
Diethatyl éthyle	0,14	0,07
Dimexano	0,89	0,88
EPTC	0,21	0,20
Fluazifop-butyl	0,13	0,14
Fluazifop-p-butyle	0,16	0,10
Fluchloraline	0,23	0,31
Flufenacet	0,61	0,46
Hexazinone	0,15	0,23
Ioxynil	2,85	2,86
Isoproturon	0,11	0,16
Mcpa-thioethyl	0,54	0,52
Mecoprop	0,18	0,19
Mepiquat	0,08	0,07
Methazole	0,34	0,36
Naptalame	0,16	0,28
Paraquat	1,69	1,13
Pendimethaline	0,09	0,05
Piperophos	1,09	1,07
Triallate	0,37	0,32

A la lumière des résultats obtenus, nous remarquons que le modèle mathématique prédit de manière satisfaisante la toxicité des herbicides. Le meilleur ajustement linéaire entre les valeurs prédites et les valeurs expérimentales est le suivant (équation 4.11) :

$$(1/DL_{50})^{predite} = 0.913\ (1/DL_{50})^{expérimentale} + 0.011 \qquad (4.11)$$

avec $R^2 = 0.956$; écart type (ou SD) = 0.116 et $p < 0.0001$

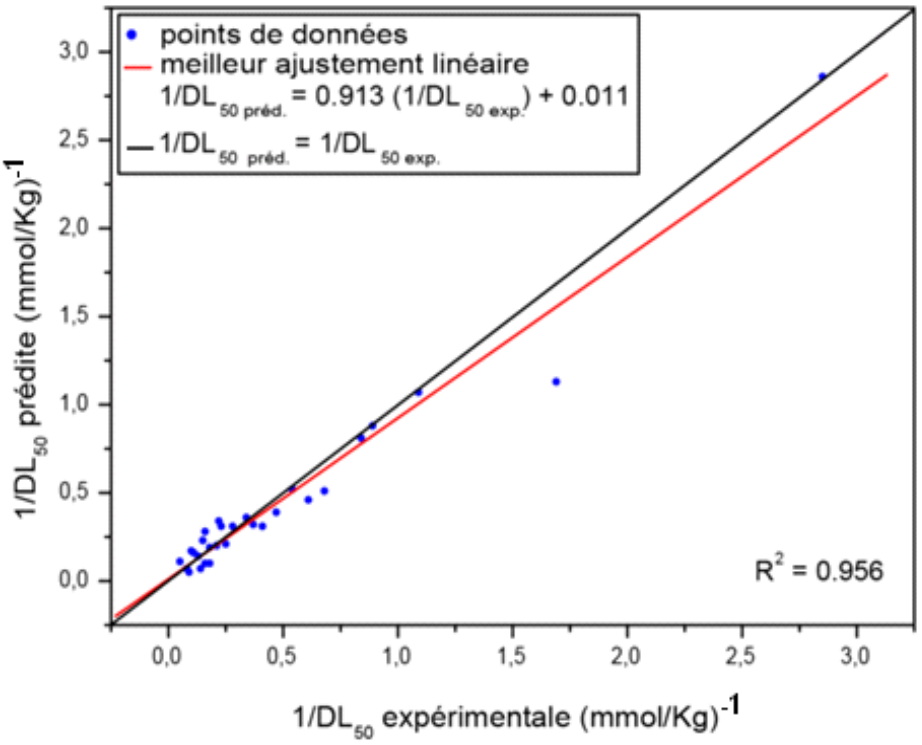


Figure 4.12 : Courbe de corrélation entre les valeurs expérimentales et prédites par le modèle mathématique.

4.2.6 Conclusion

Dans cette première partie de notre étude, nous avons établi un modèle neuronal (PMC-RNA) et un modèle de régression linéaire (RLM) pour la prédiction de la toxicité aigüe des herbicides. Ce modèle de prédiction a été obtenu suite à un traitement préliminaire de la base de données initiale, laquelle nous a permis de déterminer le nombre optimale et la nature des variables d'entrée (descripteurs moléculaires). Suite à cela, deux modèles QSAR ont été élaborés. Par comparaison avec les résultats satisfaisants du modèle MLR, ceux du modèle neuronal sont plus performants. Les différents paramètres statistiques et le domaine d'applicabilité nous ont permis de constater que le modèle PMC-RNA est performant. Par ailleurs, le modèle mathématique issu de l'étude QSAR neuronale a été appliqué à la prédiction de la toxicité aigüe de 29 herbicides ne figurant pas dans la base de données. Les résultats obtenus sont plus qu'encourageants. En effet, la toxicité de 94% de ces molécules a été correctement prédite.

4.3 PREDICTION DE LA TOXICITE DES INSECTICIDES

Dans cette partie de notre étude, nous présentons les résultats de la modélisation de la toxicité orale aigüe (DL_{50}) des insecticides par les réseaux de neurones artificiels uniquement et ce au vu des résultats de comparaison obtenus lors de l'étude précédente. Cette étude a été scindée en deux phases : la première est dédiée à l'élaboration de la base de données. Dans la seconde phase, on s'intéressera à la modélisation de la toxicité aigüe des insecticides par les réseaux de neurones de type perceptron multicouches. Enfin on conclura par une analyse et une interprétation des différents résultats obtenus.

4.3.1 Résultats et discussion

4.3..1.1 Prétraitement de la base de données

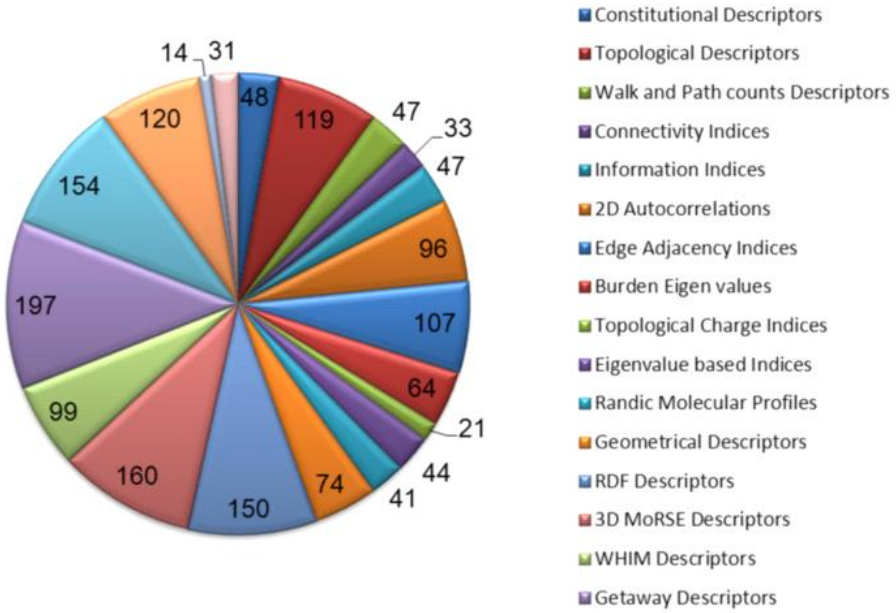
Lors de cette étape, nous avons sélectionné 139 molécules d'insecticides dont la DL_{50} est comprise entre 0.01 et 29.72 mmol/Kg (Tableau 4.10). Comme étape préliminaire, il a été question de la réduction du nombre de descripteurs. Par application de la méthode décrite au paragraphe 4.1.3, on a pu sélectionner 8 descripteurs. Cette sélection a été effectuée sur la base du nombre de descripteurs fournissant le plus grand coefficient de corrélation ($R^2_{adjusté}$) et la racine de l'erreur quadratique moyenne (REQM) la plus petite possible. Les 8 descripteurs sélectionnés et retenus à l'issue de cette étape sont consignés dans le tableau 4.10 ci-contre.

Tableau 4.10 : Descripteurs sélectionnés lors de l'étude des insecticides

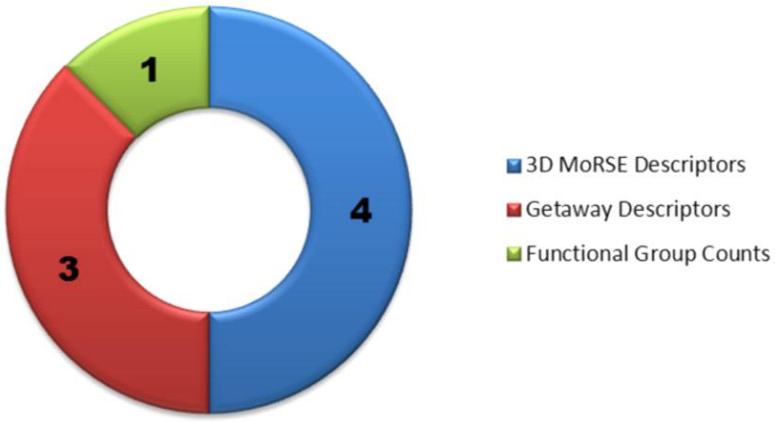
Descripteurs sélectionnés et catégorie d'appartenance	
Mor11u	3D-MoRSE - signal 11 / unweighted
Mor26e	3D-MoRSE - signal 26 / weighted by atomic Sanderson electronegativities
Mor19p	3D-MoRSE - signal 19 / weighted by atomic polarizabilities
Mor26p	3D-MoRSE - signal 26 / weighted by atomic polarizabilities
H6m	Getaway-H autocorrelation of lag 6 / weighted by atomic masses
H6v	Getaway-H autocorrelation of lag 6 / weighted by atomic van der Waals volumes
R6m+	Getaway-R maximal autocorrelation of lag 6 / weighted by atomic masses
nArX	Functional Group Counts-number of X on aromatic ring

Sur la base des résultats obtenus sur la matrice de correlation entre les descripteurs, on note que la valeur du coefficient de corrélation la plus élevée de chaque paire de descripteurs

choisis est de 0,537, ce qui signifie que les descripteurs retenus sont indépendants. De plus, la plus grande valeur du VIF est de 2.431, ce qui implique qu'il n'y'a pas de relation entre les descripteurs sélectionnés. Le nombre de descripteurs initiaux et ceux obtenus après les 5 étapes de sélection sont schématisés sur la figure 4.13 ci-contre.



(a)



(b)

Figure 4.13: Nombre de Descripteurs initiaux (a) et nombre de descripteurs pertinents obtenus après exécution des 5 étapes (b).

4.3.1.2 Normalisation et mise en forme des données

La conception et l'optimisation du réseau ont été réalisées en élaborant un programme en MATLAB. Tout d'abord, il a fallu normaliser toutes les variables d'entrée et la variable de sortie. En effet, si des entrées ont des grandeurs très différentes, celles qui sont petites n'auront pas d'influence lors de l'apprentissage. Avec cette normalisation, l'apprentissage sera plus efficace et très rapide et ce en empêchant la fonction de transfert de devenir saturée. Les entrées et les sorties sont normalisées à un intervalle de [-1, 1] en utilisant l'équation (4.12) de normalisation. La fonction « premnmx » de MATLAB est celle qui réalise la normalisation selon l'équation (4.12), alors que la dénormalisation est réalisée par la fonction « *postmnmx* ».

$$X_n = \frac{2(X_0 - X_{min})}{X_{max} - X_{min}} - 1 \qquad (4.12)$$

où X_n est la valeur normalisée du paramètre X.

X_0, X_{max} et X_{min} sont respectivement les valeurs réelles, maximales et minimales de X.

Suite à cela et après plusieurs essais, les données (139 molécules d'insecticides) ont été réparties en deux ensembles distincts (Figure 4.14), en réservant 88.5 % de données (123 insecticides) pour la phase d'apprentissage et 11.5% (16 insecticides) pour la validation croisée.

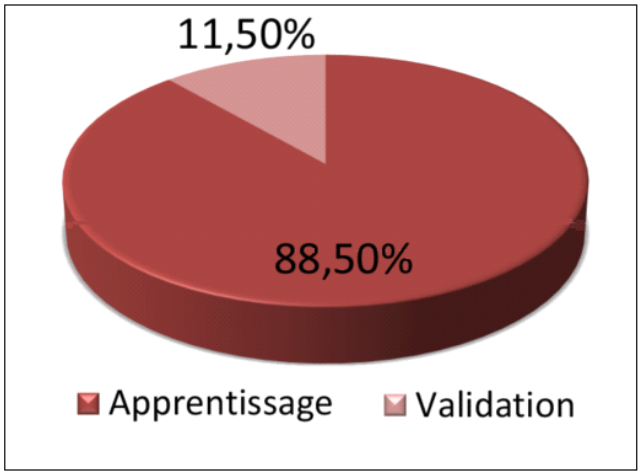


Figure 4.14. Mise en forme des données

4.3.1.3 Réseau optimal

Le nombre de neurones dans la couche cachée a 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". 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 4.11 ci-après récapitule les différents éléments structuraux du réseau de neurones optimal.

Tableau 4.11 : Structure du réseau de neurones optimal obtenu dans la modélisation de la toxicité aigüe.

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

Pour évaluer la performance du modèle neuronal obtenu, la représentation graphique des valeurs prédites de DL_{50} en fonction des valeurs expérimentales pour les ensembles d'apprentissage et de validation obtenues est donnée sur les figures 4.15 et 4.16 ci-après. Une corrélation étroite entre les valeurs calculées par le modèle PMC-RNA et les valeurs expérimentales de la toxicité aigüe a été observée.

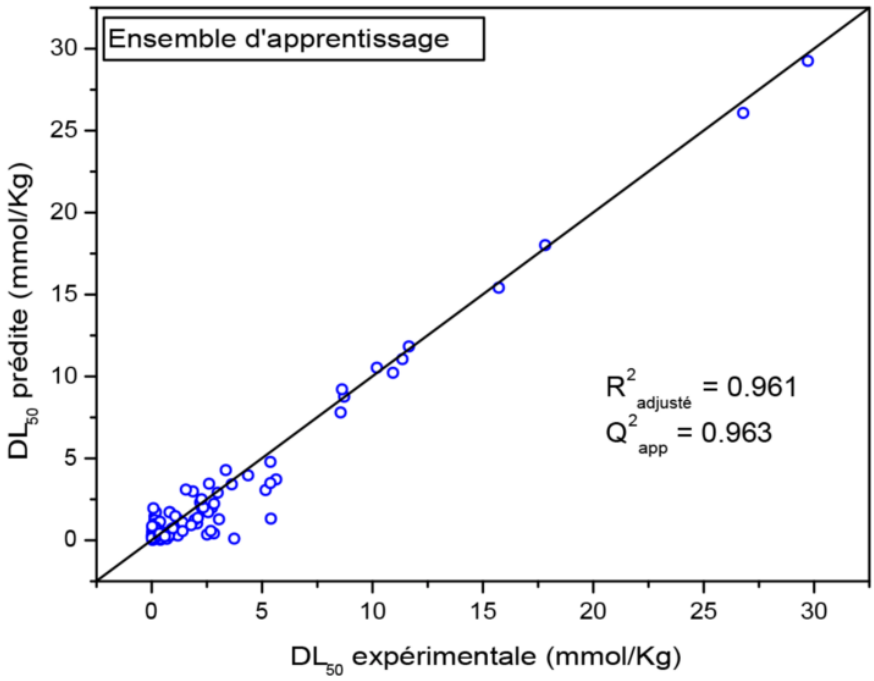


Figure 4.15 : Courbe de corrélation des valeurs prédites de DL_{50} en fonction des valeurs expérimentales pour l'ensemble d'apprentissage.

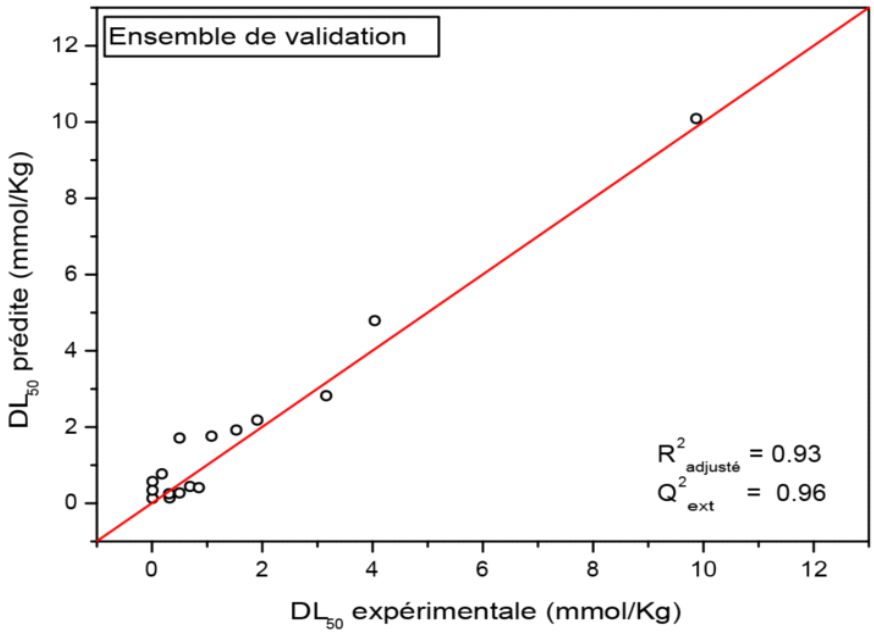


Figure 4.16 : Courbe de corrélation des valeurs prédites de DL_{50} en fonction des valeurs expérimentales pour l'ensemble de validation.

La véritable utilité des modèles QSAR n'est pas seulement leur capacité à reproduire les données connues, mais aussi avoir un potentiel prédictif. Pour cette raison, on a établi les paramètres caractéristiques du modèle qui figurent sur le tableau 4.12 ci-dessous. Sur le tableau 4.13 sont reportées les valeurs expérimentales et prédites des insecticides.

Tableau 4.12 : Valeurs des paramètres statistiques pour les ensembles d'apprentissage et de validation

Paramètres statistiques	Apprentissage	Validation
n	123	16
R	0.982	0.983
$R^2_{adjusté}$	0.961	0.930
Q^2	0.963	---
Q^2_{ext}	---	0.957
$SDEP_{ext}$	---	0.500

La valeur de 0.963 pour Q^2 suggère une cohérence de l'ensemble d'apprentissage et une bonne validation interne. Par conséquent la robustesse du modèle a été confirmée. D'autre part, la grande valeur de Q^2_{ext} (0.957) préconise un bon pouvoir prédictif du modèle.

Les résidus pour les valeurs observées dans les ensembles d'apprentissage et de validation ont été déterminés. Comme pour le cas des herbicides, le modèle n'a pas montré d'erreur proportionnelle et systématique, parce que la distribution des résidus des deux côtés de la ligne médiane est aléatoire.

Tableau 4.13 : Valeurs de DL_{50} expérimentales et prédites par PMC-RNA des insecticides

Insecticides	DL_{50} exp. (mmol/Kg)	DL_{50} préd. (mmol/Kg)	Insecticides	DL_{50} exp. (mmol/Kg)	DL_{50} préd. (mmol/Kg)
Ensemble d'apprentissage					
Acéphat	5,16	1,06	Chlorfenethol	1,87	2,98
Acetamipride	0,96	0,34	Chlorfensulphide	11,36	11,05
Alanycarbe	0,83	1,71	Chlorfenviphos	0,03	0,80
Aldoxycarbe	0,12	1,00	Chlorfluazuron	15,72	15,41
Aldrine	0,11	0,08	Chlormephos	0,03	0,42
Allethrine	2,27	2,00	Chlorobenzilate	8,56	7,80
Allyxycarb	0,32	1,21	Chloromethiuron	10,93	10,22
Alphacypermethrine	0,14	0,09	Chlorpyriphos-Me	8,72	8,76
Alphaendosulfan	0,09	0,41	Chlorthiophos	0,03	0,28
Alpha-HCCH	0,61	0,44	Cloethocarb	0,14	1,00
Amidithion	2,20	1,31	Closantel	0,40	0,03
Aminocarbe	0,14	0,36	Crotoxyphos	0,21	0,28
Amitraze	2,73	1,98	Cyanophos	2,51	0,35
Aramite	11,65	11,83	Cyhalothrine	0,31	0,10
Azametiphos	3,63	3,41	Cypermethrine	0,69	0,09
Azinphos-ethyl	0,03	0,54	Deltamethrine	0,17	0,44
Azinphos-methyle	0,03	1,46	Demeton-s-methyle	0,17	0,33
Azoxybenzene	2,60	3,45	Demeton-s-MS	0,12	0,92
Bamphéchlore	0,12	0,25	Diafenthiuron	5,38	4,78
Bendiocarbe	0,15	0,21	Dialiphos	0,18	0,44
Bensultap	2,56	1,71	Diazinon	3,74	0,10
Betacypermethrine	0,40	0,03	Dicrotophos	0,07	0,44
Bifenthrine	0,13	1,48	Dieldrine	0,12	0,34
Binapacryl	0,18	1,10	Diethyltoluamide	10,20	10,53
Bioallethrine	2,34	2,00	Diflovidazin	1,95	1,25
Bromophos	4,37	3,96	Dimethoate	1,07	1,47
Bromophos-éthyl	0,13	1,46	Dimethylvinphos	0,29	0,51
Butonate	3,36	4,28	Dioxabenzophos	0,58	0,27
Butoxycarboxime	2,06	1,03	Dioxacarbe	0,11	0,59
Cadusafos	0,11	0,42	Dioxathion	0,05	0,02
Carbanolate	0,14	0,33	Endosulfan	0,09	0,41
Carbaryl	3,05	1,28	Endothion	0,08	0,52
Carbofuran	0,03	0,25	Endrine	0,02	0,34
Carbophenothion	0,03	0,04	Epn	0,04	0,17
Carbosulfan	0,27	1,10	Esfenvalerate	0,21	0,82
Cartap	1,19	1,28	Ethiofencarbe	0,89	0,66
Chlordane	1,12	0,70	Ethion	0,54	0,35
Chlordecone	0,19	1,71	Ethiprole	17,82	18,01
Chlordimeforme	0,81	1,71	Ethoate-methyle	1,40	1,09
Chlorethoxyfos	0,13	1,23	Fenazaflor	0,64	0,26
Fenchlorphos	1,55	3,09	Methidathion	0,08	1,95

Tableau 4.13 : Valeurs de DL_{50} expérimentales et prédites par PMC-RNA des Insecticides (suite et fin).

Insecticides	DL_{50} exp. (mmol/Kg)	DL_{50} préd. (mmol/Kg)	Insecticides	DL_{50} exp. (mmol/Kg)	DL_{50} préd. (mmol/Kg)
Ensemble d'apprentissage					
Fenitrothion	1,19	0,29	Methiocarb	0,08	0,11
Fenobucarb	2,99	2,90	Methomyl	0,18	0,79
Gamma-cyhalothrine	0,11	0,13	Mexacarbate	0,02	0,30
Halfenprox	0,28	0,24	Mirex	0,03	0,39
Halofenozide	8,62	9,21	Morphothion	0,82	0,46
Heptachlore	0,39	0,09	Nithiazine	1,40	0,56
Heptenophos	0,38	0,58	Omethoate	0,23	1,05
Imiprothrin	2,83	2,23	Phenkapton	0,12	0,10
Indoxacarbe	0,51	0,24	Phenthorate	0,78	0,28
Iodofenphos	5,64	3,71	Phorate	0,01	0,14
Isobenzan	0,01	0,27	Phosalone	0,33	0,28
Isoprocarbe	2,09	1,37	Phosmet	0,36	0,42
Isoxathion	0,36	0,09	Phosnichlor	1,79	0,92
Kelevan	0,38	1,13	Phosphamidon	0,02	0,86
Lambda-cyhalothrine	0,12	0,11	Plifenate	29,72	29,25
Leptophos	0,10	0,47	Profenophos	0,96	0,73
Malathion	5,38	3,49	Prothiofos	2,68	0,57
Mécarbam	5,40	1,32	Pymetrozine	26,79	26,08
Méphospholan	0,13	0,32	Pyridafenthion	2,26	0,50
Methacrifos	2,82	0,42	Pyrimicarbe-	0,60	0,24
Methamidophos	0,21	0,29			
Ensemble de validation					
Acethion	4,04	4,79	Demeton-O-MS	1,91	2,18
Azidithion	3,16	1,82	Disulfoton	0,01	0,33
Benfuracarbe	0,50	0,27	Icaridin	9,87	11,70
Butocarboxime	0,69	0,14	Naled	0,50	3,71
Chlorfénapyr	1,08	1,76	Nitenpyram	0,31	0,25
Chlorpyriphos	0,18	2,77	Oxamyl	0,01	2,07
Cyphenothrin	0,85	0,41	Parathion	0,01	0,54
DDT	0,32	0,14	Prallethrin	1,53	2,62

4.3.1.4 Domaine d'applicabilité

Comme pour les herbicides, le domaine d'applicabilité du modèle élaboré pour la prédiction de la toxicité aigüe des insecticides a été défini par utilisation du diagramme de Williams (figure 4.17).

Le diagramme de Williams fait ressortir 2 observations (Pymetrozine et Plifenate) dans l'ensemble d'apprentissage qui se sont révélées être des valeurs aberrantes avec une valeur de

levier plus élevée que celle du levier critique h* = 0.19. Ces valeurs aberrantes peuvent être expliquées comme des composés ayant des caractéristiques particulières mal représentées dans l'ensemble d'apprentissage, ce qui pourrait avoir une incidence sur la sélection de variables pour une meilleure modélisation de ces composés. Par ailleurs, 3 composés (Diazinon, Acephate et Mecarbam) présentent des valeurs de résidus normalisés supérieures aux valeurs limites choisies (± 3s). Cependant, la majorité des échantillons (96.40%) dans le modèle se situe dans le domaine d'applicabilité ainsi défini, ce qui indique la fiabilité des prédictions.

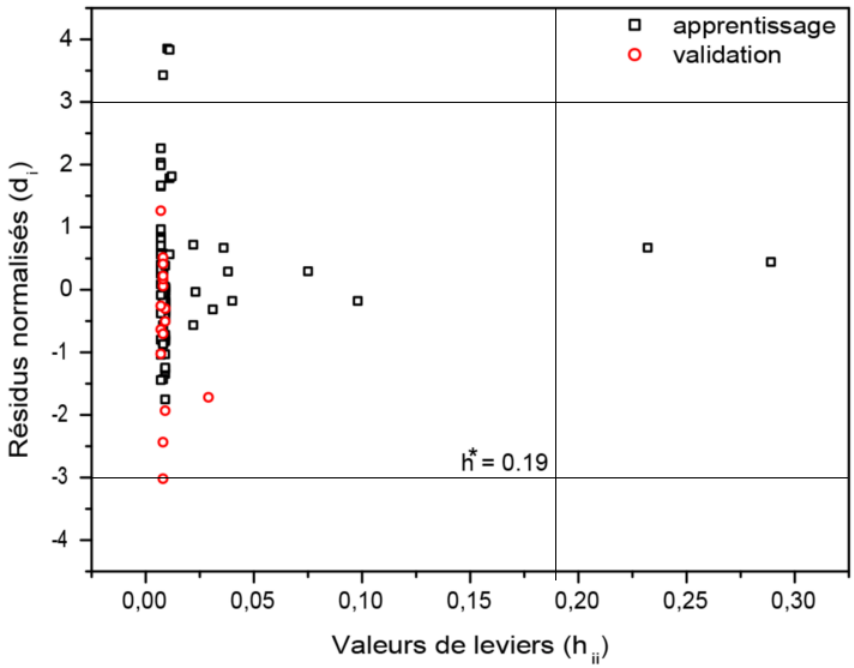


Figure 4.17 : Diagramme de Williams pour le modèle de prédiction de la toxicité des insecticides

4.3.1.5 Analyse de la contribution des descripteurs

L'un des principaux objectifs des études QSAR est la détermination des facteurs qui influent sur la toxicité des composés étudiés. Nous avons vu donc nécessaire d'évaluer la contribution des descripteurs moléculaires au modèle neuronal établi. La contribution de chaque descripteur a été déterminée au moyen du logiciel STATISTICA. Les résultats obtenus sont reportés sur la figure 4.18 ci-dessous.

L'interprétation des descripteurs de dragon n'est pas simple, ce qui la rend difficile. Les descripteurs qui apparaissent dans ce modèle appartiennent à 3 catégories de descripteurs (3D-Morse, Getaway et Functional Group counts).

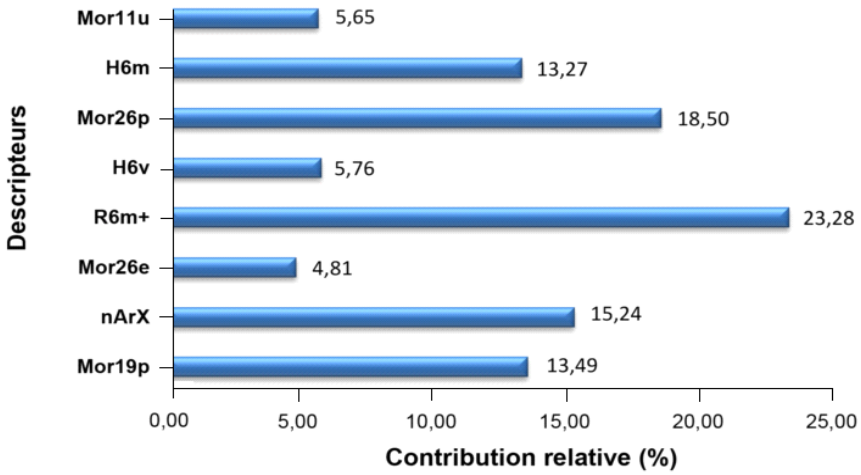


Figure 4.18 : contribution relative des descripteurs dans la prédiction de la toxicité orale aigüe des insecticides.

Les descripteurs 3D-Morse sélectionnés sont liés à des considérations électroniques et structurales des molécules utilisées comme observations (polarisabilité, électronégativité, nature des liaisons et distances entre elles). Les descripteurs Getaway sélectionnés sont relatifs à la masse atomique et aux structures tridimensionnelles des molécules et fournissent des informations sur leurs effets stériques. Le descripteur nArX appartenant à la catégorie Functional Group Counts est relatif au nombre d'atomes d'halogènes sur les cycles aromatiques des composés. La figure 4.18 indique que l'importance relative des descripteurs varie dans l'ordre suivant : R6m+ > Mor26p > nArX > Mor19p > H6m > H6v > Mor11u > Mor26e. On notera que les cinq premiers descripteurs capitalisent à eux seuls 84 % de la contribution à la toxicité des insecticides. Au vu de ces résultats, la toxicité orale aiguë (DL_{50}) des insecticides pour les rats est influencée principalement par la structure de la molécule, ses propriétés électroniques et la présence d'halogènes sur les cycles aromatiques.

4.3.2 Conclusion

Le but de cette étude a été d'élaborer un modèle QSAR pour prédire la toxicité orale aigüe des insecticides. Différents descripteurs moléculaires théoriques ont été calculés online par le logiciel DRAGON et soumis à une sélection dont la finalité est de ne garder que les descripteurs pertinents. Le modèle PMC-RNA construit a été évaluée (validations internes et externes) et tous les résultats qui sont satisfaisants indiquent que le modèle QSAR que nous avons construit est fiable et prédictif. La nature des descripteurs sélectionnés a montré que les effets électroniques en particulier jouent un rôle principal dans l'activité toxicologique des insecticides.

4.4 PREDICTION DE LA TOXICITE AIGÜE DES FONGICIDES

On présente dans cette partie du travail, les résultats de la modélisation de la toxicité des fongicides par les réseaux de neurones artificiels. On parlera de l'élaboration de la base de données et ensuite on s'emploiera à la modélisation de la toxicité par les réseaux de neurones de type perceptron multicouches. Enfin on conclura cette partie par une analyse et une interprétation des différents résultats obtenus. Nous tenons à faire remarquer que la méthodologie (élaboration et prétraitement de la base de données, normalisation et mise en forme des données) est toujours la même que celle suivie pour les insecticides.

4.4.1 Résultats et discussion

4.4.1.1 Sélection des descripteurs pertinents

La procédure de sélection des descripteurs pertinents, analogue à celle utilisée précédemment avec les herbicides et les insecticides, a abouti à la sélection de 10 descripteurs reportés sur le tableau 4.14 ci-après.

Tableau 4.14 : Descripteurs sélectionnés pour la prédiction des fongicides

Descripteurs	Catégorie
T (O...O)	Topological Descriptors
RDF020e	RDF Descriptors
RDF030e	
RDF080e	
Mor15m	3D MoRSE Descriptors
Mor26u	
Mor16u	
nHDon	Functional Group Counts
C-008	Atom centered fragments
TPSA (Tot)	Molecular Properties

4.4.1.2 Fongicides utilisés

Lors de cette étape, nous avons sélectionné 91 molécules de fongicides. Sur le tableau 4.15 ci-dessous, sont reportés les fongicides utilisés ainsi que leur écriture SMILES qui a permis le calcul online des descripteurs par le logiciel e-Dragon 1.0.

Tableau 4.15 : fongicides utilisés ainsi que leur écriture SMILES

Fongicides	Notation smiles
Acide Peracetique	C(OO) (C)=O
Ampropylfos	O=P(O) (O) C(N) CC
Anilazine	Clc1nc (nc(Cl) n1) Nc2ccccc2Cl
Benalaxil	O=C(N(c1c(cccc1C)C)C(C(=O)OC)C)Cc2ccccc2
Biphényle	c1cc(ccc1)c2ccccc2
Biphenyle-2-Ol	c1(c2c(cccc2)O)ccccc1
Boric Acid	B(O)(O)O
Bupirimate	CCCCC1=C(N=C(N=C1OS(=O)(=O)N(C)C)NCC)C
Carboxine	C1(C(Nc2ccccc2)=O)=C(OCCS1)C
Chloraniformethane	c1(N[C@@H](NC=O)C(Cl)(Cl)Cl)cc(c(Cl)cc1)Cl
Chlordecone	ClC54C(=O)C1(Cl)C2(Cl)C5(Cl)C3(Cl)C4(Cl)C1(Cl)C2(Cl)C3(Cl)Cl
Chloronebe	Clc1cc(OC)c(Cl)cc1OC
Chloroxylenol	c1(c(cc(O)cc1C)C)Cl
Climbazole	n1([C@@H](Oc2ccc(Cl)cc2)C(C(C)(C)C)=O)ccnc1
Cyprofuram	O=C3OCCC3N(C(=O)C1CC1)c2cccc(Cl)c2
Dazomet	S=C1SCN(C)CN1C
Dichlone	O=C2c1c(cccc1)C(=O)C(/Cl)=C2/Cl
Dichlorophene	Clc1cc(c(O)cc1)Cc2cc(Cl)ccc2O
Diflumetorim	Clc2c (NC(CC) c1ccc (OC(F) F) cc1) ncnc2C
Dimethomorphe	O=C(\C=C(\c1ccc(Cl)cc1)c2ccc(OC)c(OC)c2)N3CCOCC3
Diniconazole-M	Clc2ccc(/C=C(/n1ncnc1)C(O)C(C)(C)C)c(Cl)c2
Dinobuton	O=C(Oc1c(cc(cc1C(C)CC)[N+]([O-])=O)[N+]([O-])=O)OC(C)C
Dinocap	O=C(Oc1c(cc(cc1C(C)CCCCCC)[N+]([O-])=O)[N+]([O-])=O)/C=C/C
Ditalimfos	O=C2c1ccccc1C(=O)N2P(=S)(OCC)OCC
Dithianon	N#CC=1S\C2=C(/SC=1C#N)C(=O)c3c(C2=O)cccc3
Edifenphos	O=P(Sc1ccccc1)(Sc2ccccc2)OCC
Epichlorohydrine	C1[C@@H](O1)CCl
Epoxiconazole	Fc1ccc(cc1)C3(OC3c2ccccc2Cl)Cn4ncnc4
Etaconazole	Clc1ccc(c(Cl) c1) C2 (OCC(O2) CC) Cn3ncnc3
Ethoxyquine	O(c2ccc1c(\C(=C/C(N1)(C)C)C)c2)CC
Fenamidone	O=C2N(Nc1ccccc1)C(\SC)=N/[C@]2(c3ccccc3)C
Fenapanil	c1([C@@](Cn2ccnc2)(CCCC)C#N)ccccc1
Fenarimol	Clc1ccc(cc1)C(O)(c2ccccc2Cl)c3cncnc3
Fenoxanil	Clc1cc(Cl) ccc1OC(C (=O) NC(C#N) (C) C(C) C) C
Fenpropidine	c1cc(ccc1CC(CN2CCCCC2)C)C(C)(C)C
Fenpropimorphe	O2[C@H](CN(CC(C)Cc1ccc(cc1)C(C)(C)C)C[C@H]2C)C
Fluorure De Sulfuryle	S(=O)(=O)(F)F
Fluquinconazole	Clc1ccc(c(Cl)c1)N3C(=N\c2c(cc(F)cc2)C3=O)/n4ncnc4
Flusilazole	n1 (cncn1) C [Si] (c1ccc(F) cc1) (C) c1ccc (cc1) F
Flusulfamide	Clc2cc(ccc2NS(=O)(=O)c1ccc(Cl)c(c1)C(F)(F)F)[N+]([O-])=O

Tableau 4.15 : fongicides utilisés ainsi que leur écriture SMILES (suite)

Fongicides	Notation smiles
Formaldehyde	C=O
Furalaxyl	O=C(N(c1c(cccc1C)C)C(C(=O)OC)C)c2occc2
Furmecyclox	c1(C(N(C2CCCCC2)OC)=O)cc(C)oc1C
Glutaraldehyde	C(CC=O)CC=O
Hexachlorophène	OC1=C(C=C(C(=C1CC2=C(C(=CC(=C2Cl)Cl)Cl)O)Cl)Cl)Cl
Hexaconazole	Clc1ccc(c(Cl)c1)C(O)(CCCC)Cn2ncnc2
Hymexazol	O=C1/C=C(\ON1)C
Imazalil	Clc1ccc(c(Cl) c1) C (OC\C=C) Cn2ccnc2
Inezin	c1([P@@](SCc2ccccc2)(OCC)=O)ccccc1
Ipconazole	Clc1ccc (cc1) CC2C(O) (C(CC2) C(C) C) Cn3ncnc3
Iprobenfos	O=P(SCc1ccccc1)(OC(C)C)OC(C)C
Isoprothiolane	O=C(OC(C)C)C(\C(=O)OC(C)C)=C1/SCCS1
Isothiocyanate de Me	O=C(Oc1cc(c(c(c1)C)S(=O)C)C)NC
Kasugamycin	O([C@H]1[C@@H]([C@@H](O)[C@H]([C@@H]([C@@H]1O)O)O)O)[C@@H] 1[C@H] (C[C@@H]([C@H](O1)C)NC(=N)C(=O)O)N
Metalaxyl	O=C(N(c1c(cccc1C)C)C(C(=O)OC)C)COC
Metalaxyl-M	O=C(N(c1c(cccc1C)C)[C@@H](C(=O)OC)C)COC
Metconazole	Clc1ccc (cc1) CC2C(O) (C(CC2) (C) C) Cn3ncnc3
Methasulfocarb	O=C(Sc1ccc(OS(=O)(=O)C)cc1)NC
Methfuroxam	O=C(c1c(oc(c1C)C)C)Nc2ccccc2
Metominostrobin	O=C(NC)C(=N\OC)\c2ccccc2Oc1ccccc1
Metsulfovax	c1ccccc1NC(=O)c1c(C)nc(C)s1
Myclobutanil	Clc1ccc(cc1)C(C#N)(CCCC)Cn2ncnc2
Nitrothal Isopropyl	O=[N+]([O-])c1cc(cc(C(=O)OC(C)C)c1)C(=O)OC(C)C
Nuarimol	Fc1ccc(cc1)C(O)(c2ccccc2Cl)c3cncnc3
Ofurace	O=C2OCCC2N(C(=O)CCl)c1c(cccc1C)C
Octhilinone	O=C1/C=C\SN1CCCCCCCC
Oxadixyl	O=C2OCCN2N(C(=O)COC)c1c(cccc1C)C
Oxycarboxine	O=S1(=O)C(=C(/OCC1)C)\C(=O)Nc2ccccc2
Pentachlorophénol	Clc1c(Cl)c(Cl)c(Cl)c(Cl)c1O
Peroxyde D'hydrogène	OO
Prochloraz	CCCN(CCOC1=C(C=C(C=C1Cl)Cl)Cl)C(=O)N2C=CN=C2
Propiconazol	Clc1ccc(c(Cl) c1) C2 (OCC(O2) CCC) Cn3ncnc3
Prothiocarb	O=C(SCC)NCCCN(C)C
Pyrazophos	S=P(OCC)(OCC)Oc1nn2cc(c(nc2c1)C)C(=O)OCC
Pyrifénox	Clc2ccc(C (=NOC) Cc1cccnc1) c(Cl) c2
Pyrimethanil	n1c (cc (nc1Nc2ccccc2) C) C
Pyroquilone	O=C3N2c1c(cccc1CC2)CC3
Simeconazole	Fc(c2)ccc(c2)C(O)(Cn(c1)ncn1)C[Si](C)(C)C
Spiroxamine	O1CC(OC12CCC(CC2)C(C)(C)C)CN(CC)CCC
Tebuconazole	Clc1ccc (cc1) CCC(O) (C(C) (C) C) Cn2ncnc2

Tableau 4.15 : fongicides utilisés ainsi que leur écriture SMILES (suite et fin)

Fongicides	Notation smiles
Tecnazène	Clc1c([N+]([O-])=O)c(Cl)c(Cl)cc1Cl
Tetraconazole	FC(F)C(F)(F)OCC(c1ccc(Cl)cc1Cl)Cn2ncnc2
Thiabendazole	n2c1c(cccc1)nc2c3ncsc3
Thicofene	c1(C#N)c([S@@](=O)CC)sc(c1Cl)C#N
Tolfenpyrad	c1(nn(c(c1Cl)C(=O)NCc1ccc(cc1)Oc1ccc(cc1)C)C)CC
Triacétate De Guazatine	N(=C(\N)N)\CCCCCCCCNCCCCCCCC/N=C(\N)N
Triadimefon	CC(C) (C) C (=O) C (N1C=NC=N1) OC2=CC=C(C=C2) Cl
Triadimenol	Clc2ccc(OC(n1ncnc1)C(O)C(C)(C)C)cc2
Triazoxide	Clc2ccc1nc(n[n+]([O-])c1c2)n3ccnc3
Tricyclazole	n2nc3sc1cccc(c1n3c2)C
Tridemorphe	O1C(CN(CCCCCCCCCCCCC)CC1C)C
Triflumizole	FC(F)(F)c2c(\N=C(\n1ccnc1)COCCC)ccc(Cl)c2
Uniconazole	Clc2ccc(/C=C(/n1ncnc1)[C@@H](O)C(C)(C)C)cc2

Pour statuer sur la corrélation entre les descripteurs sélectionnés, la matrice de corrélation a été établie à l'aide du logiciel XLSTAT. Les résultats trouvés montrent bel et bien que ces descripteurs ne sont pas corrélés entre eux du fait que la plus grande valeur du coefficient de corrélation est de 0.657. Les descripteurs sélectionnés ne présentent pas de multicolinéarité au vu des valeurs du VIF dont la plus grande valeur équivaut à 3.382.

4.4.1.3 Normalisation et mise en forme des données

La procédure de normalisation des données est la même que celle employée lors de l'étude des insecticides. La base de données constituée de 91 molécules, a été scindée en base d'apprentissage (75 observations) et en base de validation (16 observations) comme illustré sur la figure 4.19 ci-dessous.

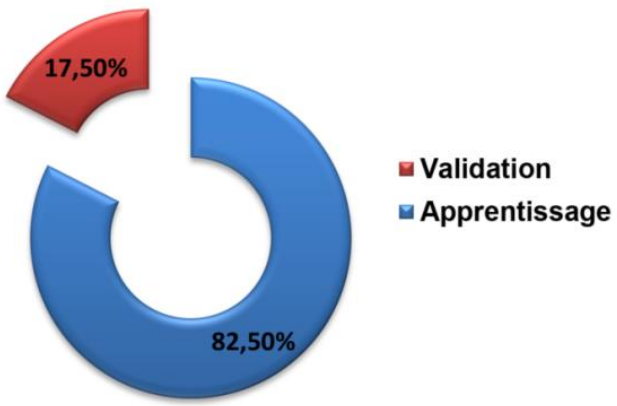


Figure 4.19 : Répartition de la base de données des fongicides

4.4.1.4 Elaboration du modèle neuronal

Comme ça a été le cas pour les insecticides, la méthodologie de l'élaboration du modèle neuronal a pour structure de base un perceptron multicouche à trois couches : une couche d'entrée, une couche cachée, une couche de sortie. Le nombre de neurones dans la couche cachée a 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 » avec option d'apprentissage atténué pour éviter le sur-apprentissage. Le tableau 4.16a ci-après récapitule les différents éléments structurels du réseau de neurones optimal.

Tableau 4.16a : Structure du RNA optimisé pour la toxicité aigüe des fongicides.

	Critère retenu pour cette étude
Architecture du Perceptron Multi-Couche (PMC)	Neurone d'entrée : 10 Neurone de sortie : 01 Neurones cachés : 11 (le plus petit nombre produisant la plus faible erreur de prédiction sur l'ensemble de validation
Fonction d'activation	
Couche cachée :	Sigmoïde Tangente Hyperbolique (Fonction tansig de MATLAB®)
Couche de sortie :	Linéaire (Fonction *purelin* de MATLAB®)
Algorithme d'apprentissage	Rétropropagation de Levenberg-Marquardt. (Fonction *trainlm* de MATLAB®)
Critère d'arrêt de l'apprentissage	Lorsque la REQM sur l'ensemble de validation atteint un minimum
Nombre d'itérations dans la phase de rétropropagation	150
Découpages des ensembles de données	Apprentissage : pour optimiser les poids des connexions Validation : pour évaluer les performances du modèle

Pour évaluer la performance du modèle neuronal obtenu, la représentation graphique des valeurs prédites de DL_{50} en fonction des valeurs expérimentales pour les ensembles d'apprentissage et de validation obtenues est donnée sur les figures 4.20 et 4.21et le tableau 4.16b ci-après. Une corrélation étroite entre les valeurs calculées par le modèle PMC-RNA et les valeurs expérimentales de la toxicité aigüe des fongicides a été observée.

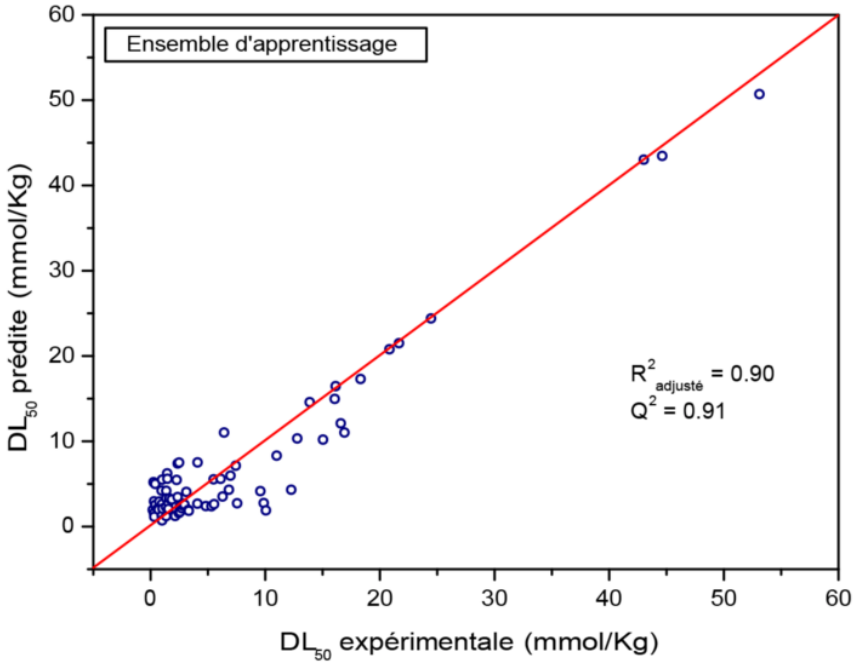


Figure 4.20 : Courbe de corrélation des valeurs prédites de DL_{50} (des fongicides) en fonction des valeurs expérimentales pour l'ensemble d'apprentissage

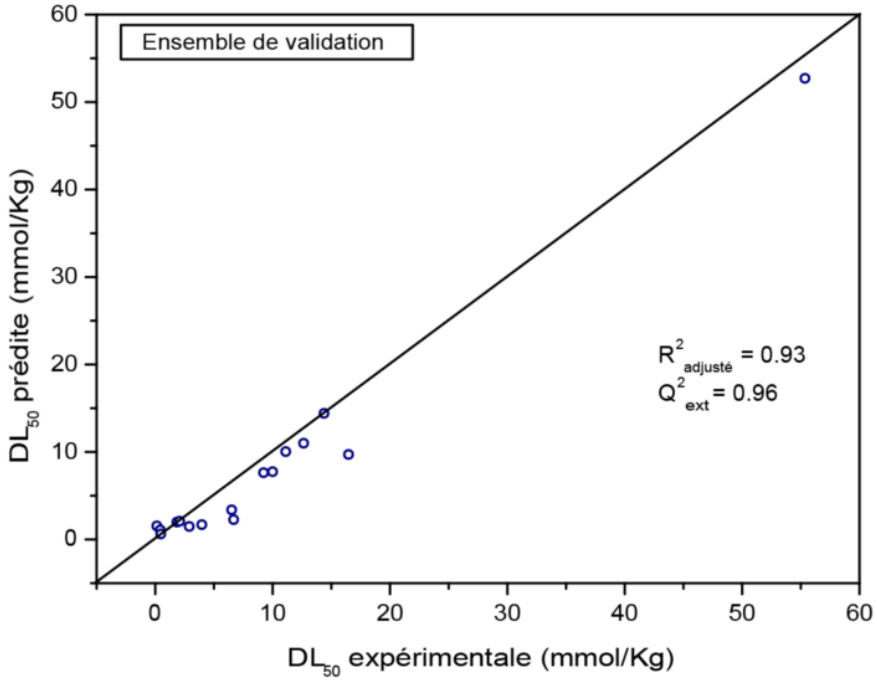


Figure 4.21 : Courbe de corrélation des valeurs prédites de DL_{50} (des fongicides) en fonction des valeurs expérimentales pour l'ensemble de validation

Tableau 4.16b : Valeurs de la toxicité (expérimentale et prédite) aigüe, des résidus normalisés et des points leviers des fongicides.

Fongicide	DL_{50} exp. (mmol/Kg)	DL_{50} préd. (mmol/Kg)	d_i	h_{ii}
Ensemble d'apprentissage				
Anilazine	16,590	12,100	1,275	0,019
Biphényle	13,880	14,580	-0,199	0,015
Biphenyle-2-ol	16,060	14,970	0,310	0,018
Boric acid	43,020	43,020	0,002	0,133
Carboxine	11,000	8,310	0,764	0,012
Chloraniformethane	7,430	7,140	0,082	0,011
Chlordecone	0,190	1,950	-0,502	0,016
Chloronebe	53,110	50,700	0,686	0,213
Chloroxylenol	24,460	24,400	0,016	0,039
Climbazole	1,370	4,190	-0,802	0,015
Cyprofuram	0,620	2,100	-0,420	0,016
Dazomet	2,560	2,360	0,057	0,013
Dichlone	0,700	1,980	-0,363	0,016
Diflumetorim	1,370	3,300	-0,549	0,015
Dimethomorphe	10,060	1,900	2,315	0,012
Diniconazole-M	1,450	6,230	-1,355	0,015
Dinocap	3,330	1,840	0,422	0,013
Dithianon	1,010	0,720	0,083	0,015
Epichlorohydrine	0,430	2,560	-0,604	0,016
Epoxiconazole	9,580	4,150	1,542	0,011
Etaconazole	4,090	2,680	0,400	0,012
Ethylene bisisothiocyanate sulphide	1,360	1,220	0,041	0,015
Fenapanil	6,280	3,520	0,783	0,011
Fenarimol	7,550	2,730	1,367	0,011
Fenoxanil	12,790	10,330	0,698	0,014
Fenpropidine	5,310	2,380	0,832	0,011
Fenpropimorphe	5,500	5,540	-0,011	0,011
Fluorure de sulfuryle	0,980	1,240	-0,075	0,015
Fluquinconazole	0,300	1,540	-0,353	0,016
Flusilazole	2,140	1,240	0,255	0,014
Flusulfamide	0,320	1,150	-0,237	0,016
Formaldehyde	18,320	17,300	0,289	0,022
Furalaxyl	3,120	4,060	-0,268	0,013
Furmecyclox	15,040	10,190	1,378	0,016
Glutaraldehyde	1,340	2,310	-0,275	0,015
Hexaconazole	6,970	5,970	0,283	0,011
Hymexazol	16,140	16,460	-0,093	0,018

Tableau 4.16b : Valeurs de la toxicité (expérimentale et prédite) aigüe, des résidus normalisés et des points leviers des fongicides (suite).

Fongicide	DL_{50} exp. (mmol/Kg)	DL_{50} préd. (mmol/Kg)	d_i	h_{ii}
Ensemble d'apprentissage				
Imazalil	0,760	2,930	-0,615	0,015
Inezin	2,460	1,590	0,247	0,014
Ipconazole	2,660	2,280	0,108	0,013
Iprobenfos	2,360	3,460	-0,314	0,014
Isoprothiolane	4,100	7,520	-0,972	0,012
Metalaxyl	2,270	5,470	-0,911	0,014
Metconazole	1,860	3,180	-0,374	0,014
Methasulfocarb	0,430	2,400	-0,559	0,016
Methfuroxam	6,410	11,020	-1,309	0,011
Metominostrobin	2,490	7,500	-1,422	0,013
Metsulfovax	16,910	11,020	1,673	0,020
Myclobutanil	5,540	2,630	0,826	0,011
Nitrothal isopropyl	21,670	21,500	0,048	0,030
Octhilinone	2,580	1,730	0,239	0,013
Oxycarboxine	6,110	5,590	0,146	0,011
Pentachlorophénol	0,300	2,960	-0,755	0,016
Peroxyde d'hydrogèn	44,630	43,460	0,332	0,145
Prochloraz	2,720	2,150	0,162	0,013
Propiconazol	2,800	2,470	0,094	0,013
Prothiocarb	6,830	4,320	0,713	0,011
Pyrazophos	0,400	5,000	-1,306	0,016
Pyrifénox	9,870	2,760	2,017	0,012
Pyrimethanil	20,840	20,790	0,016	0,028
Simeconazole	2,860	2,680	0,051	0,013
Spiroxamine	1,550	2,770	-0,347	0,014
Tebuconazole	4,820	2,400	0,689	0,012
Tecnazène	3,300	1,950	0,383	0,013
Thiabendazole	12,270	4,320	2,256	0,013
Thicyofene	1,630	3,300	-0,474	0,014
Tolfenpyrad	1,010	5,490	-1,272	0,015
Triacétate de guazatine	1,050	2,100	-0,300	0,015
Triadimefon	1,010	2,650	-0,465	0,015
Triadimenol	2,370	7,390	-1,426	0,014
Triazoxide	0,260	5,200	-1,402	0,016
Tricyclazole	0,970	4,220	-0,922	0,015
Tridemorphe	1,560	2,070	-0,144	0,014
Triflumizole	2,950	2,580	0,103	0,013
Uniconazole	1,470	5,600	-1,172	0,015

Tableau 4.16b : Valeurs de la toxicité (expérimentale et prédite) aigüe, des résidus normalisés et des points leviers des fongicides (suite et fin)

Fongicide	DL_{50} exp. (mmol/Kg)	DL_{50} préd. (mmol/Kg)	d_i	h_{ii}
Ensemble de validation				
Ampropylfos	14,380	14,410	-0,009	0,016
Benalaxil	2,090	2,080	0,002	0,014
Bupirimate	12,640	11,000	0,466	0,014
Dichlorophene	10,000	7,740	0,641	0,012
Dinobuton	0,430	2,000	-0,445	0,016
Ditalimfos	16,470	9,710	1,919	0,019
Edifenphos	0,480	0,620	-0,040	0,016
Ethoxyquine	11,120	10,040	0,306	0,012
Fenamidone	6,510	3,380	0,888	0,011
Hexachlorophène	0,140	13,540	-3,805	0,016
Kasugamycin	55,360	52,710	0,752	0,233
Nuarimol	3,970	1,690	0,649	0,012
Ofurace	9,230	1,090	2,311	0,011
Oxadixyl	6,680	2,260	1,255	0,011
Pyroquilone	1,850	1,980	-0,037	0,014
Tetraconazole	2,900	1,480	0,402	0,013

Pour déterminer la qualité de l'ajustement du modèle QSAR établi, une évaluation statistique est plus que nécessaire. Ainsi, la validation interne doit être effectuée pour évaluer la stabilité et la robustesse du modèle, tandis que la validation externe doit être mise en place pour évaluer son pouvoir prédictif. Sur le tableau 4.17, sont regroupées les valeurs des paramètres statistiques pour les deux ensembles (apprentissage et validation).

Tableau 4.17 : Valeurs des paramètres statistiques pour les ensembles d'apprentissage et de validation

Paramètres statistiques	Apprentissage	Validation
n	75	16
R	0.96	0.99
$\mathbf{R^2_{adjusté}}$	0.90	0.93
$\mathbf{Q^2}$	0.91	--
$\mathbf{Q^2_{ext}}$	--	0.96
$\mathbf{SDEP_{ext}}$	--	2.54

Une valeur de 0.91 pour Q^2 suggère une cohérence de l'ensemble d'apprentissage et une bonne validation interne. Par conséquent la robustesse du modèle a été confirmée. D'autre part, la grande valeur de Q^2_{ext} (0.96) préconise un bon pouvoir prédictif du modèle.

Par ailleurs, Les résidus pour les valeurs observées dans les ensembles d'apprentissage et de validation ont été déterminés et reproduits sur la figure 4.22 ci-dessous. Comme pour le cas des herbicides, le modèle n'a pas montré d'erreur proportionnelle et systématique, étant donné que la distribution des résidus des deux côtés de la ligne médiane est aléatoire.

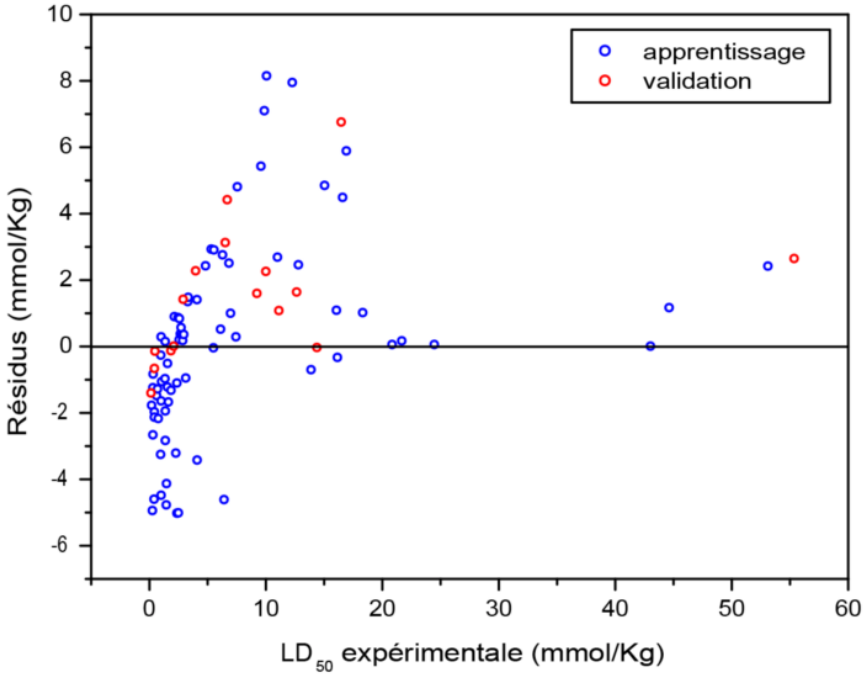


Figure 4.22 : Résidus des valeurs de DL_{50} obtenus pour les fongicides en fonction de la DL_{50} expérimentale

4.4.1.5 Domaine d'applicabilité du modèle

Comme ça a été le cas pour les deux classes de pesticides que sont les herbicides et les insecticides, le domaine d'applicabilité du modèle généré pour la prédiction de la toxicité orale aigüe des fongicides a été défini. Les valeurs des leviers des produits en question et celles de leurs résidus normalisés sont reportées sur le tableau 4.16 ci-dessus. Le diagramme de Williams ainsi obtenu est illustré sur la figure 4.23 ci-contre.

Le diagramme de Williams fait ressortir un domaine d'applicabilité à l'intérieur d'une zone rectangulaire avec une valeur limite de résidus égale à "±3s" et une valeur de levier critique

h* = 0.36. Tous les composés des deux ensembles d'apprentissage et de validation se trouvent à l'intérieur de cette zone, à l'exception d'une seule observation (Hexachlorophène) dans l'ensemble de validation qui s'est révélée avoir une valeur aberrante avec une valeur de levier plus petite que celle du levier critique. La globalité des fongicides (98.90%) dans le modèle se situe dans le domaine d'applicabilité ainsi défini, ce qui indique la fiabilité des prédictions.

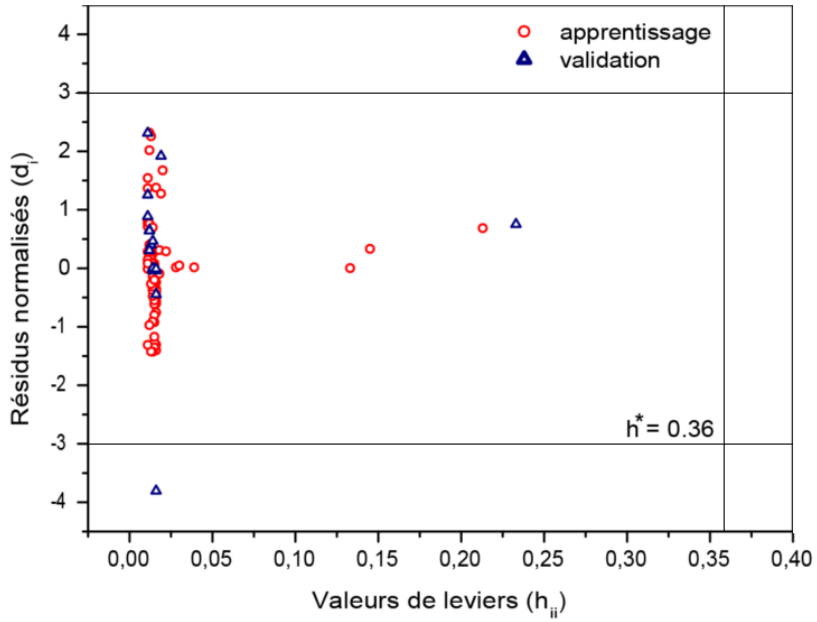


Figure 4.23 : Diagramme de Williams pour le modèle de prédiction de la toxicité des fongicides

4.4.1.6 Analyse de la contribution des descripteurs

Pour voir l'importance de chaque variable sur la prédiction de la toxicité aiguë, une analyse de sensibilité à l'aide du logiciel STATISTICA a été menée. Cette méthode fournit une quantification de l'importance relative des différentes entrées (variables) sur la sortie du réseau de neurones. L'importance des descripteurs du modèle RNA est indiquée sur la figure 4.24 ci-dessous. On notera que les descripteurs ayant une contribution conséquente dans la prédiction de la toxicité sont les suivants : Mor15m, Mor26u, Mor16u et nHDon. Par ailleurs, la contribution par catégorie de descripteurs est à l'actif de la catégorie 3D MoRSE Descriptors, suivie par RDF Descriptors et à un degré moindre celle appelée Functional Group Counts (Figure 4.25).

En plus des catégories de descripteurs déjà rencontrées lors de l'étude des herbicides et insecticides, deux autres catégories apparaissent dans l'étude sur les fongicides. La

première désignée sous le nom de "descripteur topologique" est celle à laquelle appartient le descripteur T(O…O). Il est défini comme étant la somme des distances entre les atomes d'oxygène dans la molécule (Todeschini et Consonni, 2000). La seconde catégorie nommée "propriétés moléculaires" est représentée par le descripteur TPSA(Tot). TPSA(Tot) représente la surface polaire des molécules contenant les atomes de phosphore et de soufre (comme les pesticides et les herbicides). Notons que TPSA (Tot) a toujours été inclus dans les modèles QSAR, car il s'est avéré pertinent pour la modélisation de la toxicité (Cassoti et al., 2014).

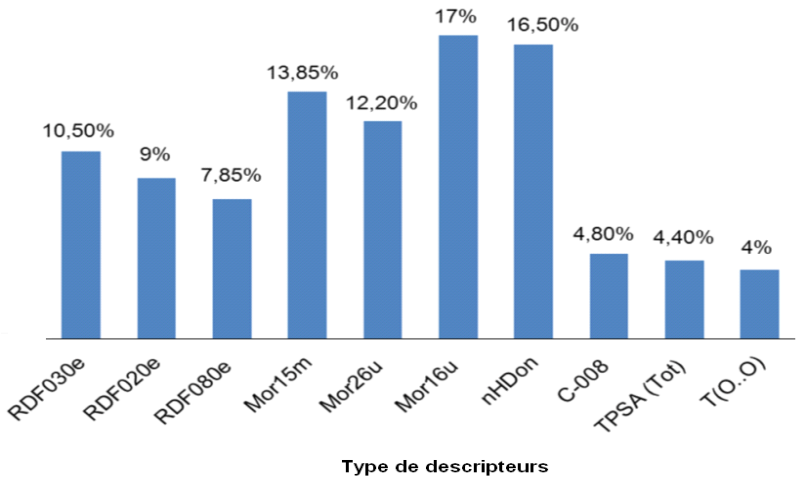


Figure 4.24 : Contribution (en %) à la toxicité orale aigüe des fongicides par type de descripteurs.

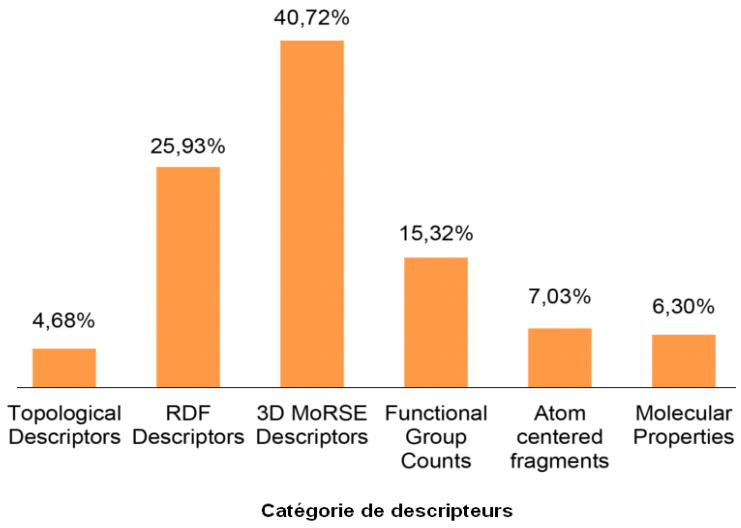


Figure 4.25 : Contribution (en %) à la toxicité orale aigüe des fongicides par catégorie de descripteurs.

4.4.2 Conclusion

Cette étude a abordé le problème de la prédiction de la toxicité. Un modèle QSAR conforme aux principes de l'OCDE a été développé pour prédire la toxicité orale aigüe des fongicides sur les rats. Les données sur la toxicité (DL_{50}) sur les rats ont été recueillies à partir de bases de données fiables et sûres. L'ensemble de données final a été répartis dans un ensemble d'apprentissage et un ensemble de validation externe. Le modèle QSAR final a montré de bonnes performances dans la prédiction (en apprentissage ou en validation). Une analyse du domaine d'applicabilité sur les deux ensembles d'apprentissage et de validation a montré que 98.90% de molécules d'insecticides se situent dans le domaine. Le modèle comprend dix descripteurs moléculaires qui recèlent des informations sur la zone polaire de la surface, l'électronégativité et la formation de liaisons H.

4.5 MODELISATION NEURONALE DE LA TOXICITE DES PESTICIDES.

Dans cette quatrième partie de cette thèse et après les résultats obtenus pour la prédiction de la toxicité aigüe des 3 classes de pesticides séparément, on a tenté de rechercher éventuellement un modèle de prédiction de cette toxicité par les réseaux de neurones et ce pour les pesticides pris dans leur ensemble. Pour ce faire, on a opté pour un modèle neuronal de type "perceptron multicouches : PMC" avec une couche d'entrée, une couche cachée et une couche de sortie.

Dans une première phase, nous présenterons les matériels et méthodes utilisés. Nous parlerons d'abord de la base de données collectées. Ensuite, nous présenterons la procédure de développement et d'optimisation du modèle neuronal de prédiction de la toxicité aigüe. Par la suite, la seconde phase de cette partie est réservée aux résultats obtenus. Aussi, nous présenterons l'architecture optimale du modèle neuronal, ses performances de prédiction, ainsi que son domaine d'applicabilité. Ensuite, on donnera l'expression mathématique du modèle optimisé ainsi que ses performances d'interpolation.

4.5.1 Résultats et discussion

4.5.1.1 Collecte des données et prétraitement

La toxicité orale aiguë chez le rat de 416 pesticides exprimée en dose létale (DL_{50} - mmol / kg) a été recueillie. Pour ce faire, la base des 416 pesticides a été scindée en 2 lots : le premier comportant 329 pesticides est dédiée à l'élaboration du modèle optimal (un ensemble d'apprentissage avec 258 pesticides et un deuxième pour la validation comportant 71 pesticides), le second constitué de 87 pesticides est laissé à des fins d'application de l'équation mathématique du modèle obtenu. Ces données de toxicité ont été collectées à partir des sources citées précédemment au paragraphe 4.1.1. Lors de la collecte de ces toxicités, nous avons tenu à ce qu'ils s'agissent de celles concernant les rats appartenant à la même souche et obtenues selon le même protocole expérimental.

Toutes les variables d'entrée et de sortie ont été normalisées dans l'intervalle [-1, 1] par utilisation de l'équation 4.12.

4.5.1.2 Obtention et selection des descripteurs

La procédure d'obtention et de sélection des descripteurs est la même que celle décrite aux paragraphes 4.1.2 et 4.1.3. Le nombre optimal de descripteurs retenus et qui donnent un

écart minimal entre les valeurs expérimentales et les valeurs calculées est de 17. Toutes les étapes de sélection des descripteurs sont reproduites sur le tableau 4.18 ci-dessous.

Tableau 4.18 : Etapes de la sélection des descripteurs pertinents

Nombre de descripteurs					
Nombre initial	Après 1ére sélection	Après 2éme sélection	Après 3éme sélection	Après Stepwise MLR	Après étude statistique (R^2 et REQM)
1664	620	275	95	29	17

Les 17 descripteurs sélectionnés appartiennent aux classes suivantes (Annexe C) : 2D Autocorrelations indices (MATS2p et MATS1m), Atom- centred fragments (N-072, H-046), Geometrical Descriptors (PJI3), Getaway Descriptors (H6m, HATSe et HATS0m), RDF descriptor (RDF030e et RDF020e), 3D-Morse descriptor (Mor15m, Mor23u et Mor26u), Whim Descriptors (Du et E1u), Functional group counts (nArX) et Constitutional descriptors (nS) .

Pour étudier sur la corrélation entre les descripteurs sélectionnés, la matrice de corrélation a été établie à l'aide du logiciel XLSTAT. Les résultats trouvés montrent que ces descripteurs ne sont pas corrélés entre eux du fait que la plus grande valeur du coefficient de corrélation est de 0.512. Les descripteurs sélectionnés ne présentent pas de multicolinéarité au vu des valeurs du VIF dont la plus grande valeur équivaut à 2.873.

4.5.1.3 Méthodologie d'élaboration du modèle neuronal optimal

L'objectif principal assigné à cette phase de l'étude est de trouver l'architecture du réseau de neurones optimale permettant de prédire la toxicité orale aigüe des pesticides sur les rats. L'utilisation d'une régression neuronale passe par le choix des paramètres d'entrée mais aussi par l'optimisation de l'architecture du réseau de neurones lui-même. L'optimisation à la fois de la répartition de la base de données, du nombre de couches cachées, du nombre de neurones par couche cachée, des fonctions de transfert (aussi bien pour les neurones cachés que les neurones de sortie) ainsi que des algorithmes a été faite après de nombreux essais. La conception du modèle neuronal consiste à évaluer les éléments

constitutifs du réseau en fonction de la performance de la modélisation souhaitée. Les performances du modèle sont évaluées, en termes d'erreur quadratique moyenne (*EQM*) (Lee et al., 1991) dont l'équation est la suivante :

$$EQM = \frac{\sum_{i=1}^{n}(y_i^{exp} - y_i^{pred})^2}{n} \qquad (4.13)$$

où n est le nombre total de données ; y^{pred} représente la toxicité prédite par le modèle neuronal et y^{exp} la valeur de la toxicité expérimentale.

a- Sélection de la répartition de la base de données

La première option est reliée à la répartition de la base de données en deux ensembles : un ensemble d'apprentissage et un ensemble de validation. Cinq types de répartition de la base de données ont été utilisés. Les résultats des cinq essais effectués sont consignés dans le tableau 4.19 et la figure 4.26.

Tableau 4.19 : Différents types de répartition de la base de données

	EQM				
	Essai 1	Essai 2	Essai 3	Essai 4	Essai 5
Base d'apprentissage : 77% Base de validation : 23%	0.041	0.038	0.040	0.041	0.039
Base d'apprentissage : 78% Base de validation : 22%	0.034	0.036	0.037	0.034	0.035
Base d'apprentissage : 78.5% Base de validation: 21.5%	**0.029**	**0.032**	**0.029**	**0.031**	**0.030**
Base d'apprentissage : 79% Base de validation: 21%	0.032	0.031	0.033	0.032	0.031
Base d'apprentissage : 80% Base de validation: 20%	0.033	0.034	0.031	0.035	0.033

D'après le calcul de l'erreur quadratique moyenne (Tableau 4.19), nous avons remarqué que la meilleure répartition est celle accordant 78.5% pour l'apprentissage et 21.5% pour la validation. Aussi, c'est cette répartition de la base de données qui sera utilisée dans la suite de nos travaux.

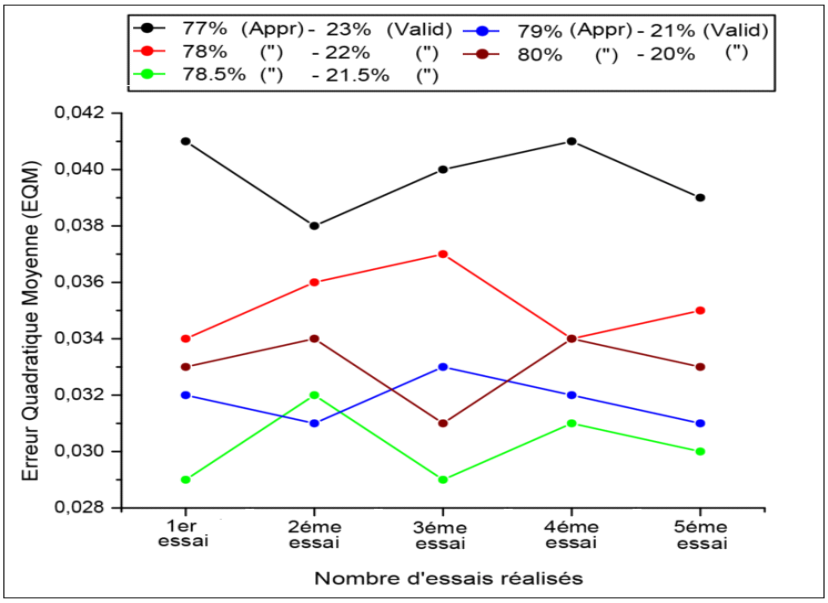


Figure 4.26 : Effet de la répartition de la base de données sur l'erreur quadratique moyenne

b-Sélection du nombre de couches cachées

La seconde optimisation consistait à sélectionner le nombre optimal de couches cachées. On a fait varier ce nombre de 1 à 2. Les résultats de l'EQM obtenus sont reportés sur le tableau 4.20.

Tableau 4.20 : Optimisation du nombre de couches cachées

	EQM				
	Essai 1	Essai 2	Essai 3	Essai 4	Essai 5
1 couche cachée	**0.0293**	**0.0291**	**0.0311**	**0.0310**	**0.0293**
2 couches cachées	0.0289	0.0295	0.0295	0.0288	0.0291

D'après les valeurs de l'EQM, il s'est avéré que le meilleur choix est d'utiliser une seule couche cachée par rapport à l'utilisation de 2 couches du fait que l'augmentation du nombre des couches cachées augmente la charge des calculs sans aucun gain substantiel de performance.

c- Sélection de la fonction d'activation

Plusieurs fonctions d'activation sont disponibles, dont les plus importantes sont :

- La fonction d'activation **linéaire**.
- la fonction **sigmoïde** ou la fonction log-sigmoïde.
- La fonction **tangente hyperbolique** ou bipolaire.

Pour connaître la fonction correspondante pour chaque couche, nous avons réalisé trois essais et testé différentes combinaisons dans la couche cachée et la couche de sortie du réseau en utilisant la meilleure répartition de la base de données avec une seule couche cachée (voir tableau 4.21).

Tableau 4.21 : Valeurs de l'EQM pour différentes combinaisons de fonctions d'activation

Couche cachée	Couche de sortie	EQM		
		Essai 1	Essai 2	Essai 3
Sigmoïde (Sigm)	Sigmoïde	0,02915	0,02934	0,02923
Sigmoïde	Linéaire	0,02943	0,02953	0,02938
Sigmoïde	Tangente hyperbolique	0,10542	0,10603	0,10597
Tangente hyperbolique (Tanh)	Sigmoïde	0,17196	0,57214	0,57310
Tangente hyperbolique	Linéaire	**0,02911**	**0,02910**	**0,02907**
Tangente hyperbolique	Tangente hyperbolique	0,02934	0,03012	0,02967
Linéaire (Lin)	Sigmoïde	0,15632	0,52563	0,54231
Linéaire	Tangente hyperbolique	0,03117	0,03096	0,03064
Linéaire	Linéaire	0,02997	0,03057	0,03071

Les résultats trouvés ont montré que la meilleure combinaison pour les fonctions d'activation dans respectivement la couche cachée et la couche de sortie est la suivante : Tangente hyperbolique-Linéaire. La figure 4.27 ci-dessous montre bien que cette dernière combinaison est celle qui fournit la plus petite erreur quadratique moyenne.

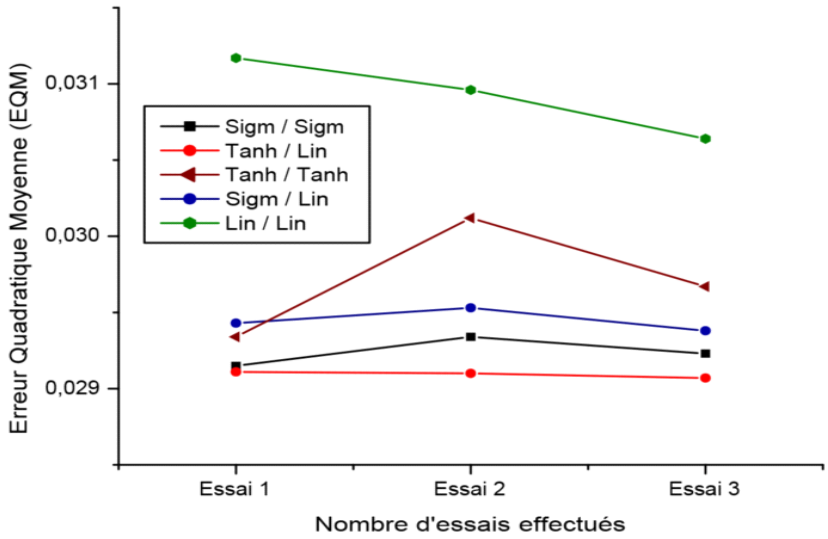


Figure 4.27 : Erreur quadratique moyenne pour différentes combinaisons des fonctions d'activation pour la couche cachée et la couche de sortie

d- Sélection du nombre de neurones dans la couche cachée

Pour définir le meilleur nombre de neurones dans la couche cachée, nous avons fait varier ce nombre entre un et quinze en prenant en compte les paramètres sélectionnés auparavant (meilleure répartition de la base de données, une couche cachée et comme fonction d'activation, la combinaison Tanh/Linéaire). Les résultats trouvés à l'issue de cette étude sont reproduits sur la figure 4.28 ci-dessous. Au vu de la figure 4.28 ci-dessus, le nombre de neurones donnant la plus faible EQM est de 9. Aussi, on optera pour ce nombre dans la suite de nos travaux.

e- Sélection de l'algorithme d'apprentissage optimum

Le choix de l'algorithme d'apprentissage est une autre tâche aussi compliquée que le choix des fonctions d'activation. En outre, il y a un grand nombre d'algorithmes d'apprentissages qui peuvent aboutir à l'obtention d'un modèle de RNA fiable. Afin de choisir lequel des algorithmes correspond le mieux à la prédiction de la toxicité, on a testé quatre algorithmes représentés dans le tableau 4.22 ci-dessous

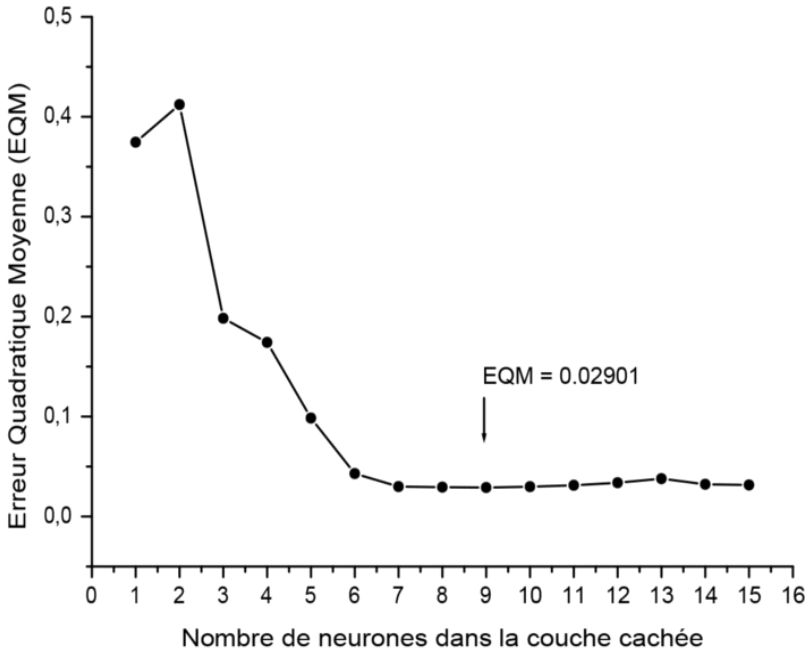


Figure 4.28 : Effet du nombre de neurones de la couche cachée sur l'erreur quadratique moyenne (EQM)

Tableau 4.22 : Algorithmes d'apprentissages utilisés.

Nom de l'algorithme d'apprentissage	Code Matlab	Acronyme
Rétropropagation de Quasi-Newton.	Trainbfg	BFGS
Rétropropagation de Levenberg –Marquardt.	Trainlm	LM
Rétropropagation de l'échelle de gradient conjugué.	Trainscg	SCG
Rétropropagation de Régularisation Bayésienne.	Trainbr	BR

Les résultats des 4 essais pour les quatre algorithmes utilisés sont illustrés sur la figure 4.29. Selon cette figure, on constate que l'algorithme d'apprentissage qui donne la meilleure performance en termes d'EQM minimale est l'algorithme de Rétropropagation de Quasi-Newton (BFGS).

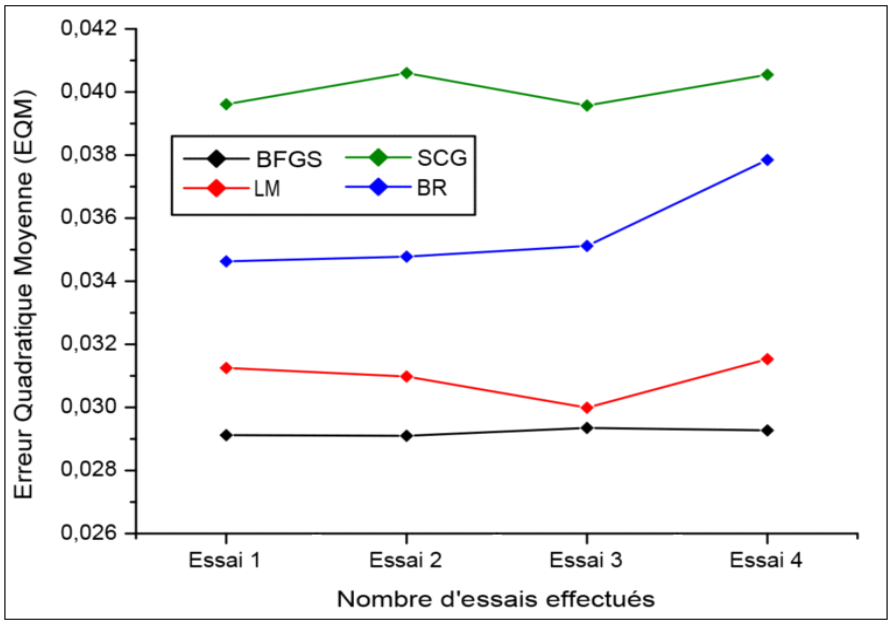


Figure 4.29 : Erreur quadratique moyenne obtenue par utilisation de 4 types d'algorithme d'apprentissage.

4.5.1.4 Structures du réseau neuronal optimum obtenu

Sur le tableau 4.23, sont récapitulés les différents éléments structurels du réseau optimal de modélisation de l'activité toxicologique (toxicité orale aigüe sur les rats) des pesticides.

Tableau 4.23 : Structure du RNA optimisé pour la toxicité aigüe des fongicides.

	Critère retenu pour cette étude
Architecture du Perceptron Multi-Couche (PMC)	Neurone d'entrée : 17 Neurone de sortie : 01 Neurones cachés : 9 (le plus petit nombre produisant la plus faible erreur de prédiction sur l'ensemble de validation
Fonction d'activation	
Couche cachée :	Tangente Hyperbolique (Fonction tansig de MATLAB®)
Couche de sortie :	Linéaire (Fonction *purelin* de MATLAB®)
Algorithme d'apprentissage	Rétropropagation de Quasi Newton (BFGS)
Critère d'arrêt de l'apprentissage	Lorsque la REQM sur l'ensemble de validation atteint un minimum
Nombre d'itérations dans la phase de rétropropagation	200
Découpages des ensembles de données	Apprentissage : 78.5% Validation : 21.5%

4.5.2 Application du réseau de neurone optimal

Une fois l'architecture du réseau de neurones connu, nous nous sommes attelés à la prédiction de la toxicité aigüe orale des pesticides sur les rats. Les résultats obtenus pour les ensembles d'apprentissage et de validation sont illustrés sur les figures 4.30 et 4.31 ci-contre.

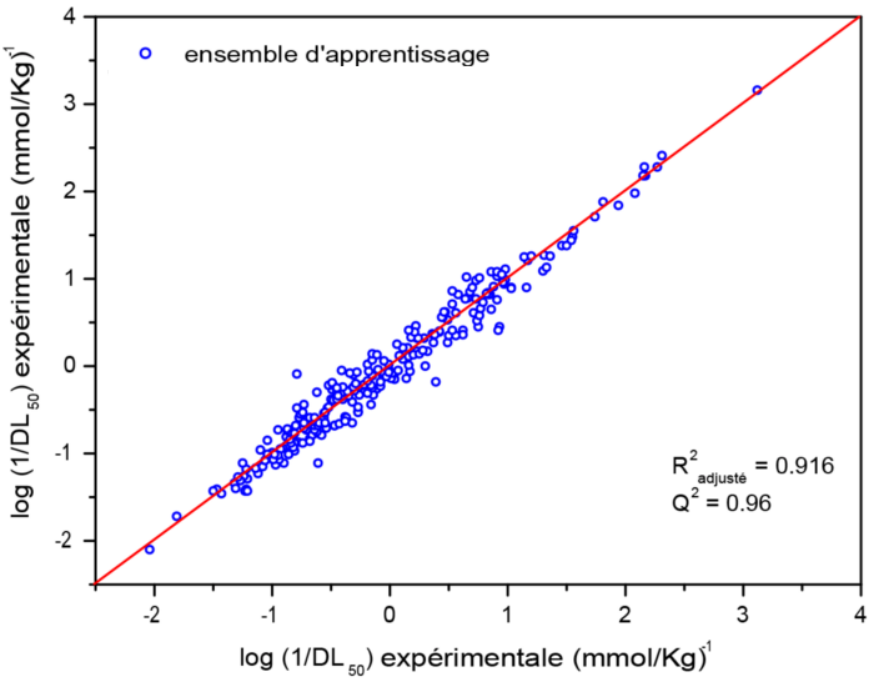


Figure 4.30 : Courbe de corrélation des valeurs prédites de log(1/DL_{50}) des pesticides en fonction des valeurs expérimentales pour l'ensemble d'apprentissage

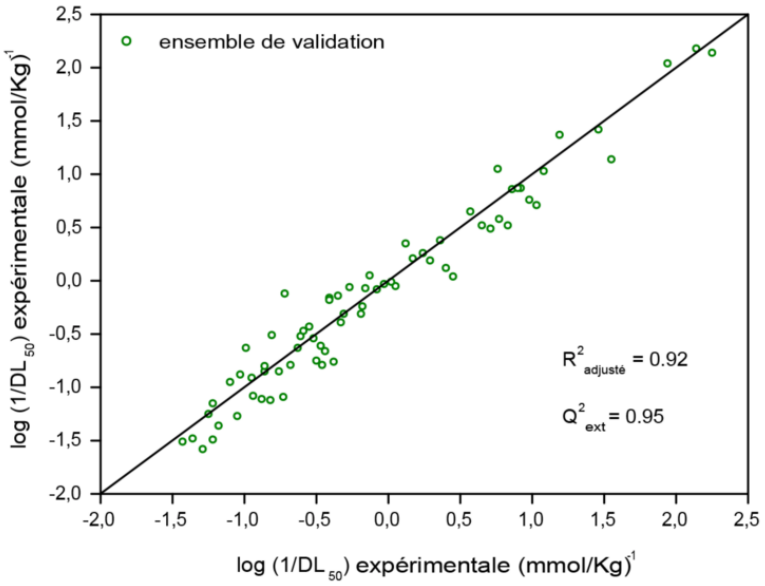


Figure 4.31 : Courbe de corrélation des valeurs prédites de log(1/DL_{50}) des pesticides en fonction des valeurs expérimentales pour l'ensemble de validation

L'architecture du réseau neuronal optimum obtenue suite aux différentes optimisations effectuées sur les éléments constitutifs d'un réseau de neurones est donnée sur la figure 4.32 ci-après.

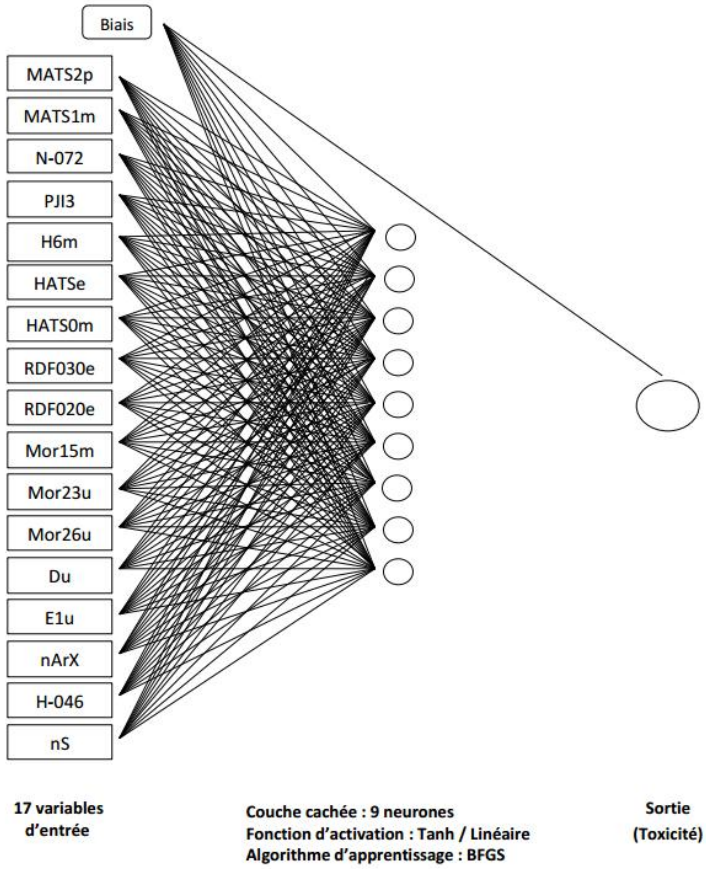


Figure 4.32 : Architecture finale du réseau optimal pour la prédiction de la toxicité orale aigüe des pesticides

Comme le montrent les figures 4.30 et 4.31, une bonne corrélation existe entre les valeurs expérimentales et prédites par le modèle. De plus, la qualité du modèle (robustesse et pouvoir prédictif) est estimée par les coefficients caractéristiques reportés sut le tableau 4.24 ci-dessous.

Tableau 4.24 : Valeurs des paramètres statistiques pour les ensembles d'apprentissage et de validation

Paramètres statistiques	Apprentissage	Validation
n	258	71
R	0.98	0.97
$R^2_{adjusté}$	0.92	0.92
Q^2	0.96	---
Q^2_{ext}	---	0.95
$SDEP_{ext}$	---	0.20

Une excellente valeur de 0.96 pour Q^2 suggère une cohérence de l'ensemble d'apprentissage et une bonne validation interne. Par conséquent la robustesse et la stabilité du modèle a été confirmée. D'autre part, la grande valeur de Q^2_{ext} (0.95) et la racine de l'erreur quadratique moyenne ($SDEP_{ext}$ = 0.20) préconise, elle aussi, un bon pouvoir prédictif du modèle. En outre, le tracé de la valeur résiduelle pour les valeurs prédites de log(1/DL_{50}) à la fois pour les ensembles d'apprentissage et de validation par rapport aux valeurs de log(1/DL_{50}) expérimentales est présenté sur la figure 4.33 ci-dessous. Comme on le voit, le modèle n'a pas montré d'erreur proportionnelle et systématique, puisque la propagation des résidus des deux côtés du zéro est aléatoire.

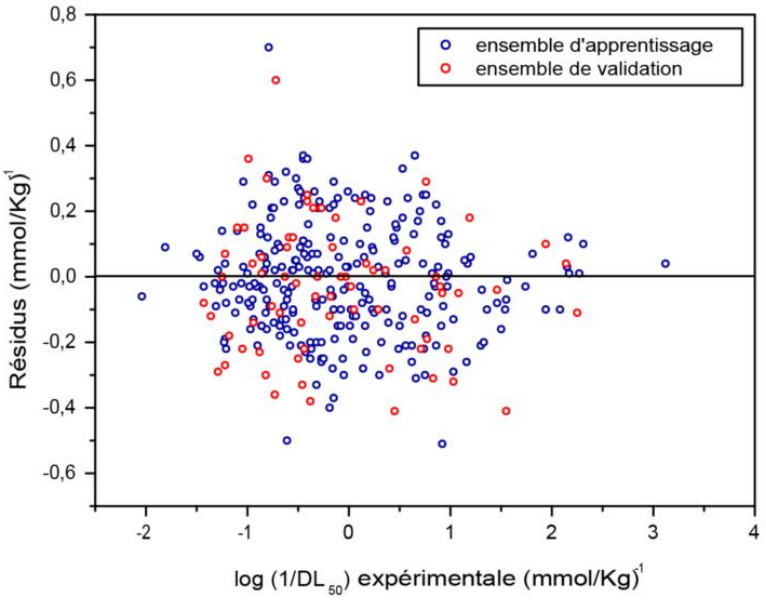


Figure 4.33 : Résidus des valeurs de log(1/DL_{50}) obtenus pour les pesticides en fonction de log(1/DL_{50}) expérimentale

4.5.3 Domaine d'applicabilité du modèle

Le domaine d'applicabilité du modèle généré pour la prédiction de la toxicité orale aigüe des pesticides a été défini. Les valeurs des leviers des produits en question et celles de leurs résidus normalisés sont reportées sur l'annexe B. Le diagramme de Williams ainsi obtenu est illustré sur la figure 4.34 ci-contre.

Le diagramme de Williams fait ressortir un domaine d'applicabilité à l'intérieur d'une zone rectangulaire avec une valeur limite de résidus égale à "±3s" et une valeur de levier critique h* = 0.16. Tous les composés des deux ensembles d'apprentissage et de validation se trouvent à l'intérieur de cette zone, à l'exception d'une seule observation (Hexachlorophène) dans l'ensemble de validation et de deux observations dans l'ensemble d'apprentissage qui se distinguent par des valeurs aberrantes mais avec une valeur de levier plus petite que celle du levier critique. La globalité des pesticides (99.10%) dans le modèle se situe dans le domaine d'applicabilité ainsi défini, ce qui témoigne de la fiabilité des prédictions et par la même de la robustesse et la performance du modèle neuronal établi.

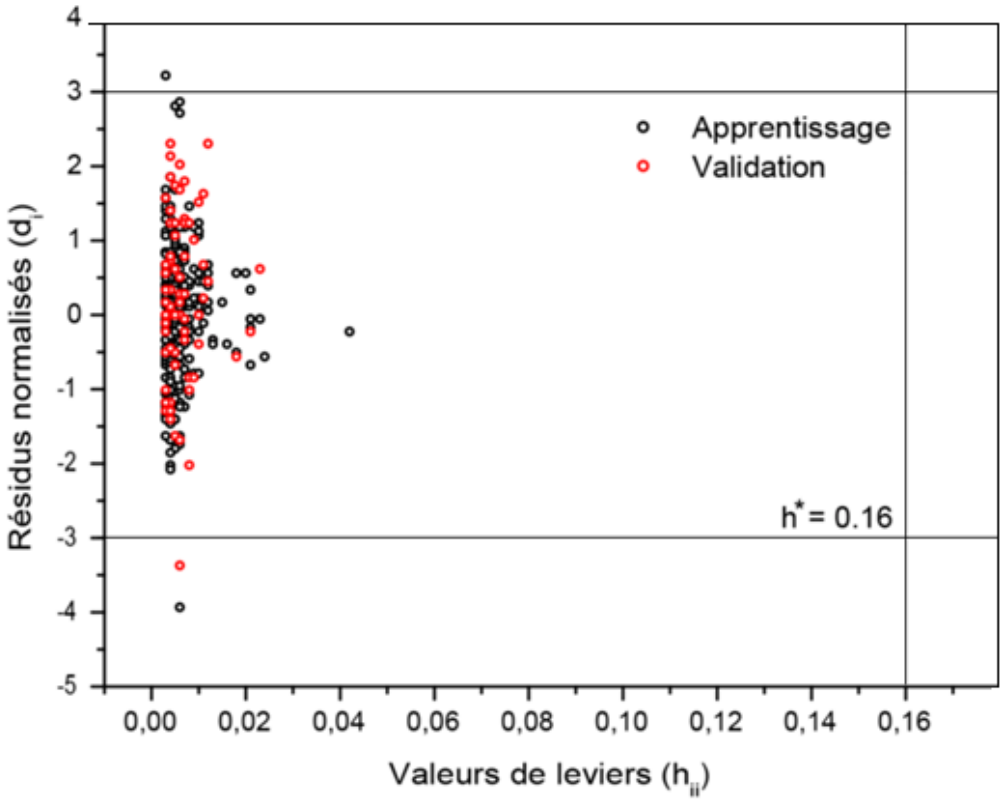


Figure 4.34 : Diagramme de Williams pour le modèle de prédiction de la toxicité des pesticides

4.5.4 EXPRESSION MATHEMATIQUE DU MODELE PMC-RNA OBTENU

L'architecture du réseau PMC-RNA optimisé est semblable à celle de la figure 4.11 concernant les herbicides. La seule différence réside dans le nombre des entrées (descripteurs) et celui du nombre de neurones dans la couche cachée. Les expressions mathématiques des sorties Z_j de la couche cachée et S de la couche de sortie sont données par les équations ci-après :

$$Z_j = f\left[\sum_{i=1}^{17} w_{ij}^I E_i + b_j^H\right] = \frac{exp\left(\sum_{i=1}^{17} w_{ij}^I E_i + b_j^H\right) - exp\left(-\sum_{i=1}^{17} w_{ij}^I E_i + b_j^H\right)}{exp\left(\sum_{i=1}^{17} w_{ij}^I E_i + b_j^H\right) + exp\left(-\sum_{i=1}^{17} w_{ij}^I E_i + b_j^H\right)} \quad (4.14)$$

$$S = f_0\left[\sum_{j=1}^{9} w_{1j}^H Z_j + b_1^0\right] = \sum_{j=1}^{9} w_{1j}^H Z_j + b_1^0 \quad (4.15)$$

4.5.5 Prédiction de la toxicité à partir de l'expression mathématique

Dans cette partie, nous avons voulu tester, une fois de plus, la performance du modèle neuronal optimisé par l'application de l'équation mathématique (4.15) décrivant ce modèle et ce à des fins de prédiction de la toxicité log(1/DL_{50}) de 87 pesticides non utilisés lors de la phase d'élaboration du modèle. Les valeurs prédites de la toxicité sont répertoriées dans le tableau 4.25 ci-dessous.

Tableau 4.25: Valeurs de log(1/DL_{50}) des pesticides prédites à partir du modèle mathématique

Pesticides	Type	log(1/DL_{50}) Expérimentale $(mmol/Kg)^{-1}$	log(1/DL_{50}) Prédite $(mmol/Kg)^{-1}$
Acethion	Insecticide	0,61	0,72
Acifluorfen	Herbicide	0,58	0,55
Alidochlore	Herbicide	0,22	0,11
Ametryne	Herbicide	-0,31	-0,42
Aramite	Insecticide	1,07	0,95
Atrazine	Herbicide	0,94	0,78
Bifenthrine	Insecticide	-0,89	-0,63
Bioallethrine	Insecticide	0,37	0,23
Biphenyle	Fongicide	1,14	1,27
Butroxydim	Herbicide	0,61	0,49
Cartap	Insecticide	0,07	0,17
Chlordimeforme	Insecticide	-0,09	-0,27
Chlorfenapyr	Insecticide	0,03	0,18

Tableau 4.25: Valeurs de log(1/DL_{50}) des pesticides prédites à partir du modèle Mathématique (suite)

Pesticides	Type	log(1/DL_{50}) Expérimentale (mmol/Kg)$^{-1}$	log(1/DL_{50}) Prédite (mmol/Kg)$^{-1}$
Chlorfensulphide	Insecticide	1,06	1,23
Chlorfenvinphos	Insecticide	-1,48	-1,38
Chloronebe	Fongicide	1,73	2,17
Chlorthiamide	Herbicide	0,57	0,42
Chlorthiophos	Insecticide	-1,56	-1,38
Cumyluron	Herbicide	0,50	0,68
Cycluron	Herbicide	0,88	0,89
Cyhalothrine	Insecticide	-0,51	-0,42
Demeton-o-methylsulfone	Insecticide	0,28	0,38
Diamidafos	Insecticide	-0,16	-0,09
Diazinon	Insecticide	0,57	0,69
Dicamba	Herbicide	0,85	0,68
Dichlofenthion	Insecticide	-0,26	-0,14
Dichlorophene	Fongicide	1,00	0,85
Dieldrine	Insecticide	-0,92	-1,23
Difenoconazole	Fongicide	0,55	0,71
Dimefox	Insecticide	-2,19	-2,52
Dimepiperate	Herbicide	0,56	0,74
Dimethipin	Herbicide	0,34	0,21
Diniconazole	Fongicide	0,16	0,22
Dioxacarbe	Insecticide	-0,95	-0,83
Disulfoton	Insecticide	-2,02	-1,62
Endothion	Insecticide	-1,09	-1,23
Epichlorohydrine	Insecticide	-0,36	-0,69
Epoxiconazole	Fongicide	0,98	1,36
Esfenvalerate	Insecticide	-0,68	-0,46
Ethion	Insecticide	-0,27	-0,13
Ethoprophos	Insecticide	-0,87	-1,02
Ethoxyquine	Fongicide	0,90	1,05
Fenazaflor	Insecticide	-0,19	-0,34
Fenpyroximate	Acaricide	-0,24	-0,33
Fenthion	Insecticide	-0,05	0,11
Fenuron	Herbicide	1,59	1,71
Fluazolate	Herbicide	1,05	1,14
Fluorure de sulfuryle	Insecticide	-0,01	0,08
Flusulfamide	Fongicide	-0,50	-1,09
Furametpyr	Fongicide	0,25	0,23
Furilazole	Fongicide	0,49	0,63
Haloxyfop	Herbicide	-0,03	-0,16

Tableau 4.25: Valeurs de log(1/DL_{50}) des pesticides prédites à partir du modèle mathématique (suite et fin)

Pesticides	Type	log(1/DL_{50}) Expérimentale $(mmol/Kg)^{-1}$	log(1/DL_{50}) Prédite $(mmol/Kg)^{-1}$
Indanofan	Herbicide	0,13	0,41
Iodofenphos	Insecticide	0,75	0,56
Ipconazole	Fongicide	0,42	0,63
Isazofos	Insecticide	-1,07	-1,37
Isobenzan	Insecticide	-1,93	-1,69
Methacrifos	Insecticide	0,45	0,32
Methamidophos	Insecticide	-0,67	-0,59
Methasulfocarb	Fongicide	-0,37	-0,32
Methiocarb	Insecticide	-1,07	-0,67
Mexacarbate	Insecticide	-1,20	-1,15
Monalide	Herbicide	1,04	1,14
Myclobutanil	Fongicide	0,74	0,54
N-nitrosodimethylamine	Insecticide	-0,31	-0,19
Nuarimol	Fongicide	0,60	0,83
Omethoate	Insecticide	-0,63	-0,74
Pentachlorophenol	Fongicide	-0,52	-0,65
Phorate	Insecticide	-2,11	-1,99
Phosnichlor	Insecticide	0,25	0,34
Phosphamidon	Insecticide	-1,63	-1,82
Prochloraz	Fongicide	0,43	0,55
Propachlore	Herbicide	0,41	0,35
Propafos	Insecticide	-0,70	-0,74
Prothiocarb	Fongicide	0,83	0,63
Pyraclofos	Insecticide	-0,18	-0,28
Pyridafenthion	Insecticide	0,35	0,19
Simeconazole	Fongicide	0,46	0,52
Tebupirimfos	Insecticide	-2,25	-2,37
Temephos	Insecticide	1,09	0,98
Thiocyclame	Insecticide	0,02	0,17
Triacetate de guazatine	Fongicide	0,02	0,11
Triadimefon	Fongicide	0,01	0,07
Triazophos	Insecticide	-0,57	-0,36
Triazoxide	Fongicide	-0,59	-0,86
Triclopyr	Herbicide	0,39	0,38
Zirame	Fongicide	-0,10	-0,08

A la lumière des résultats obtenus, nous remarquons que le modèle mathématique prédit de manière satisfaisante la toxicité des pesticides. Par ailleurs, le graphe des valeurs

prédites en fonction des valeurs expérimentales (figure 4.35) a été établi pour déterminer le meilleur ajustement linéaire. Ce dernier est donné par l'équation 4.16 ci-contre.

$$log(1/DL_{50})^{predite} = 0.015 + 1.017\ log(1/DL_{50})^{expérimentale} \quad (4.16)$$

avec R^2 = 0.96 ; écart type (ou SD) = 0.186 et p < 0.0001

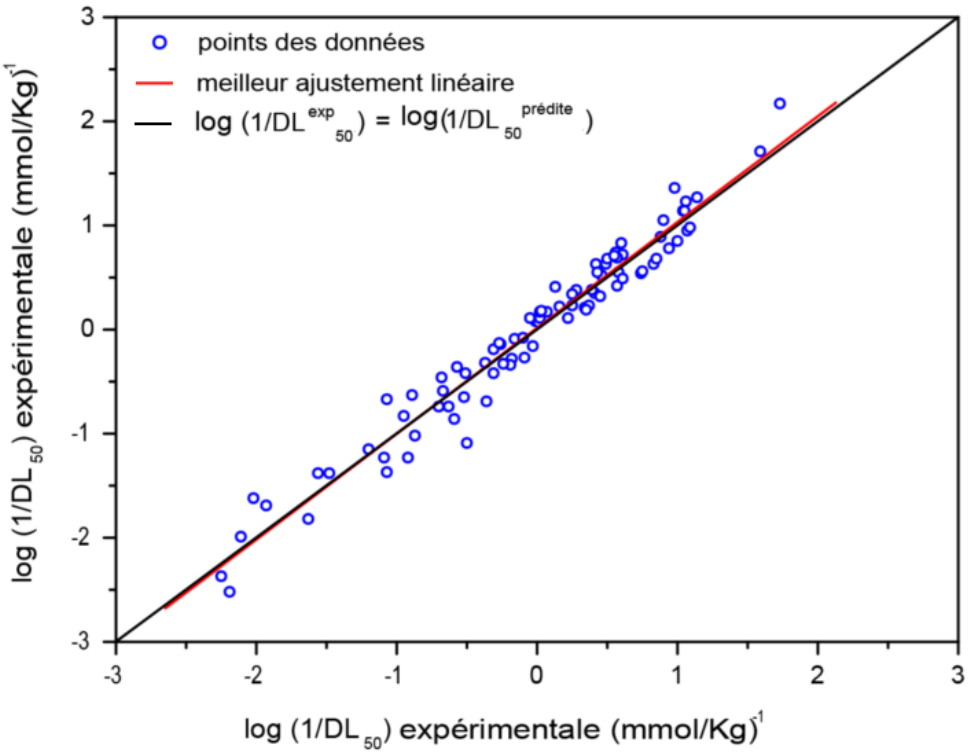


Figure 4.35 : Courbe de corrélation entre les valeurs expérimentales et celles prédites par le modèle mathématique.

4.5.6 Analyse de la contribution des descripteurs

Pour voir l'importance de chaque descripteur sur la prédiction de la toxicité aiguë, une analyse de leur contribution à l'aide du logiciel STATISTICA a été menée. Cette méthode fournit une quantification de l'importance relative des différentes entrées (descripteurs) sur la sortie du réseau de neurones. La contribution des descripteurs sélectionnés du modèle PMC-RNA des pesticides est indiquée sur la figure 4.36 ci-dessous.

Les descripteurs sélectionnés se répartissent selon quatre groupes de par leur contribution dans la prédiction de la toxicité. Le premier groupe est constitué du descripteur HATS0m qui contribue à hauteur de 12.81%. Notons que ce même descripteur possédait la plus grande contribution dans le modèle de prédiction élaboré pour les herbicides.

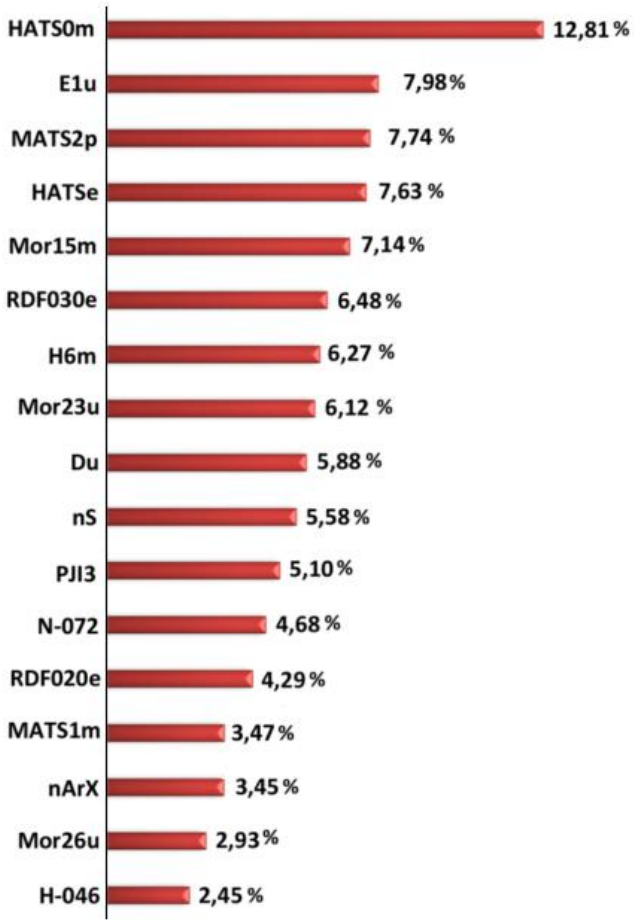


Figure 4.36 : Contribution relative des descripteurs dans la prédiction de la toxicité orale aigüe des pesticides

Le second groupe est constitué quant à lui de sept descripteurs (E1u, MATS2p, HATSe, Mor15m, RDF030e, H6m et Mor23u) dont la contribution dépasse les 6% pour chacun d'entre eux. La plupart de ces descripteurs ont été sélectionnés lors de l'établissement des modèles de prédiction pour les herbicides, les fongicides et les insecticides. Cinq descripteur (Du, nS, PJI3, N-072 et RDF020e) dont la contribution dépasse les 4% se retrouvent dans le troisième groupe. Enfin, un dernier groupe est constitué de quatre descripteurs (MATS1m, nArX, Mor26u et H-046) possédant la plus faible contribution relative.

Il est utile de rappeler que 75% des descripteurs sélectionnés du modèle établi pour les pesticides, l'ont été lors de l'établissement des 3 modèles précédent. Ce résultat appuie la thèse selon laquelle la toxicité orale aiguë (DL_{50}) pour les rats est régie essentiellement par trois la structure de la molécule, ses propriétés électroniques et un certain nombre d'atomes (soufre et halogènes, … etc.).

4.5.7 Conclusion

Dans cette partie de la thèse, la méthodologie QSAR a été utilisée pour développer un modèle de prédiction de la toxicité orale aigüe des pesticides sur les rats moyennant l'utilisation de descripteurs. Un modèle non linéaire de dix-sept descripteurs a été développé par les réseaux de neurones (perceptron multicouche). Les valeurs des paramètres statistiques et le domaine d'applicabilité pour les deux ensembles (d'apprentissage et validation) nous ont amené à constater la robustesse et le pouvoir prédictif du modèle élaboré. Des résultats encourageants de prédiction par utilisation de l'équation mathématique du modèle ont été obtenus. Le modèle proposé est prédictif et pourrait être utilisés pour estimer la toxicité orale aigüe de nouveaux composés pesticides sachant que les descripteurs impliqués peuvent être directement calculés à partir de la structure moléculaire.

4.6 COMPARAISON ENTRE MODELES

4.6.1 Modèles établis dans ce travail

Pour statuer sur le modèle le plus performant d'entre ceux obtenus successivement pour les herbicides, les insecticides, les fongicides et les pesticides, nous avons regroupé dans le tableau 4.26 ci-dessous, quelques paramètres statistiques ainsi que certaines données y afférentes.

Tableau 4.26 : Comparaison entre les 4 modèles de notre travail

		Modèle neuronal			
		Herbicides	Insecticides	Fongicides	Pesticides
Nombre de produits (n)		62	139	91	329
$R^2_{adjusté}$	apprentissage	0.89	0.96	0.90	0.91
	validation	0.95	0.93	0.93	0.92
Q^2		0.98	0.96	0.91	0.96
Q^2_{ext}		0.78	0.96	0.96	0.95
$SDEP_{ext}$		0.20	0.50	2.54	0.20
Nombre de descripteurs Pertinents (k)		12	8	10	17
Rapport : n / k		5.17	17.38	9.1	19.35
Intervalle des DL_{50} (mmol/Kg)		0.29 - 50.00	0.01 – 29.72	0.14 – 55.36	0.01 – 109.30
Pourcentage de molécules appartenant au domaine d'applicabilité		98.00 %	96.40 %	98.90 %	99.10 %

Les données du tableau 4.26 montrent clairement que le modèle de prédiction pour les pesticides est le plus fiable et ce pour plusieurs raisons :

- L'intervalle des toxicités est le plus large, ce qui peut permettre la prédiction par interpolation d'un grand nombre de nouvelles molécules.
- La racine de l'erreur quadratique moyenne pour la validation est la plus petite comparée aux 3 autres erreurs.
- Le rapport du nombre de molécules sur le nombre de descripteurs est le plus grand (19.35).

4.6.2 Comparaison avec les modèles de la littérature

Comme mentionné dans le chapitre 1, un nombre limité de modèles QSAR sur la prédiction de la toxicité orale aigüe des pesticides sur les rats ont déjà été mis au point. Certains d'entre eux étaient basés sur une classe de pesticides tels que les herbicides amides, les herbicides sulfonylurées ou les pesticides organophosphorés. D'autres couvrent différentes catégories de produits toxiques.

Pour comparer les performances du modèle qu'on a établi pour la prédiction de la toxicité orale aigüe des pesticides sur les rats, nous l'avons comparé avec quelques-uns des modèles parus dans la littérature scientifique. Les résultats sont présentés dans le tableau 4.27 ci-dessous.

Tableau 4.27 : Comparaison entre notre modèle et ceux parus dans la littérature

Composés étudiés	Nombre de composés	Méthode QSAR	Références	R^2
Pesticides	**329**	**RNA**	**Notre travail**	**0.96**
Herbicides	27	AR	Can et al. (2013)	0.93
Organophosphorés	62	AR	Garcia *et al*. (2007)	0.82
Organophosphorés	30	CoMFA	Guo *et al*. (2006)	0.90
Organophosphorés	67	RNA	Zahouily (2002)	0.93
Organophosphorés	54	RA / RNA	Eldred and Jurs (1999)	0.64 à 0.81
Herbicides amides	44	RNA	Zakarya *et al*. (1996)	0.90 à 0.92
Variés	525	AR	Enslein, (1978)	0.33

La seule donnée disponible sur les articles parus dans la littérature a trait malheureusement au seul coefficient de corrélation R^2**.** Néanmoins, les données disponibles sur ce tableau nous laisse penser que le modèle de prédiction de la toxicité des pesticides est apparemment plus précis et plus fiable que ceux parus jusqu'à présent.

CONCLUSION GENERALE ET PERSPECTIVES

L'humanité est, de nos jours, confrontée à la problématique de la pollution généralisée de son environnement par toute sorte de polluants. Les pesticides sont un cas particulier de ces polluants. Outre leur intérêt primordial en agriculture notamment, ils sont néanmoins nocifs de par leur toxicité aigüe ou chronique. Cette toxicité a engendré des problèmes aussi bien pour les humains que pour toutes les composantes de l'écosystème. D'ailleurs, la revue bibliographique élaborée dans le cadre de cette thèse est édifiante quant au danger réel que présente cette classe de produits chimiques. Aussi, la determination de la toxicité de ces produits est une condition essentielle avant leur mise sur le marché. Une fois que la toxicité est connue, la substance en question est soit interdite ou soit autorisée à être utilisée sous certaines conditions.

Dans ce contexte, l'objectif de cette thèse était de mettre au point un modèle statistique, de prédiction de la toxicité orale aigüe des pesticides sur les rats. Cette étude s'intègre dans le cadre de la quantification de cette toxicité sans recourir à l'expérimentation animale. Cette dernière, en plus du fait qu'elle est coûteuse et laborieuse, est décriée par une multitude d'organisations de défense des animaux. Pour atteindre cet objectif, les modèles QSAR ont été établis en utilisant deux méthodes d'apprentissage statistique : la régression linéaire multiple et les réseaux de neurones artificiels à 3 couches (une couche d'entrées, une couche cachée et une couche de sortie). Plusieurs modèles statistiques ont été générés suite à l'application de différents algorithmes d'apprentissage. Dans ce contexte, quatre axes de recherche ont été menés.

Le premier axe de notre travail concerne tout d'abord l'élaboration d'une base de données. Dans toute recherche de modèles QSAR, l'établissement d'une base de données fiable est l'une des clés de réussite. D'ailleurs, d'après la réglementation REACH, c'est la première des conditions à satisfaire pour avoir un modèle valide. Compte tenu de cette condition, nous avons élaboré une base de données avec comme fonction à modéliser la toxicité orale aigüe des pesticides sur les rats. Un soin particulier a été apporté au recueil des

données expérimentales de toxicité et ce en vérifiant que ces valeurs ont été obtenues dans les mêmes conditions expérimentales d'une part, et avec le même type d'animaux d'autre part. Au total, on a recueilli la toxicité aigüe de 416 pesticides appartenant à différentes classes. Le choix des variables nécessaires à la description de la toxicité a été fait sur la base du concept qui stipule que le phénomène à modéliser (toxicité) est intimement lié à la structure moléculaire. A cette fin, nous avons calculé 1664 variables ou descripteurs moléculaires pour chaque pesticide étudié.

Le second axe à être abordé a été l'élaboration d'un modèle de prédiction de la toxicité aigüe des herbicides. 62 herbicides et 12 descripteurs sélectionnés comme étant les plus pertinents (c'est-à-dire ceux qui sont susceptibles de décrire au mieux le phénomène de la toxicité) ont été utilisés pour développer deux modèles QSAR prédictifs, l'un linéaire et l'autre non linéaire. Dans un premier temps, on a opté pour la modélisation par régression linéaire multiple. Le modèle *RLM* obtenu pour la prédiction de la toxicité aiguë par voie orale des herbicides sur des rats est un modèle linéaire dont les coefficients statistiques sont les suivants : R^2 =0.855, $R^2_{adj.}$ =0.809, $R^2_{prediction}$ = 0.880, RMSE=0.270, F=18.664, $p < 0.0001$. Ces derniers, ajoutés à la concordance entre les valeurs prédites pour la LD_{50} et les valeurs expérimentales, nous amène à conclure que le modèle *RLM* a un bon pouvoir prédictif de la toxicité orale aigüe.

Une modélisation par utilisation des réseaux de neurones (*PMC-RNA*) a été menée à l'effet de développer un modèle non linéaire. Les réseaux de neurones possèdent une propriété remarquable pour la modélisation des processus non linéaires et sont devenus une technique de modélisation importante. Une fois que le choix des conditions de travail (répartition de la base de données en ensemble d'apprentissage et de validation, algorithme d'apprentissage, fonctions d'activation de la couche cachée) eut été fait, un modèle neuronal a été élaboré. Les paramètres statistiques, le domaine d'applicabilité et le calcul des toxicités par utilisation de l'équation mathématique du modèle nous a permis de conclure que ce modèle est fiable et possède une capacité prédictive intéressante.

Pour comparer à la fois les performances et la qualité de la prédiction des deux modèles obtenus dans ce travail (*RLM et PMC-RNA*), une évaluation statistique des deux modèles a été faite. Cette comparaison nous a permis de constater que la prédiction déterminée par le modèle *PMC-RNA* est meilleure que celle fournie par le modèle *RLM*. Une amélioration substantielle des paramètres statistiques pour le modèle *PMC-RNA* a été notée. Dans une seconde comparaison basée cette fois ci sur les résidus, on a constaté que ceux donnés par le modèle *PMC-RNA* sont plus près de la ligne médiane représentant l'égalité entre

la toxicité observée et celle prédite. Aussi, on peut donc conclure que le modèle *PMC-RNA* a un meilleur pouvoir prédictif que le modèle *RLM*. Cela signifie que le modèle obtenu avec le réseau de neurones permet, dans une large mesure, de confirmer qu'une relation non linéaire relie la toxicité aiguë des pesticides à leur structure moléculaire.

Le deuxième axe de notre travail a trait à la recherche d'un modèle neuronal de prédiction de la toxicité aigüe des insecticides sur les rats. Le choix du modèle non linéaire est motivé par le résultat du premier axe et par les données bibliographiques qui consacrent la supériorité des modèles neuronaux par rapport aux modèles de régression multiple. A cet effet, 139 molécules ont été utilisées et 8 descripteurs furent sélectionnés comme étant les plus pertinents. Le modèle ainsi obtenu a fourni des résultats encourageants avec un coefficient de prédiction Q^2_{ext} = 0.96. Par ailleurs, l'analyse du domaine d'applicabilité de ce modèle nous a permis de conclure que la toxicité de plus de 96% a été correctement prédite. Tous les résultats obtenus nous amène à considérer que ce modèle a donné des résultats fiables et qu'il est doté d'un pouvoir prédictif plus que satisfaisants.

L'élaboration d'un modèle de prédiction de la toxicité orale aigüe des fongicides a été le but du troisième axe de notre travail. Le modèle neuronal qui a été obtenu avait comme base de données 91 molécules de fongicide et 10 descripteurs pertinents. La valeur de Q^2 = 0.90 pour l'ensemble d'apprentissage suggère une cohérence de l'ensemble d'apprentissage et une bonne validation interne. Par ailleurs, le bon pouvoir prédictif du modèle et sa robustesse a été confirmée par l'excellente valeur de Q^2_{ext} (0.96). D'autre part, l'étude des résidus a confirmé que le modèle n'a pas montré d'erreur proportionnelle et systématique. Le domaine d'applicabilité, qui a été défini, a montré que plus de 98% de valeurs de toxicité ont été correctement prédites.

Les résultats obtenus pour les trois familles de pesticides (herbicides, insecticides et fongicides) nous ont amené à essayer de développer un modèle neuronal qui puisse prédire la toxicité des pesticides en général. Ce fut le quatrième et dernier axe de notre recherche. Pour ce faire, on a utilisé 416 pesticides et 17 descripteurs sélectionnés parmi les 1664 calculés. Contrairement aux 3 précédents modèles, celui des pesticides a été construit suite à une série d'optimisations sur tous les éléments entrant dans l'élaboration d'un modèle. Le modèle final ainsi obtenu a donné des résultats très encourageants. En effet, 99% des 329 pesticides utilisés pour l'apprentissage et la validation se situe dans le domaine d'applicabilité de ce modèle. Par comparaison avec les 3 modèles précédents, celui des pesticides est plus performant. De plus, la fiabilité du modèle a été confirmée par le calcul, à partir de l'équation mathématique

dérivant de ce modèle. Les valeurs des descripteurs de 87 molécules, ne figurait ni dans l'ensemble d'apprentissage ni dans celui de la validation ont été injectées dans l'équation mathématique du modèle donnant la toxicité en fonction des variables (descripteurs dans notre cas). La comparaison des valeurs de toxicité calculées avec celles obtenues expérimentalement fait ressortir une très bonne corrélation. Plus de 90% des valeurs ont été prédites avec des erreurs très minimes.

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

- ✓ Les résultats auxquels on a abouti sont meilleurs que ceux parus jusqu'à présent dans les publications scientifiques. Dans les articles traitant la prédiction de la toxicité, l'établissement du modèle non linéaire a été effectué par utilisation d'un seul algorithme d'apprentissage, d'une seule fonction de transfert, d'un nombre restreint de descripteurs et de molécules pour la plupart appartenant à la même famille chimique. Dans le cadre de cette thèse, on a utilisé plusieurs familles de pesticides, un très grand nombre de descripteurs et quatre algorithmes d'apprentissage.
- ✓ L'expression mathématique du modèle neuronal optimisé a été établie. Cet état de fait nous a permis de valider ce modèle et ce par l'utilisation d'un ensemble de pesticides n'ayant pas été utilisé lors de son établissement. Cette validation a donné des résultats de prédiction prometteurs.
- ✓ la toxicité orale aiguë (DL_{50}) est influencée principalement par trois facteurs importants : la structure de la molécule, les propriétés électroniques et le nombre de certains atomes (soufre, halogènes et oxygène).
- ✓ Le but qu'on s'est assigné a été atteint, car la modélisation par réseaux de neurones artificiels peut effectivement servir d'alternative aux méthodes expérimentales de détermination de la toxicité des molécules.

En conclusion, tout en estimant avoir atteint l'objectif tracé pour cette étude, de nombreuses voies de recherche ont émergé. Ces perspectives concernent plus particulièrement la prédiction d'autres phénomènes biologiques. D'autre part, il faudrait s'intéresser à d'autres méthodes d'apprentissage statistique pour voir s'il est possible d'améliorer le modèle de prédiction de la toxicité des pesticides.

REFERENCES BIBLIOGRAPHIQUES

Adamson, G.W., Bawden, D. et Saggers, D.T. (1984). "Quantitative structure–activity relationship studies of acute toxicity (LD50) in a large series of herbicidal benzimidazoles". *Pestic. Sci.* 15: 31–39.

Agrawal, A., Pandey, R.S., Sharma, B. (2010). "Water pollution with special reference to pesticide contamination in India". *Journal of Water Resources and Protection* 2:432-448.

Ahouangninou, C., Thibaud, M., Edorh, P., Sahabi Bio-Bangana, Onil S., Louis St-L., Dion, S. et Fayomi, B. (2012). "Characterization of Health and Environmental Risks of Pesticide Use in Market-Gardening in the Rural City of Tori-Bossito in Benin, West Africa". *Journal of Environmental Protection*, 3:241-248

AIRPARIF : "Surveillance de la qualité de l'air en Ile de France. Evaluation des concentrations en pesticides dans l'air ambiant francilien: campagne exploratoire". Juin 2007.

Akan, J.C., Mohammed, Z., Jafiya, L. et Ogugbuaja, V.O. (2013) "Organochlorine Pesticide Residues in Fish Samples from Alau Dam, Borno State, North Eastern Nigeria". *J Environ. Anal. Toxicol* 3: 171. doi:10.4172/2161-0525.1000171.

Alix, A., Barriuso, E., Bedos, C., Bonicelli, B., Caquet, T., Dubus, I., Gascuel, C., Gril, J.J.et Voltz, M. (2005) "Expertise scientifique collective "Pesticides, agriculture et environnement" Chapitre 3 : Devenir et transfert des pesticides dans l'environnement et impacts biologiques". INRA.

Andersen, H.R., Ida M. Schmidt, Philippe Grandjean, Tina K. Jensen, Esben Budtz-Jørgensen, Mia B. Kjærstad, Jesper Bælum, Jesper B. Nielsen, Niels E. Skakkebæk, Katharina M. Main. (2008) "Impaired Reproductive Development in Sons of Women Occupationally Exposed to Pesticides during Pregnancy", *Environmental Health Perspectives*. 116(4): 566–572.

Arzi A, Hemmati A.A, Nazari Khorasgani Z. (2011). "Determination and comparison of the organochlorine pesticide residue levels among benni fish of Shadegan, mahshahr and susangerd cities, Khozestan province in Iran". *Jundishapur Journal of Natural Pharmaceutical Products* 6(1): 24-31

Asman, W.A.H. et al. (2005). "Wet deposition of pesticides and nitrophenols at two sites in Denmark: measurements and contributions from regional sources". *Chemosphere*, 59: 1023-1031.

Bakhtiyor, R., Kusic, H., Leszczynska, D., Leszczynski, J. and Koprivanac, N. (2010). "QSAR modeling of acute toxicity on mammals caused by aromatic compounds: the case study using oral LD50 for rats". *J. Environ. Monit.* 12: 1037.

Baldi I, Lebailly P, Rondeau V, Bouchart V, Blanc-Lapierre A, Bouvier G, Canal-Raffin M, Garrigou A (2012). "Levels and determinants of pesticide exposure in operators involved in treatment of vineyards: results of the PESTEXPO Study". *Journal of exposure science and environmental epidemiology* **22**(6): 593-600.

Baldi Isabelle, Pierre Lebailly, Brahim Mohammed-Brahim, Luc Letenneur, Jean-François Dartigues, and Patrick Brochard. (2003) "Neurodegenerative Diseases and Exposure to Pesticides in the Elderly". *American Journal of Epidemiology*. 157:409–414

Baldi, I (1998). "Effets retardés des pesticides sur la santé : état des connaissances épidémiologiques". *Rev.Epidem. et Santé Publ.* 46: 134-142.

Band, P.R., Abanto, Z., Bert, J., Lang, B., Fang, R., Gallagher, R.P., Le, N.D. (2011). "Prostate cancer risk and exposure to pesticides in British Columbia farmers". *Prostate* 71(2):168–183.

Barjhoux Iris. (2011) thèse de Doctorat "Étude de la biodisponibilité et de la toxicité de polluants chimiques à risque dans les sédiments aquatiques vis-à-vis des premiers stades de développement d'un poisson modèle, *Oryzias latipes*", Université Bordeaux 1, France.

Bedos, C., Cellier, P., Calvet, R. and Barriuso, E. (2002). "Occurrence of pesticides in the atmosphere in France". *Agronomie*. 22: 35-49.

Bempah Crentsil Kofi, Jacob Asomaning, Juliana Boateng. (2012). "Market basket survey for some pesticides residues in fruits and vegetables from Ghana". *Journal of Microbiology, Biotechnology and Food Sciences*. 2 (3): 850-871

Bhhatarai, B., and Gramatica, P., (2011). "Oral LD50 toxicity modeling and prediction of per- and polyfluorinated chemicals on rat and mouse". *Mol. Divers*. 15: 467.

Blanchoud H., Farrugia, F. & Mouchel J.M. (2004). "Pesticide uses and transfers in urbanised catchments". *Chemosphere*. 55: 905-913

Borges, E. (2012). "Modeling physical and toxicity endpoints of alkyl (1-phenylsulfonyl) cycloalkane carboxylates using the Ordered Predictors Selection (OPS) for variable selection and descriptors derived with SMILES". *Chemometr. Intell. Lab.* 118: 79.

Bottoni, P. P. Grenni , L. Lucentini , A. Barra Caracciolo (2013). "Terbuthylazine and other triazines in Italian water resources". *Microchemical Journal* 107: 136–142.

Bouvier G, Blanchard O, Momas I, Seta N. (2006). "Pesticide exposure of non-occupationally exposed subjects compared to some occupational exposure: A French pilot study". *Science of the total environment* 366(1): 74-91.

Bouziani Mustapha, (2007). "L'usage immodéré des pesticides : De graves conséquences sanitaires". *santemaghreb.com*

Briand, O. (2003). "Influence des facteurs environnementaux et des pratiques agricoles sur les variations spatio-temporelles des niveaux de contamination de l'atmosphère par les pesticides". Thèse Rennes 1, 297 pp.

Brophy James T. ; Margaret M. Keith; Kevin M. Gorey; Ethan Laukkanen; Deborah Hellyer; Andrew Watterson; Abraham Reinhartz; Michael Gilberston. (2002). "Occupational Histories of Cancer Patients in a Canadian Cancer Treatment Center and the Generated Hypothesis Regarding Breast Cancer and Farming". *International Journal of Occupational and Environmental Health.* 8:346-353.

Cabidoche, Y. M., Lesueur-Jannoyer, M. (2012). "Contamination of Harvested Organs in Root Crops Grown on Chlordecone-Polluted Soils". *Pedosphere* 22(4): 562–571.

Can A, Yildiz I, Guvendik G. (2013). "The determination of toxicities of sulphonylurea and phenylurea herbicides with quantitative structure-toxicity relationship (QSTR) studies". *Environ. Toxicol Pharmacol.* 35(3):369-79.

Carine Floch, Anne-Céline Chevremont , Karine Joanico , Yvan Capowiez , Stéven Criquet.(2011). "Indicators of pesticide contamination: Soil enzyme compared to functional diversity of bacterial communities via Biolog. Ecoplates". *European Journal of Soil Biology* 47 : 256-263

Cassotti, M., Ballabio, D., Consonni, V., Mauri, A., Tetko, I.V., Todeschini, R. (2014) "Prediction of Acute Aquatic Toxicity Toward Daphnia magna by using the GA kNN Method" ATLA 42, 31–41.

Chevrier Cécile, Gwendolina Limon, Christine Monfort, Florence Rouget, Ronan Garlantézec, Claire Petit, Gaël Durand, Sylvaine Cordier.(2011). "Urinary Biomarkers of Prenatal Atrazine Exposure and Adverse Birth Outcomes in the PELAGIE Birth Cohort". *Environmental Health Perspectives* 119:1034-1041.

Chrisman Juliana de Rezende, Sergio Koifman, Paula de Novaes Sarcinelli, Josino Costa Moreira, Rosalina Jorge Koifman, Armando Meyer.(2009). "Pesticide sales and adult male cancer mortality in Brazil". *Int. J. Hyg. Environ. Health* 212 : 310–321

Christiansen S., Boberg J., Nelleman C., Axelstad M., Jacobsen P.R. and Hass U. (2010). "A cocktail of endocrine disrupting pesticides affects sexual differentiation in rats". Free communication / *Reproductive Toxicology*, 30: 229.

Chunfa Wu, Yongming Luo, Tong Gui, Yujuan Huang (2014). "Concentrations and potential health hazards of organochlorine pesticides in shallow groundwater of Taihu Lake region, China". *Science of the Total Environment* 470–471: 1047–1055.

COM 2010 – "Sixième rapport sur les statistiques concernant le nombre d'animaux utilisés à des fins expérimentales et à d'autres fins scientifiques dans les États membres de l'Union européenne". Pages 4-7.

Confais, J. Le Guen, M. (2006) "Premiers pas en regression linéaire avec SAS". Revue MODULAD, Numéro 35

Coscollà C., Colin P., Yahyaoui A., Petrique O., Yusà V., Mellouki A. and Pastor A. (2010). "Occurrence of currently used pesticides in ambient air of Centre Region (France)". *Atmospheric Environment*. 44: 3915-3925.

Costello Sadie, Myles Cockburn, Jeff Bronstein, Xinbo Zhang, and Beate Ritz.(2009). "Parkinson's Disease and Residential Exposure to Maneb and Paraquat From Agricultural Applications in the Central Valley of California". *American Journal of Epidemiology* 169:919–926

Coupe R.H. and Blomquist J.D. (2004). "Water-soluble pesticides in finished water of community water supplies". *J. AWWA*. 96: 56-58.

Dalvie Mohamed Aqiel, London Leslie. (2009) "Risk assessment of pesticide residues in South African raw wheat". *Crop Protection* 28: 864–869

Dan Liu, Shungeng Min. (2012) "Rapid analysis of organochlorine and pyrethroid pesticides in tea samples by directly suspended droplet microextraction using a gas

chromatography–electron capture detector". *Journal of Chromatography A*. 1235: 166– 173

Davodi Morteza, Abbas Esmaili-Sari, Nader Bahramifarr. (2011) "Concentration of polychlorinated biphenyls and organochlorine pesticides in some edible fish species from the Shadegan Marshes (Iran)". *Ecotoxicology and Environmental Safety* 74: 294–300.

Devillers, J. (2004). "Prediction of mammalian toxicity of organophosphorus pesticides from QSTR modeling". *SAR QSAR Environ. Res.* 15: 501–510.

Devillers, J. (2013) "Computational Toxicology: Volume II, Methods in Molecular Biology". Brad Reisfeld and Arthur N. Mayeno (eds.), *Springer Science* vol. 930, DOI 10.1007/978-1-62703-059-5_1

Dorothee Provost, Anne Cantagrel, Pierre Lebailly, Anne Jaffre, Veronique Loyant, Hugues Loiseau, Anne Vital, Patrick Brochard, Isabelle Baldi. (2007). "Brain tumours and exposure to pesticides: a case–control study in south western France" . *Occup Environ Med.* 64:509–514

Duchowicz, P. R., Marrugo, J. Erlinda, J. H., Ortiz, V., Castro, E. A. and Vivas-Reyes, R. (2009). "Qsar study for the fish toxicity of benzene derivatives". *J.Argentine Chem. Soc.* 97(2): 116.

Dupupet JL, Adjemian A, Grillet JP, Garnier R (2010) "Etude d'ergoexpologie aux fongicides dithiocarbamates auprès de professionnels de trois secteurs agricoles". *Archives des maladies professionnelles et de l'environnement* 71(4) : 638-643.

Eldred, D.V. et P.C. Jurs,(1999). "Prediction of acute mammalian toxicity of organophosphorus pesticide compounds from molecular structure". *SAR QSAR Environ. Res.* 10: 75–99.

Eriksson, L. et al.(2003). " Methods for reliability and uncertainty assessment and for applicability evaluations of classification and regression-based QSARs ". Environ. Health Perspect. 111, 1361–1375.

FAO organisation des nations unies pour l'alimentation et l'agriculture. Site statistique http://www.fao.org/statistics/fr/ consulté le 04 Mai 2014.

Farajzadeh Mir Ali, Leila Khoshmaram , Ali Akbar Alizadeh Nabil (2014). "Determination of pyrethroid pesticides residues in vegetable oils using liquid–liquid extraction and dispersive liquid–liquid microextraction followed by gas chromatography–flame ionization detection". *J. Food Compos. Anal.* 34(2) : 128-135.

Fianko Joseph R., Augustine Donkor, Samuel T. Lowor, Paul O. Yeboah, Eric T. Glover, Theodosia Adom, Augustine Faanu. (2011). "Health Risk Associated with Pesticide Contamination of Fish from the Densu River Basin in Ghana". *Journal of Environmental Protection*. 2: 115-123

Fischer W.J., B. Schilter, A.M. Tritscher, R.H. Stadler (2011). "Contaminants of milk and dairy products: Contamination Resulting from Farm and Dairy Practices" *Encyclopedia of Dairy Sciences* (Second Edition). Pages 887-897

Freire, C. et Koifman, S. (2012). "Pesticide exposure and Parkinson's disease: epidemiological evidence of association". *Neurotoxicology* 33 (5): 947–971.

Galiulin R.V., Bashkin V.N., Galiulina R.A. et Kucharski R. (2002). "Behavior of persistent organic pollutants in the airplant-soil system". *Water Air and Soil* Pollution, 37:179-191

Garcıa-Domenech, R., P. Alarcon-Elbal, G. Bolas, R. Bueno-Marı, F.A. Chorda-Olmos, S.A. Delcour, M.C. Mourino, A. Vidal, and J. Galvez. (2007). "Prediction of acute toxicity of organophosphorus pesticides using topological indices" *SAR QSAR Environ. Res*. 18 : 745–755.

Garson, G.D. (1991). "Interpreting neural network connection weights". *Artif. Intell. Expert.* 6(4): 47.

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

Goh, A.T.C. (1995). "Back-propagation neural networks for modeling complex systems". *Artif. Intell. Eng.* 9: 143.

Golbraikh, A., Tropsha, A. (2002). " Beware of q2! " J. Mol. Graph. Model. 2, 269 276

Gough, J.D. et L.H. Hall. (1999). "Modeling the toxicity of amide herbicides using the electrotopological state". *Environ. Toxicol. Chem.* 18 : 1069–1075.

Grote K., Niemann L., Selzsam B., Haider W., Gericke C., Herzler M. et Chahoud I. (2008). "Epoxiconazole causes changes in testicular histology and sperm production in the japanese quail (*Coturnix coturnix japonica*)". *Environmental Toxicology and Chemistry*. 27: 2368-2374.

Grynkiewicz, M., Polkowska, Z., Gorecki, T. et Namiesnik, J. (2001). "Pesticides in precipitation in the Gdansk region (Poland)". *Chemosphere* 43(3): 303-312.

Gunier Robert B. , Mary H. Ward, Matthew Airola, Erin M. Bell, Joanne Colt, Marcia Nishioka, Patricia A. Buffler, Peggy Reynolds, Rudolph P. Rull, Andrew Hertz, Catherine Metayer, John R. Nuckols. (2011). "Determinants of Agricultural Pesticide Concentrations in Carpet Dust". *Environmental Health Perspectives*, 119:970-976

Hayden, K.M., Norton, M.C., Darcey, D., Ostbye, T., Zandi, P.P., Breitner, J.C., Welsh-Bohmer, K.A. (2010). "Occupational exposure to pesticides increases the risk of incident AD: the Cache County study". *Neurology* 74 (19): 1524–1530.

Hina Ahsan, Nasim Karim, Syed Sanwer Ali, Alia B. Munshi, Sohail Shaukat. (2013). "Impact of Pesticides Contamination on Nutritional Values of Marine fishery from Karachi Coast of Arabian Sea". *Food and Nutrition Sciences*. 4: 924-932

Holmes B. (2010). "Weed resistance could mean herbicide is futile". *New Scientist*, 206: 12.

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

Hongying, D., Wang, J., Zhide, H., Xiaojun, Y., et Xiaoyun, Z. (2008). "Prediction of Fungicidal Activities of Rice Blast Disease Based on Least-Squares Support Vector Machines and Project Pursuit Regression". *J. Agric. Food Chem.* 56: 10785.

Hoppin JA, Umbach DM, London SJ, Alavanja MC (2002). "Chemical predictors of wheeze among farmer pesticide applicators in the Agricultural Health Study". *Am. J. Respir. Crit Care Med.* 165 (5): 683-89.

Hoppin JA, Valcin M, Henneberger PK, Kullman GJ (2007) "Pesticide use and chronic bronchitis among farmers in the Agricultural Health Study". *Am J Ind Med.* 50 (12): 969-79.

Huynh Thi Ngoc Phuong (2007). "Synthèse et étude des relations structure/activité quantitatives (qsar/2d) d'analogues benzo[c]phénanthridiniques". Thèse de Doctorat en Chimie organique et thérapeutique, Faculté de Pharmacie d'Angers.

Jackel H. et W. Klein. (1991) "Prediction of mammalian toxicity by quantitative structure activity relationships: Aliphatic amines and anilines". *Quant. Struct.-Act. Relat.* 10 : 198–204.

Jasmine G., Yogeshwer Shukla (2011). "Pesticides and cancer: Insights into toxicoproteomic-based findings". *Journal of Proteomics* 74: 2713 – 2722.

Johnson, S.R. et P.C. Jurs (1997). "Prediction of acute mammalian toxicity from molecular structure for a diverse set of substituted anilines using regression analysis and computational neural networks" in Computer-Assisted Lead Finding and Optimization. Current Tools for Medicinal Chemistry, H. van de Waterbeemd, B. Testa, and G. Folkers, eds., Wiley-VCH, Basel. pp. 31–48.

Juranic, I.O. B.J. Drakulic, S.D. Petrovic, D.Z. Mijin, et M.V. Stankovic. (2006). "A QSAR study of acute toxicity of N-substituted fluoroacetamides to rats". *Chemosphere* 62 : 641–649.

Kjaerstad M.B., Taxvig C., Nellemann C., Vinggaard A.M. et Andersen H.E. (2010). "Endocrine disrupting effects *in vitro* of conazole antifungals used as pesticides and pharmaceuticals". *Reproductive Toxicology*, 30: 573-582.

Konovalov, D.A., Sim, N., Deconinck, E., Heyden, Y.V., and Coomans, D. (2008). "Statistical Confidence for variable selection in QSAR models". *J. Chem. Inf. Model*. 48, 370.

Konovalov, D.A., Sim, N., Deconinck, E., Heyden, Y.V., et Coomans, D. (2008). "Statistical Confidence for variable selection in QSAR models". *J. Chem. Inf. Model*. 48: 370.

Levin, M.J. (2007). "Pesticides: a toxic time bomb in our midst". *Praegers Publishers*.

Lian-Kuet Chai, Mee-Hua Wong , Hans C. B. Hansen (2013). "Degradation of chlorpyrifos in humid tropical soils". *Journal of Environmental Management* 125: 28-32.

Liess M., Brown C., Dohmen P., Duquesne S., Hart A., Heimbach F., Kreuger J., Lagadic L., Reinert W., Maund S., Streloke M. et Tarazona J. (2005). "Effects of Pesticides in the Field". *EU & SETAC Europe Workshop*, Octobre 2003, Le Croisic, France.

Lifei Zhang, Liang Dong, Shuangxin Shi, Li Zhou, Ting Zhang, Yeru Huang. (2009) "Organochlorine pesticides contamination in surface soils from two pesticide factories in Southeast China". *Chemosphere* 77: 628–633

Ma XX, Ran Y. (2009). "The research for organochlorine pesticides in soils of the Pearl River Delta". *Ecology and Environmental Sciences*. 18(1): 134-137

Mast M.A., Campbell D.H., Ingersoll G.P., Foreman W.T. et Krabbenhoft D.P. (2003). "Atmospheric deposition of nutrients, pesticides, and mercury in Rocky Mountain National Park, Colorado, 2002". US Department of the Interior, *U.S. Geological Survey Water-Resources Investigations Report* 03-4241.

Mills Paul K. et Yang, Richard C. (2007). "Agricultural exposures and gastric cancer risk in Hispanic farm workers in California". *Environmental Research* 104: 282–289

Montgomery MP, Kamel F, Saldana TM, Alavanja MCR et Sandler DP. (2008). "Incident diabetes and pesticide exposure among licensed pesticide applicators: Agricultural Health Study 1993 – 2003". *Amer J Epidemiol*. 67:1235-1246.

Moussaoui K.M., R. Boussahel, Y. Tchoulak, O. Haouchine, M. Benmami, N. Dalachi. (2001). "Utilisation, évaluation et impacts des pesticides en Algérie". *Ecole Nationale Polytechnique*

Moussaoui Yacine, Ludovic Tuduri, Yacine Kerchich , B.Y. Meklati et Gauthier Eppe. (2012). "Atmospheric concentrations of PCDD/Fs, dl-PCBs and some pesticides in northern Algeria using passive air sampling". *Chemosphere* 88: 270–277.

Multigner, L. (2005). "Effets retardés des pesticides sur la santé humaine". *Environnement, Risques et Santé 3* : 187-194.

Munaron D. (2004). "Etude des apports en herbicides et en nutriments par la CHARENTE : modélisation de la dispersion de l'atrazine dans le bassin de MARENNES-LERON" thèse de Doctorat, Université Pierre et Marie Curie, Paris VI.

Nendza, M. B. Dittrich, A. Wenzel, et W. Klein. (1991) "Predictive QSAR models for estimating ecotoxic hazard of plant-protecting agents: Target and non-target toxicity" in QSAR in Environmental Toxicology–IV, J.L.P. Hermens and A. Opperhuizen, eds., Elsevier, Amsterdam. 527–535.

Netzeva, T.I.,Worth,A.P., Aldenberg, T., Benigni,R., Cronin,M.T.D., Gramatica, P., Jaworska, J.S., Kahn, S., Klopman, G., Marchant, C.A.,Myatt, G., Nikolova-Jeliazkova, N., Patlewicz, G.Y., Perkins, R., Roberts, D.W., Schultz, T.W., Stanton, D.T., Van De Sandt, J.J.M., Tong, W., Veith, G., Yang, C. (2005). " Current status of methods for defining the applicability domain of (quantitative) structure–activity relationships ". Altern. Lab. Anim. 33, 155–173.

Neumann M., Liess M. et Schulz R. (2003). "A qualitative method for monitoring water quality in temporary channels or point sources and its application to pesticide contamination". *Chemosphere*. 51: 509-513

Ning-Xin, T., Ping, Li., Han-Bing, R., Ze-Rong, Li., and Xiang-Yuan, Li. (2010). Prediction of the acute toxicity of chemical compounds to the fathead minnow by machine learning approaches. *Chemom. Intell. Lab. Syst*. 100, 66.

Nougadère Alexandre , Véronique Sirot , Ali Kadar , Antony Fastier, Eric Truchot, Claude Vergnet, Frédéric Hommet, Joëlle Baylé, Philippe Gros, Jean-Charles Leblanc. (2012). "Total diet study on pesticide residues in France: Levels in food as consumed and chronic dietary risk to consumers". *Environment International* 45: 135–150

OCDE (2009). "Principles for the Validation, for Regulatory Purposes, of (Quantitative) Structure-Activity Relationship Models", Organisation de Coopération et de Développement Economique, Paris, France.

Olden, J.D., Jackson, D.A. (2002) " Illuminating the black box: understanding variable contributions in artificial neural networks". Ecological Modelling, 154, 135-150.

Olden J. D., Joy M.K. and R-G (2004). "Death, An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data", Journal Ecological Modeling and systems ecology, 178, pp. 389–397.

Orton F, Rosivatz E, Scholze M, Kortenkamp A. (2011). "Widely used pesticides with previously unknown endocrine activity revealed as in vitro antiandrogens". *Environmental Health Perspectives*. 119(6):794-800.

Oukali-Haouchine Ouzna, Enrique Barriuso, Yamina Mayata et Khadija M. Moussaoui. (2013). "Factors affecting Métribuzine retention in Algerian soils and assessment of the risks of contamination". *Environ Monit Assess* . 185:4107–4115

Parron, T., Requena, M., Hernandez, A.F., Alarcon, R. (2011). "Association between environmental exposure to pesticides and neurodegenerative diseases". *Toxicol. Appl. Pharmacol.* 256 (3) : 379–385.

Peter Willett. (2011). "Chemoinformatics: a history. Wiley Interdisciplinary Reviews: *Computational Molecular Science*, vol. 1, n° 1, pages 46-56.

Pigram, G. M., MacDonald, T. R. (2001). " Use of neural network models to predict industrial bioreactor effluent quality." Environmental Science & Technology 35(1): 157-162.

Poezevara G. (2011). "Fouille de graphes pour la découverte de contrastes entre classes : application l'estimation de la toxicité des molecules". Thèse de Doctorat Université de Caen.

Poezevara Guillaume. (2011) "Fouille de graphes pour la découverte de contrastes entre classes : application à l'estimation de la toxicité des molécules". Spécialité Informatique et applications, Thèse de doctorat, École Doctorale SIMEN , Université de Caen Basse-Normandie.

Polkowska, Z., Kot, A., Wiergowski, M., Wolska, L., Wolowska, K. et Namiesnik, J. (2000). "Organic pollutants in precipitation: determination of pesticides and polycyclic aromatic hydrocarbons in Gdansk, Poland". *Atmospheric Environment* 34(8): 1233-1245.

Prouvost H., Declerc C. (2005) "Exposition de la population aux pesticides dans la région Nord Pas de Calais : apport du programme PHYTO AIR".

Raevsky OA, Grigorev VJ, Modina AE et Worth AP (2010). "Prediction of acute toxicity to mice by the arithmetic mean toxicity (AMT) modelling approach". *SAR and QSAR in Environmental Research*, 21(3-4): 265-275.

Rebich, R. A., Coupe, R. H. et Thurman, E. M. (2004). "Herbicide concentrations in the Mississippi River Basin, the importance of chloroacetanilide herbicide degradates". *The Science of the Total Environment* 321: 189–199

Reini W. Bretveld, Mariette Hooiveld, Gerhard A. Zielhuis, Annelies Pellegrino, Iris A.L.M. van Rooij, Nel Roeleveld. (2008) "Reproductive disorders among male and female greenhouse workers". *Reproductive Toxicology* 25:) 107–114

Roberts Eric M., Paul B. English, Judith K. Grether, Gayle C. Windham, Lucia Somberg et Craig Wolff. (2007) "Maternal Residence Near Agricultural Pesticide Applications and Autism Spectrum Disorders among Children in the California Central Valley". *Environmental Health Perspectives* 115(10):1482-1489

Rousseau, P., Clément, M., Seux, R. and Briand, O. (2004). "Dix ans de mesure des niveaux de contamination de l'atmosphère par les pesticides en France". XXXIVéme Congrès du Groupe Français des Pesticides (GFP)

S. Kacew (1996) "Role of rat strain in the differential sensitivity to pharmaceutical agents and naturally occurring substances". *J. Toxicol. Environ. Health, Part A* 47: 1–30.

Sanagi Mohd Marsin, Syairah Salleh, Wan Aini Wan Ibrahim, Ahmedy Abu Naim , Dadan Hermawan , Mazidatulakmam Miskam , Iqbal Hussain , Hassan Y. Aboul-Enein.(2013) "Molecularly imprinted polymer solid-phase extraction for the analysis

of organophosphorus pesticides in fruit samples". *Journal of Food Composition and Analysis* 32: 155–161.

Sanderson, H., Johnson, D.J., Reitsma, T., Brain, R.A., Wilson ,C.J., et Solomon, K.R. (2004). "Ranking and prioritization of environmental risks of pharmaceuticals in surface waters". *Reg. Tox. Pharm*. 39: 158.

Sarkhosh, M., Jahan, B., et Ghasemi, M. A. (2012). "A quantitative structure- property relationship of gas chromatographic/mass spectrometric retention data of 85 volatile organic compounds as air pollutant materials by multivariate methods". *Chem. Cent. J*. 6(Suppl 2):S4

Sazonovas, A., P. Japertasa et R. Didziapetrisa (2010). "Estimation of reliability of predictions and model applicability domain evaluation in the analysis of acute toxicity (LD50)". *SAR and QSAR in Environmental Research* 21: 127–148.

Scheyer, A., Graeff, C., Morville, S., Mirabel, P. et Millet, M. (2005). "Analysis of some organochlorine pesticides in an urban atmosphere (Strasbourg, east of France)". *Chemosphere*. 58: 1517-1524.

Schulz R. (2001). "Rainfall-induced sediment and pesticide input from orchards into the Lourens River, Western Cape, South Africa: importance of a single event". *Water Research.* 35: 1869-1876

Schulz R. (2004). "Field studies on exposure, effects, and risk mitigation of aquatic nonpoint-source insecticide pollution: a review". *Journal of Environmental Quality*, 33: 419-448

Shojaei Saadi, H., Abdollahi, M. (2012). "Is there a link between human infertilities and exposure to pesticides". *Int. J. Pharmacol*. 8(8): 708–710.

Sophie Coat , Dominique Monti, Pierre Legendre , Claude Bouchon , Félix Massat , Gilles Lepoint (2011). "Organochlorine pollution in tropical rivers (Guadeloupe): Role of ecological factors in food web bioaccumulation". *Environmental Pollution* 159: 1692-1701

Soto, A. J., Cecchini, R. L., Vazquez, G. E., et Ponzoni, I. (2009). "Multi-Objective Feature Selection in QSAR Using a Machine Learning Approach". *QSAR Comb. Sci*. 28(11-12): 1509.

Srinivas Reddy, A., Kumar, S., Garg, R. (2010). "Hybrid-genetic algorithm based descriptor optimization and QSAR models for predicting the biological activity of Tipranavir analogs for HIV protease inhibition". *J. Mol. Graph. Model.* 28: 852-862.

Srinivas Reddy, A., Kumar, S., Garg, R. (2010). "Hybrid-genetic algorithm based descriptor optimization and QSAR models for predicting the biological activity of Tipranavir analogs for HIV protease inhibition". *J. Mol. Graph. Model.* 28: 852-862.

Steen, R. J. C. A., Van der Vaart, J., Hiep, M., Van Hattum, B., Cofino, W. P. et Brinkman, U. A. T. (2001). "Gross fluxes and estuarine behaviour of pesticides in the Scheldt Estuary (1995-1997)". *Environmental Pollution* 115(1): 65-79.

Tagert Mary Love M., Joseph H. Massey, David R. Shaw. (2014). "Water quality survey of Mississippi's Upper Pearl River". *Science of the Total Environment* 481: 564–573.

Talat Saeed, Wajih N. Sawaya, Nisar Ahmad, Sangeetha Rajagopal, et Ali Al-Omair (2005). Organophosphorus pesticide residues in the total diet of Kuwait".*The Arabian Journal for Science and Engineering*. 30(1A): 17-27

Tanii H. S. Saito, et K. Hashimoto (1992). "Structure–toxicity relationship of ethylene glycol ether". *Arch. Toxicol*. 66: 368–371.

Tanii, H. J. Huang, and K. Hashimoto (1995) "Structure–acute toxicity relationship of aromatic hydrocarbons in mice". *Toxicol. Lett*. 76: 27–31.

Todeschini, R., Consonni, V. (2000). "Handbook of Molecular Descriptors". Wiley-VCH, Weinheim.

Touré K, Coly M, Toure D, Fall M, Sarr MD. (2011) "Investigation of Death Cases by Pesticides Poisonning in a Rural Community, Bignona, Senegal". *Epidemiol* 1:105. doi:10.4172/2161-1165.1000105

Tropsha, A., Gramatica, P., Gombar, V.K.(2003). " The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models ". QSAR Comb. Sci. 22, 69–77.

Tugcu, G., Turker, M. S., Vracko, M., Novic, M., et Minovski, N. (2012). "QSTR modelling of the acute toxicity of pharmaceuticals to fish". *SAR QSAR Environ. Res.* 23(3-4): 297.

UIPP Union des Industries de la Protection des Plantes, protéger les plantes c'est protéger l'avenir ; rapport d'activités 2011/2012.

Van der Mark, M., Brouwer, M., Kromhout, H., Nijssen, P., Huss, A., Vermeulen, R. (2012). "Is pesticide use related to Parkinson disease? Some clues to heterogeneity in study results". *Environ. Health Perspect.* 120 (3): 340–347.

Van Maele-Fabry, G., Hoet, P., Vilain, F., Lison, D. (2012). "Occupational exposure to pesticides and Parkinson's disease: a systematic review and meta-analysis of cohort studies". *Environ. Int.* 46: 30–43.

Walker S., Bell K., Robinson G. et Widderick M. (2011). "Flaxleaf fleabane (*Conyza bonariensis*) populations have developped glyphosate resistance in north-east Australian cropping fields". *Crop Protection*. 30: 311-317.

Wan YW, Tang TF, Zhou ZL, (2009). "Distribution and sources of organochlorine pesticides in Beijing Guanting Reservoir". *Journal of Ecology and Rural Environment*. 25(1): 53-56, 68

Wang Jun, Michael M. Kliks , Soojin Jun , Qing X. Li (2010) "Residues of organochlorine pesticides in honeys from different geographic regions". *Food Research International* 43: 2329–2334.

Wang, G. et N. Bai (1998). "Structure–activity relationships for rat and mouse LD50 of miscellaneous alcohols". *Chemosphere* 36: 1475–1483.

Wessel, M.D., et Jurs, P.C. (1994). "Prediction of reduced ion mobility constants from structural information using multiple linear regression analysis and computational neural networks". *Anal. Chem.* 66: 2480–2487.

Wigle Donald T., Michelle C. Turner et Daniel Krewski. (2009) "A Systematic Review and Meta-analysis of Childhood Leukemia and Parental Occupational Pesticide Exposure". *Environmental Health Perspectives*. 117: 1505-1513

Winter C. K. (1992). "Pesticide Tolerances and Their Relevance as Safety Standards". *Regulatory Toxicology and Pharmacology* 15: 137–150.

Wold S, Eriksson L (1995). "Statistical validation of QSAR results". In Waterbeemd, Han van de. *Chemometric methods in molecular design*. Weinheim: VCH. pp. 309–318. ISBN 3-527-30044-9.

Xu, J., Zhu, L., Fang, D., Wang, L., Xiao, S., Liu, Li., et Xu, W. (2012). "QSPR studies of impact sensitivity of nitro energetic compounds using three-dimensional descriptors". *J. Mol.Graph.Model.* 36: 10–19.

Xu, L., et Zhang, W. J. (2001). "Comparison of different methods for variable selection". *Anal. Chim. Acta* 446 : 477.

Xu, L., et Zhang, W. J. (2001). "Comparison of different methods for variable selection". *Anal. Chim. Acta* 446: 477.

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

Yao Y., Harner T., Blanchard P., Tuduri L., Waite D., Poissant L., Murphy C., Belzer W., Aulagnier F. et Sverko E. (2008). "Pesticides in the atmosphere across Canadian agricultural regions". *Environmental Science and Technology*. 42: 5931-5937.

Zahouily, M., A. Rhihil, H. Bazoui, S. Sebti, et D. Zakarya (2002). "Structure–toxicity relationships study of a series of organophosphorus insecticides". *J. Mol. Model*. 8 : 168–172.

Zakarya, D., E.M. Larfaoui, A. Boulaamail, and T. Lakhlifi (1996). "Analysis of structure–toxicity relationships for a series of amide herbicides using statistical methods and neural network". *SAR QSAR Environ. Res*. 5: 269–279.

Zhang Lifei, Liang Dong, Wenlong Yang, Li Zhou, Shuangxin Shi, Xiulan Zhang, Shan Niu, Lingling Li, Zhongxiang Wu, Yeru Huang (2013). "Passive air sampling of organochlorine pesticides and polychlorinated biphenyls in the Yangtze River Delta, China: Concentrations, distributions, and cancer risk assessment". *Environmental Pollution* 181: 159-166

Zhao L, Hou H, Guo PY (2009). "Distribution of organochlorine pesticides in soils in Haihe River and Haihe estuary area". *China Environmental Science*. 30(2): 543-550

Zhou Q., Sun X., Gao R., Zhang Q. et Wang W. (2010). "Mechanism study on OH-initiated atmospheric degradation of the organophosphorus pesticide chlorpyrifos". *Journal of Molecular Structure*: THEOCHEM. 952: 8-15.

Zhu H, Martin TM, Ye L, Sedykh A, Young DM et Tropsha A (2009a). "Quantitative Structure−Activity Relationship Modeling of Rat Acute Toxicity by Oral Exposure". *Chemical Research in Toxicology* 22(12): 1913–1921.

Zhu H, Ye L, Richard A, Golbraikh A, Wright FA, Rusyn I et Tropsha A (2009b). "A novel two-step hierarchical quantitative structure-activity relationship modeling work flow for predicting acute toxicity of chemicals in rodents". *Environmental Health Perspectives* 117(8): 1257-1264.

ANNEXE A

Référence : Pesticide Properties Data Base, PPDB–University of Hertfordshire (http://www.herts.ac.uk/aeru/projects/ppdb/index.htm)

Pesticide	Type	LD50 mg/Kg	Masse Molaire (g/mol)	N° de CAS
1,2 DICHLOROPROPANE	Insecticide	1947	112.99	78-87-5
2,4,5-TRICHLOROPHENOL	Herbicide	820	197.45	95-95-4
2,4-D	Herbicide	469	221.04	94-75-7
2,4-DB	Herbicide	877	249.09	94-82-6
2,4-DIMETHYLPHENOL	Fongicide	242	122.17	105-67-9
2-AMINO BUTANE	Fongicide	350	73.14	13952-84-6
ACEPHATE	Insecticide	945	183.17	30560-19-1
ACETAMIPRIDE	Insecticide	213	222.67	135410-20-7
ACETHION	Insecticide	1100	272.32	919-54-0
ACETOCHLORE	Herbicide	1929	269.77	34256-82-1
ACIDE 4 CHLOROPHENOXYACETIQUE	Herbicide	850	186.59	122-88-3
ACIFLUORFEN	Herbicide	1370	361.66	50594-66-6
ACROLEINE	Herbicide	29	56.1	107-02-8
ALACHLORE	Herbicide	930	269.77	15972-60-8
ALANYCARBE	Insecticide	330	399.53	83130-01-2
ALDICARBE	Insecticide	0,9	190.26	116-06-3
ALDOXYCARBE	Insecticide	27	222.26	1646-88-4
ALDRINE	Insecticide	39	364.91	309-00-2
ALIDOCHLORE	Herbicide	290	173.64	93-71-0
ALLETHRINE	Insecticide	685	302.41	584-79-2
ALLYXYCARB	Insecticide	89	274.36	6392-46-7
ALPHACYPERMETHRINE	Insecticide	57	416.30	67375-30-8
ALPHAENDOSULFAN	Insecticide	38	406.93	959-98-8
AMETRYNE	Herbicide	110	227.12	834-12-8
AMICARBAZONE	Herbicide	1015	241.29	129909-90-6
AMIDITHION	Insecticide	600	273.31	919-76-6
AMINOCARBE	Insecticide	30	208.26	2032-59-9
AMIPROFOS-METHYL	Herbicide	309	304.31	36001-88-4
AMITRAZE	Insecticide	800	293.41	33089-61-1
ANCYMIDOLE	Herbicide	1721	256.30	12771-68-5
ANILAZINE	Fongicide	4570	275.53	101-05-3

Pesticide	Type	LD50 mg/Kg	Masse Molaire (g/mol)	N° de CAS
ANILOFOS	Herbicide	472	367.85	64249-01-0
ARAMITE	Insecticide	3900	334.87	140-57-8
ASOMATE	Fongicide	335	435.57	3586-60-5
ATRAZINE	Herbicide	1869	215.68	1912-24-9
AZACONAZOLE	Fongicide	308	300.14	60207-31-0
AZAMETIPHOS	Insecticide	1180	324.68	35575-96-3
AZINPHOS-ETHYL	Insecticide	12	345.38	2642-71-9
AZINPHOS-METHYL	Insecticide	9	317.32	86-50-0
BARBANE	Herbicide	527	258.10	101-27-9
BENALAXIL	Fongicide	680	325.40	71626-11-4
BENDIOCARBE	Insecticide	34	223.23	22781-23-3
BENFURACARBE	Insecticide	205	410.53	82560-54-1
BENQUINOX	Fongicide	100	241.25	495-73-8
BENSULIDE	Herbicide	270	397.51	741-58-2
BENSULTAP	Insecticide	1105	431.6	17606-31-4
BENTAZONE	Herbicide	500	240.3	25057-89-0
BENZTHIAZURON	Herbicide	1280	207.25	1929-88-0
BETACYPERMETHRINE	Insecticide	93	416.3	65731-84-2
BIFENTHRINE	Insecticide	54,5	422.88	82657-04-3
BINAPACRYL	Fongicide	58	322.31	485-31-4
BIOALLETHRINE	Insecticide	709	302.41	584-79-2
BIPHENYLE	Fongicide	2140	154.21	92-52-4
BRODIFACOUM	Rodonticide	0,4	523.4	56073-10-0
BROMACIL	Herbicide	1300	261.12	314-40-9
BROMOCYCLENE	Insecticide	12500	393.75	1715-40-8
BROMOPHOS	Insecticide	1600	366.00	2104-96-3
BROMOPHOS-ETHYL	Insecticide	52	394.0	4824-78-6
BROMOXYNIL	Herbicide	81	276.9	1689-84-5
BROMOXYNIL HEPTANOATE	Herbicide	291	389.08	56634-95-8
BROMOXYNIL OCTANOATE	Herbicide	238	403.00	1689-99-2
BROMUCONAZOLE	Fongicide	328	377.06	116255-48-2
BRONOPOL	Fongicide	254	200.0	52-51-7
BUPIRIMATE	Fongicide	4000	316.42	41483-43-6
BUTACHLORE	Herbicide	2000	311.9	23184-66-9
BUTAMIFOS	Herbicide	630	332.36	36335-67-8
BUTILATE	Herbicide	3500	217.4	2008-41-5
BUTOCARBOXIME	Insecticide	153	222.26	34681-23-7
BUTONATE	Insecticide	1100	327.52	126-22-7
BUTOXYCARBOXIME	Insecticide	458	222.26	34681-23-7
BUTRALINE	Herbicide	1049	295.33	33629-47-9
BUTROXYDIM	Herbicide	1635	399.5	138164-12-2
CADUSAFOS	Insecticide	30,1	270.39	95465-99-9
CAMPHECHLORE	Insecticide	50	413,96	8001-35-2

Pesticide	Type	LD50 mg/Kg	Masse Molaire (g/mol)	N° de CAS
CARBANOLATE	Insecticide	30	213.66	671-04-5
CARBARYL	Insecticide	614	201.22	63-25-2
CARBETAMIDE	Herbicide	1718	236.27	16118-49-3
CARBOFURAN	Insecticide	7	221.26	1563-66-2
CARBOPHENOTHION	Insecticide	10	342.9	786-19-6
CARBOSULFAN	Insecticide	101	380.5	55285-14-8
CARBOXINE	Fongicide	2588	235.30	5234-68-4
CARTAP	Insecticide	325	273.80	15263-53-3
CHLORBROMURON	Herbicide	2150	293.5	13360-45-7
CHLORBUFAME	Herbicide	2380	223.7	1967-16-4
CHLORDANE	Insecticide	460	409.78	57-74-9
CHLORDECONE	Insecticide	91,3	490.64	143-50-0
CHLORDIMEFORME	Insecticide	160	196.68	6164-98-3
CHLORETHOXYFOS	Insecticide	1,8	336.0	54593-83-8
CHLORFENAC	Herbicide	1780	239.48	85-34-7
CHLORFENAPYR	Insecticide	441	407.62	122453-73-0
CHLORFENETHOL	Insecticide	500	267.15	80-06-8
CHLORFENSULPHIDE	Insecticide	4000	352.07	2274-74-0
CHLORFENVINPHOS	Insecticide	12	359.6	470-90-6
CHLORIDAZONE	Herbicide	2140	221.6	1698-60-8
CHLOROBENZILATE	Insecticide	2784	325.19	510-15-6
CHLOROMETHIURON	Insecticide	25000	228.75	28217-97-2
CHLORONEBE	Fongicide	11000	207.1	2675-77-6
CHLOROPHACINONE	Rodonticide	3,15	374.82	3691-35-8
CHLOROPICRINE	Insecticide	250	164.37	76-06-2
CHLORPROPHAME	Herbicide	4200	213.66	101-21-3
CHLORPYRIPHOS	Insecticide	64	350.89	2921-88-2
CHLORPYRIPHOS-METHYL	Insecticide	2814	322.53	5598-13-0
CHLORTHIAMIDE	Herbicide	757	206.09	1918-13-4
CHLORTHIAMIDE	Herbicide	757	206.09	1918-13-4
CHLORTHION	Insecticide	880	297.66	500-28-7
CHLORTHIOPHOS	Insecticide	10	361.25	60238-56-4
CLETHODIM	Herbicide	1133	359.92	99129-21-2
CLODINAFOP-PROPARGYL	Herbicide	1392	349.8	105512-06-9
CLOETHOCARB	Insecticide	35,5	259.69	51487-69-5
CLOMAZONE	Herbicide	1369	239.7	81777-89-1
CLOSANTEL	Insecticide	262	663.08	57808-65-8
COUMACHLORE	Rodonticide	16	342.77	81-82-3
CRIMIDINE	Rodonticide	1,25	171.63	535-89-7
CROTOXYPHOS	Insecticide	66	314.27	7700-17-6
CUMYLURON	Herbicide	961	302.80	99485-76-4
CYANAZINE	Herbicide	288	240.69	21725-46-2
CYANOPHOS	Insecticide	610	243.2	636-26-2

Pesticide	Type	LD50 mg/Kg	Masse Molaire (g/mol)	N° de CAS
CYCLOXYDIME	Herbicide	3940	325.46	101205-02-1
CYCLURON	Herbicide	1500	198.31	2163-69-1
CYHALOTHRINE	Insecticide	140	449.85	68085-85-8
CYHEXATIN	Insecticide	265	385.17	13121-70-5
CYMOXANIL	Fongicide	760	198.18	57966-95-7
CYPERMETHRINE	Insecticide	287	416.3	52315-07-8
CYPHENOTHRIN	Insecticide	318	375.46	39515-40-7
CYPROFURAM	Fongicide	174	279.72	69581-33-5
CYROMAZINE	Insecticide	3387	166.18	66215-27-8
DALAPON	Herbicide	9330	143.0	75-99-0
DAZOMET	Insecticide	415	162.3	533-74-4
DELTAMETHRINE	Insecticide	87	505.2	52918-63-5
DEMETON-O-METHYLSULFONE	Insecticide	500	262.28	25476-48-6
DEMETON-S-METHYL	Insecticide	40	230.29	919-86-8
DEMETON-S-METHYL SULFONE	Insecticide	32,4	262.89	17040-19-6
DESMETRYNE	Herbicide	1390	213.30	1014-69-3
DIAFENTHIURON	Insecticide	2068	384.58	80060-09-9
DIALLATE	Herbicide	395	270.2	2303-16-4
DIAMIDAFOS	Insecticide	140	200,17	1754-58-1
DIAZINON	Insecticide	1139	304.35	333-41-5
DIBROMOCHLOROPROPANE	Insecticide	170	236.36	96-12-8
DICAMBA	Herbicide	1581	221.04	1918-00-9
DICHLOFENTHION	Insecticide	172	315.15	97-17-6
DICHLONE	Fongicide	160	227.04	117-80-6
DICHLOROPHENE	Fongicide	2690	269.1	97-23-4
DICHLORPROP	Herbicide	825	235.06	120-36-5
DICHLORPROP-P	Herbicide	567	235.06	15165-67-0
DICHLORVOS	Insecticide	80	220.98	62-73-7
DICOFANE	Insecticide	113	354,49	50-29-3
DICOFOL	Insecticide	578	370.49	115-32-2
DICROTOPHOS	Insecticide	17	237.19	141-66-2
DIELDRINE	Insecticide	46	380.91	60-57-1
DIENOCHLORE	Insecticide	3160	474.8	2227-17-0
DIETHATYL ETHYL	Herbicide	2300	311.5	38727-55-8
DIFENAMIDE	Herbicide	970	239.31	957-51-7
DIFENOCONAZOLE	Fongicide	1453	406.26	119446-68-3
DIFLOVIDAZIN	Insecticide	594	304.69	162320-67-4
DIFLUMETORIM	Fongicide	448	327.76	130339-07-0
DIMEFOX	Insecticide	1	154.12	115-26-4
DIMEPIPERATE	Herbicide	946	263.39	61432-55-1
DIMETACHLORE	Herbicide	1600	255.74	50563-36-5
DIMETHENAMID	Herbicide	397	275.79	87674-68-8
DIMETHENAMID-P	Herbicide	429	275.8	163515-14-8

Pesticide	Type	LD50 mg/Kg	Masse Molaire (g/mol)	N° de CAS
DIMETHIPIN	Herbicide	458	210.27	55290-64-7
DIMETHOATE	Insecticide	245	229.26	60-51-5
DIMETHOMORPHE	Fongicide	3900	387.86	110488-70-5
DIMETHYLVINPHOS	Insecticide	97,5	331.52	2274-67-1
DIMEXANO	Herbicide	240	214.35	1468-37-7
DINICONAZOLE	Fongicide	474	326.22	83657-24-3
DINOBUTON	Fongicide	140	326.30	973-21-7
DINOCAP	Fongicide	1212	364.18	39300-45-3
DINOSEBE	Herbicide	25	240.22	88-85-7
DINOTERBE	Insecticide	25	240.21	1420-07-1
DIOXABENZOPHOS	Insecticide	125	216.20	3811-49-2
DIOXACARBE	Insecticide	25	223.23	6988-21-2
DIOXATHION	Insecticide	23	456.54	78-34-2
DIPHACINONE	Rodonticide	2,3	340.37	82-66-6
DIQUAT	Herbicide	214	184.24	2764-72-9
DISULFOTON	Insecticide	2,6		298-04-4
DITALIMPHOS	Fongicide	4930	299.28	5131-24-8
DITHIANON	Fongicide	300	296.32	3347-22-6
DIURON	Herbicide	437	233.09	330-54-1
DNOC	Herbicide	25	198.13	534-52-1
EDIFENPHOS	Fongicide	150	310.37	17109-49-8
ENDOSULFAN	Insecticide	38	406.93	115-29-7
ENDOTHAL	Herbicide	51	186.16	145-73-3
ENDOTHION	Insecticide	23	280.23	03/04/2718
EPICHLOROHYDRINE	Insecticide	40	92.52	106-89-8
EPN	Insecticide	14	323.30	2104-64-5
EPOXICONAZOLE	Fongicide	3160	329.76	133855-98-8
EPTC	Herbicide	915	189.3	759-94-4
ESFENVALERATE	Insecticide	88,5	419.90	66230-04-4
ETACONAZOLE	Fongicide	1343	328.20	60207-93-4
ETHANEDIAL	Herbicide	119	58.04	557-30-2
ETHIOFENCARBE	Insecticide	200	225.31	29973-13-5
ETHION	Insecticide	208	384.48	563-12-2
ETHIPROLE	Insecticide	7080	397.20	181587-01-9
ETHOATE-METHYLE	Insecticide	340	243.29	116-01-8
ETHOPROPHOS	Insecticide	32,9	242.3	13194-48-4
ETHOXYQUINE	Fongicide	1726	217.31	91-53-2
ETHOXYSULFURON	Herbicide	3270	398.39	126801-58-9
FENAMIDONE	Fongicide	2028	311.40	161326-34-7
FENARIMOL	Fongicide	2500	331.20	60168-88-9
FENAZAFLOR	Insecticide	240	375.13	14255-88-0
FENAZAQUIN	Acaricide	134	306.40	120928-09-8
FENCHLORPHOS	Insecticide	500	321.55	299-84-3

Pesticide	Type	LD50 mg/Kg	Masse Molaire (g/mol)	N° de CAS
FENITROTHION	Insecticide	330	277.23	122-14-5
FENOBUCARB	Insecticide	620	207.27	3766-81-2
FENOPROP	Herbicide	650	269.51	93-72-1
FENPROPATHRINE	Insecticide	870	349.42	39515-41-8
FENPROPIDINE	Fongicide	1452	273.46	67306-00-7
FENPROPIMORPHE	Fongicide	1670	303.48	67564-91-4
FENPYROXIMATE	Acaricide	245	421.49	134098-61-6
FENSULFOTHION	Insecticide	2,2	308.35	115-90-2
FENTHION	Insecticide	250	278.33	55-38-9
FENTINE ACETATE	Fongicide	140	410.07	900-95-8
FENURON	Herbicide	6400	164.20	101-42-8
FENVALERATE	Insecticide	451	419.90	51630-58-1
FIPRONIL	Insecticide	92	437.15	120068-37-3
FLONICAMIDE	Insecticide	884	229.16	158062-67-0
FLORASULAM	Herbicide	5000	359.28	145701-23-1
FLUAZIFOP-BUTYL	Herbicide	3030	383.36	69806-50-4
FLUAZIFOP-P-BUTYL	Herbicide	2451	383.36	69806-50-4
FLUAZOLATE	Herbicide	5000	443.62	174514-07-9
FLUCHLORALINE	Herbicide	1550	355.70	33245-39-5
FLUCYTHRINATE	Insecticide	67	451.46	70124-77-5
FLUFENACET	Herbicide	598	363.33	142459-58-3
FLUMORPH	Fongicide	2710	371.40	211867-47-9
FLUOROACETAMIDE	Insecticide	13	77.06	640-19-7
FLUOROGLYCOPHENE	Herbicide	1480	419.69	77501-60-1
FLUORURE DE SULFURYLE	Insecticide	100	102.06	2699-79-8
FLUQUINCONAZOLE	Fongicide	112	376.17	136426-54-5
FLUSILAZOLE	Fongicide	674	315.39	85509-19-9
FLUSULFAMIDE	Fongicide	132	415.17	106917-52-6
FLUVALINATE	Insecticide	261	502.91	69409-94-5
FOMESAFEN	Herbicide	1250	438.76	72178-02-0
FONOFOS	Insecticide	6,8	246.3	944-22-9
FORMETANATE	Insecticide	14,8	221.26	22259-30-9
FORMOTHION	Insecticide	365	257.27	2540-82-1
FOSPIRATE	Insecticide	869	306.47	5598-52-7
FOSTHIAZATE	Insecticide	57	283.35	98886-44-3
FURALAXYL	Fongicide	949	301.34	57646-30-7
FURAMETPYR	Fongicide	590	333.81	123572-88-3
FURATHIOCARBE	Insecticide	53	382.47	65907-30-4
FURFURAL	Fongicide	65	96.08	98-01-1
FURILAZOLE	Fongicide	869	278.14	121776-33-8
FURMECYCLOX	Fongicide	3780	251.32	60568-05-0
GAMMA-CYHALOTHRINE	Insecticide	55	449.85	76703-62-3
GLUFOSINATE	Herbicide	1625	181.13	51276-47-2

Pesticide	Type	LD50 mg/Kg	Masse Molaire (g/mol)	N° de CAS
GLUTARALDEHYDE	Fongicide	134	100.12	111-30-8
HALFENPROX	Insecticide	132	477.34	111872-58-3
HALOFENOZIDE	Insecticide	2850	330.81	112226-61-6
HALOSULFURON-METHYL	Herbicide	7758	434.81	100784-20-1
HALOXYFOP	Herbicide	337	361.70	69806-34-4
HEPTENOPHOS	Insecticide	96	250.6	23560-59-0
HEXACONAZOLE	Fongicide	2189	314.21	79983-71-4
HEXAZINONE	Herbicide	1690	252.31	51235-04-2
HYMEXAZOL	Fongicide	1600	99.15	10004-44-1
ICARIDIN	Insecticide	2236	229.32	119515-38-7
IMAZALIL	Fongicide	227	297.18	35554-44-0
IMIPROTHRIN	Insecticide	900	318.37	72963-72-5
INDANOFAN	Herbicide	460	340.80	133220-30-1
INDOXACARBE	Insecticide	268	527.83	173584-44-6
IODOFENPHOS	Insecticide	2330	413.00	18181-70-9
IOXYNIL	Herbicide	130	370.91	1689-83-4
IPCONAZOLE	Fongicide	888	333.9	125225-28-7
IPROBENFOS	Fongicide	680	288.34	26087-47-8
ISAZOFOS	Insecticide	27	313.74	42509-80-8
ISOBENZAN	Insecticide	4,8	411.75	297-78-9
ISOCARBOPHOS	Insecticide	50	289.29	24353-61-5
ISOPHENPHOS-METHYLE	Insecticide	21,5	331.37	99675-03-3
ISOPROCARBE	Insecticide	403	193.24	2631-40-5
ISOPROTHIOLANE	Fongicide	1190	290.40	50512-35-1
ISOPROTURON	Herbicide	1826	206.28	34123-59-6
ISOXATHION	Insecticide	112	313.31	18854-01-8
KELEVAN	Insecticide	240	634.81	4234-79-1
LAMBDA-CYHALOTHRINE	Insecticide	56	449.85	91465-08-6
LEPTOPHOS	Insecticide	43	412.07	21609-90-5
LINDANE	Insecticide	163	290,82	210-168-9
LINURON	Herbicide	1146	249.09	330-55-2
MALATHION	Insecticide	1778	330.36	121-75-5
MCPA	Herbicide	962	200.62	94-74-6
MCPA-THIOETHYL	Herbicide	450	244.74	25319-90-8
MCPB	Herbicide	4300	228.67	94-81-5
MECARBAM	Insecticide	36	329.37	2595-54-2
MECOPROP	Herbicide	1166	214.65	7085-19-0
MEPIQUAT	Herbicide	1490	114.21	15302-91-7
METALAXYL	Fongicide	633	279.33	57837-19-1
METAMITRONE	Herbicide	1183	202.21	41394-05-2
METAZACHLORE	Herbicide	3480	277.75	67129-08-2
METCONAZOLE	Fongicide	595	319.83	125116-23-6
METHACRIFOS	Insecticide	678	240.22	62610-77-9

Pesticide	Type	LD50 mg/Kg	Masse Molaire (g/mol)	N° de CAS
METHAMIDOPHOS	Insecticide	30	141.13	10265-92-6
METHASULFOCARB	Fongicide	112	261.32	66952-49-6
METHAZOLE	Herbicide	777	261.1	20354-26-1
METHIDATHION	Insecticide	25	302.3	950-37-8
METHIOCARB	Insecticide	19	225.31	2032-65-7
METHOMYL	Insecticide	30	162.21	16752-77-5
METOLACHLOR	Herbicide	1200	283.8	51218-45-2
METOMINOSTROBIN	Fongicide	708	284.31	133408-50-1
METRIBUZINE	Herbicide	32	214.29	21087-64-9
METSULFOVAX	Fongicide	3929	232.31	21452-18-6
MEVINPHOS	Insecticide	3,5	224.1	7786-34-7
MEXACARBATE	Insecticide	14	222.28	315-18-4
MOLINATE	Herbicide	483	187.3	2212-67-1
MONALIDE	Herbicide	2600	239.7	7287-36-7
MONOCROTOPHOS	Insecticide	14	223.16	6923-22-4
MONOLINURON	Herbicide	2100	214.65	1746-81-2
MORPHOTHION	Insecticide	190	285.33	144-41-2
MYCLOBUTANIL	Fongicide	1600	288.78	88671-89-0
NALED	Insecticide	83	380.79	300-76-5
NAPTALAME	Herbicide	1770	291.30	132-66-1
NITENPYRAM	Insecticide	1575	270.72	150824-47-8
NITHIAZINE	Insecticide	303	216.24	58842-20-9
NITRAPYRINE	Bactericide	713	230.91	1929-82-4
NITROFENE	Herbicide	2630	284.1	1836-75-5
N-NITROSODIMETHYLAMINE	Insecticide	36	74.08	62-75-9
NUARIMOL	Fongicide	1250	314.7	63284-71-9
OCTHILINONE	Fongicide	550	213.34	26530-20-1
OFURACE	Fongicide	2600	281.74	58810-48-3
OMETHOATE	Insecticide	50	213.2	1113-02-6
OXADIXYL	Fongicide	1860	278.3	77732-09-3
OXAMYL	Insecticide	2,5	219.26	23135-22-0
OXYCARBOXINE	Fongicide	1632	267.31	5259-88-1
OXYDEMETON-METHYLE	Insecticide	48	246.3	301-12-2
PARAQUAT	Herbicide	110	186.25	4685-14-7
PARATHION	Insecticide	2	291.26	56-38-2
PARATHION METHYL	Insecticide	3	263.21	298-00-0
PEBULATE	Herbicide	1120	203.34	1114-71-2
PENDIMETHALINE	Herbicide	3189	281.31	40487-42-1
PENTACHLOROPHENOL	Fongicide	80	266.34	87-86-5
PETHOXAMID	Herbicide	983	295.81	106700-29-2
PHENKAPTON	Insecticide	44	377.31	2275-14-1
PHENTHOATE	Insecticide	249	320.39	07/03/2597
PHORATE	Insecticide	2	260.4	298-02-2

Pesticide	Type	LD50 mg/Kg	Masse Molaire (g/mol)	N° de CAS
PHOSALONE	Insecticide	120	367.8	2310-17-0
PHOSMET	Insecticide	113	317.3	732-11-6
PHOSNICHLOR	Insecticide	500	279.65	5826-76-6
PHOSPHAMIDON	Insecticide	7	299.69	13171-21-6
PICLORAME	Herbicide	4012	241.46	1918-02-1
PIPEROPHOS	Herbicide	324	353.48	24151-93-7
PIRIMICARB	Insecticide	142	238.39	23103-98-2
PLIFENATE	Insecticide	10000	336.43	21757-82-4
PRALLETHRIN	Insecticide	460	300.40	23031-36-9
PRETILACHLORE	Herbicide	6099	311.85	51218-49-6
PROCHLORAZ	Fongicide	1023	376.7	67747-09-5
PROFENOPHOS	Insecticide	358	373.63	41198-08-7
PROMECARBE	Insecticide	35	207.3	2631-37-0
PROMETONE	Herbicide	1518	225.3	1610-18-0
PROPACHLORE	Herbicide	550	211.69	1918-16-7
PROPAFOS	Insecticide	61	304.34	7292-16-2
PROPANIL	Herbicide	960	218.08	709-98-8
PROPARGITE	Insecticide	2639	350.47	2312-35-8
PROPAZINE	Herbicide	3840	229.71	139-40-2
PROPICONAZOL	Fongicide	958	342.22	60207-90-1
PROPOXUR	Insecticide	50	209.24	114-26-1
PROSULFOCARBE	Herbicide	1820	251.39	52888-80-9
PROSULFURON	Herbicide	546	419.38	94125-34-5
PROTHIOCARB	Fongicide	1300	190.31	19622-08-3
PROTHIOFOS	Insecticide	925	345.25	34643-46-4
PROTHOATE	Insecticide	8	285.36	2275-18-5
PYMETROZINE	Insecticide	5820	217.23	123312-89-0
PYRACLOFOS	Insecticide	237	360.8	77458-01-6
PYRAZOPHOS	Fongicide	151	373.36	13457-18-6
PYRAZOXYFENE	Herbicide	1644	403.26	71561-11-0
PYRIDABENE	Insecticide	161	364.93	96489-71-3
PYRIDAFENTHION	Insecticide	769	340.33	119-12-0
PYRIDAFENTHION	Insecticide	769	340.33	119-12-0
PYRIFENOX	Fongicide	2912	295.16	88283-41-4
PYRIMETHANIL	Fongicide	4150	199.11	53112-28-0
PYROQUILONE	Fongicide	321	173.21	57369-32-1
QUINALPHOS	Insecticide	71	298.3	13593-03-8
QUINCLORAC	Herbicide	2680	242.06	84087-01-4
SETHOXYDIME	Herbicide	2676	232.32	1982-49-6
SIMECONAZOLE	Fongicide	611	213.30	1014-70-6
SIMETRYNE	Herbicide	750	310.47	65733-16-6
SULFOTEP	Insecticide	5	277.27	946578-00-3
SULFOXAFLOR	Insecticide	1000	322.45	35400-43-2

Pesticide	Type	LD50 mg/Kg	Masse Molaire (g/mol)	N° de CAS
SULPROFOS	Insecticide	176	307.82	107534-96-3
TEBUCONAZOLE	Fongicide	1700	352.47	112410-23-8
TEBUPIRIMFOS	Insecticide	1,3	233.35	35256-85-0
TEBUTAME	Herbicide	6210	228.31	34014-18-1
TEBUTHIURON	Herbicide	644	447.91	76280-91-6
TECLOFTALAM	Fongicide	2340	260.89	117-18-0
TECNAZENE	Fongicide	1256	381.11	83121-18-0
TEFLUTHRINE	Insecticide	21,8	440.82	335104-84-2
TEMEPHOS	Insecticide	4204	341.8	149979-41-9
TEPRALOXYDIM	Herbicide	5000	216.07	5902-51-2
TERBUPHOS	Insecticide	1,3	229.71	5915-41-3
TETRACONAZOLE	Fongicide	1031	356.06	116-29-0
THIACLOPRID	Insecticide	444	291.71	153719-23-4
THIOBENCARBE	Herbicide	560	201.25	25171-63-5
THIOCARBOXIME	Insecticide	12,6	181.3	31895-21-3
THIOCYCLAME	Insecticide	370	354.47	59669-26-0
THIODICARBE	Insecticide	50	218.32	39196-18-4
THIOFANOX	Insecticide	8,5	246.35	640-15-3
THIOMETON	Insecticide	40	248.24	297-97-2
TOLFENPYRAD	Insecticide	386	347.27	731-27-1
TRALKOXYDIME	Herbicide	934	665.0	66841-25-6
TRALOMETHRINE	Insecticide	99	371.15	118712-89-3
TRIACETATE DE GUAZATINE	Fongicide	308	293.8	43121-43-3
TRIADIMEFON	Fongicide	300	295.76	55219-65-3
TRIALLATE	Herbicide	1100	401.8	82097-50-5
TRIAZAMATE	Insecticide	61	313.3	24017-47-8
TRIAZOPHOS	Insecticide	66	247.64	72459-58-6
TRIAZOXIDE	Fongicide	98	381.37	106040-48-6
TRIBUFOS	Herbicide	234	257.4	52-68-6
TRICHLORFON	Insecticide	212	333.6	327-98-0
TRICHLORONATE	Insecticide	23,7	256.5	55335-06-3
TRICLOPYR	Herbicide	630	256.5	55335-06-3
TRICYCLAZOLE	Fongicide	290	297.52	81412-43-3
TRIDEMORPHE	Fongicide	500	320.43	58138-08-2
TRIDIPHANE	Herbicide	1740	229.71	1912-26-1
TRIETAZINE	Herbicide	494	408.37	141517-21-7
TRIFLUMIZOLE	Fongicide	1057	358.70	64628-44-0
TRIMETHACARBE	Insecticide	566	317.81	131983-72-7
VAMIDOTHION	Insecticide	64	287.34	1929-77-7
VERNOLATE	Herbicide	1500	286.11	50471-44-8
ZIRAME	Fongicide	267	336.64	156052-68-5

ANNEXE B

Pesticide	Type	di	hii
Ensemble d'apprentissage			
1,2 DICHLOROPROPANE	Insecticide	0,112	0,010
2,4,5-TRICHLOROPHENOL	Herbicide	-1,798	0,005
2,4-DB	Herbicide	-0,112	0,005
2,4-DIMETHYLPHENOL	Fongicide	-0,112	0,004
2-AMINO BUTANE	Fongicide	-0,506	0,005
ACEPHATE	Insecticide	0,112	0,006
ACETAMIPRIDE	Insecticide	0,562	0,003
ACETOCHLORE	Herbicide	-0,056	0,007
ACIDE 4 CHLOROPHENOXYACETIQUE	Herbicide	-0,169	0,005
ACROLEINE	Herbicide	-0,169	0,003
ALACHLORE	Herbicide	0,955	0,005
ALANYCARBE	Insecticide	-0,225	0,003
ALDICARBE	Insecticide	-0,562	0,024
ALDRINE	Insecticide	0,169	0,006
ALLYXYCARB	Insecticide	1,236	0,004
ALPHAENDOSULFAN	Insecticide	0,730	0,007
AMICARBAZONE	Herbicide	0,730	0,005
AMIDITHION	Insecticide	-1,461	0,004
AMINOCARBE	Insecticide	0,112	0,006
AMIPROFOS-METHYL	Herbicide	0,169	0,003
AMITRAZE	Insecticide	-0,393	0,004
ANCYMIDOLE	Herbicide	-0,337	0,006
ANILAZINE	Fongicide	1,067	0,010
ANILOFOS	Herbicide	-1,348	0,003
ASOMATE	Fongicide	-0,562	0,003
AZACONAZOLE	Fongicide	-0,169	0,003
AZAMETIPHOS	Insecticide	0,169	0,005
AZINPHOS-METHYL	Insecticide	0,393	0,012
BENALAXIL	Fongicide	1,854	0,004
BENDIOCARBE	Insecticide	-0,056	0,005
BENFURACARBE	Insecticide	1,685	0,003
BENQUINOX	Fongicide	0,127	0,003

Pesticide	Type	di	hii
BENTAZONE	Herbicide	-0,269	0,004
BENZTHIAZURON	Herbicide	-1,742	0,006
BINAPACRYL	Fongicide	1,685	0,005
BRODIFACOUM	Rodonticide	-0,225	0,042
BROMACIL	Herbicide	-0,562	0,005
BROMOCYCLENE	Insecticide	-0,393	0,013
BROMOPHOS	Insecticide	0,506	0,005
BROMOPHOS-ETHYL	Insecticide	-0,169	0,006
BROMOXYNIL	Herbicide	-1,854	0,004
BROMOXYNIL HEPTANOATE	Herbicide	0,205	0,003
BROMOXYNIL OCTANOATE	Herbicide	0,331	0,003
BROMUCONAZOLE	Fongicide	-1,063	0,003
BRONOPOL	Fongicide	-0,337	0,003
BUPIRIMATE	Fongicide	-0,787	0,009
BUTACHLORE	Herbicide	0,899	0,006
BUTAMIFOS	Herbicide	-1,180	0,003
BUTILATE	Herbicide	0,449	0,010
BUTOCARBOXIME	Insecticide	0,843	0,003
BUTONATE	Insecticide	-0,281	0,004
BUTOXYCARBOXIME	Insecticide	-1,180	0,004
BUTRALINE	Herbicide	0,618	0,005
CADUSAFOS	Insecticide	-0,562	0,006
CAMPHECHLORE	Insecticide	2,865	0,006
CARBANOLATE	Insecticide	0,169	0,006
CARBARYL	Insecticide	-0,506	0,004
CARBETAMIDE	Herbicide	0,843	0,007
CARBOFURAN	Insecticide	0,674	0,012
CARBOPHENOTHION	Insecticide	0,562	0,012
CARBOSULFAN	Insecticide	-1,348	0,004
CARBOXINE	Fongicide	-1,075	0,008
CHLORDANE	Insecticide	-0,618	0,003
CHLORDECONE	Insecticide	-0,225	0,005
CHLORETHOXYFOS	Insecticide	-0,056	0,023
CHLORFENAC	Herbicide	-0,843	0,007
CHLORFENETHOL	Insecticide	1,461	0,003
CHLORIDAZONE	Herbicide	0,169	0,008
CHLOROBENZILATE	Insecticide	0,112	0,007
CHLOROMETHIURON	Insecticide	0,337	0,021
CHLOROPHACINONE	Rodonticide	0,562	0,020
CHLOROPICRINE	Insecticide	0,225	0,003
CHLORPYRIPHOS	Insecticide	1,236	0,005

Pesticide	Type	di	hii
CHLORPYRIPHOS-METHYL	Insecticide	0,730	0,007
CHLORTHIAMIDE	Herbicide	-0,112	0,005
CHLORTHION	Insecticide	1,180	0,004
CLETHODIM	Herbicide	-0,663	0,004
CLODINAFOP-PROPARGYL	Herbicide	0,281	0,005
CLOETHOCARB	Insecticide	-1,236	0,006
CLOMAZONE	Herbicide	-1,180	0,006
COUMACHLORE	Rodonticide	1,124	0,010
CROTOXYPHOS	Insecticide	-0,955	0,005
CYANAZINE	Herbicide	0,843	0,003
CYANOPHOS	Insecticide	-0,562	0,004
CYCLOXYDIME	Herbicide	0,393	0,008
CYHEXATIN	Insecticide	-1,404	0,003
CYMOXANIL	Fongicide	1,180	0,005
CYPERMETHRINE	Insecticide	0,281	0,003
CYPHENOTHRIN	Insecticide	0,674	0,003
CYPROFURAM	Fongicide	-1,031	0,003
CYROMAZINE	Insecticide	0,506	0,011
DALAPON	Herbicide	-0,506	0,018
DAZOMET	Insecticide	-2,022	0,004
DELTAMETHRINE	Insecticide	-1,404	0,005
DEMETON-S-METHYL SULFONE	Insecticide	-0,955	0,006
DESMETRYNE	Herbicide	0,393	0,006
DIAFENTHIURON	Insecticide	-0,449	0,006
DIALLATE	Herbicide	1,573	0,003
DIBROMOCHLOROPROPANE	Insecticide	1,573	0,003
DICHLONE	Fongicide	-0,337	0,003
DICHLORPROP	Herbicide	0,562	0,005
DICHLORVOS	Insecticide	-0,674	0,004
DICOFANE	Insecticide	0,843	0,004
DICOFOL	Insecticide	-1,180	0,003
DICROTOPHOS	Insecticide	-0,591	0,008
DIENOCHLORE	Insecticide	0,618	0,006
DIETHATYL ETHYL	Herbicide	-0,730	0,007
DIFENAMIDE	Herbicide	0,393	0,005
DIFLOVIDAZIN	Insecticide	-0,505	0,004
DIFLUMETORIM	Fongicide	1,067	0,003
DIMETACHLORE	Herbicide	-0,674	0,006
DIMETHENAMID	Herbicide	0,337	0,003
DIMETHENAMID-P	Herbicide	0,664	0,003
DIMETHOMORPHE	Fongicide	-0,056	0,008

Pesticide	Type	di	hii
DIMETHYLVINPHOS	Insecticide	-1,011	0,004
DIMEXANO	Herbicide	0,225	0,003
DINOBUTON	Fongicide	0,562	0,003
DINOSEBE	Herbicide	-0,730	0,007
DINOTERBE	Insecticide	-0,056	0,007
DIOXATHION	Insecticide	1,180	0,009
DIPHACINONE	Rodonticide	-0,056	0,021
DIQUAT	Herbicide	0,254	0,003
DITHIANON	Fongicide	0,365	0,003
DIURON	Herbicide	1,124	0,003
EDIFENPHOS	Fongicide	0,820	0,003
ENDOTHAL	Herbicide	1,180	0,004
EPN	Insecticide	0,562	0,010
EPTC	Herbicide	1,011	0,005
ETHANEDIAL	Herbicide	-0,393	0,004
ETHOATE-METHYLE	Insecticide	-1,629	0,003
ETHOXYSULFURON	Herbicide	0,169	0,007
FENAMIDONE	Fongicide	-0,225	0,006
FENCHLORPHOS	Insecticide	-0,112	0,003
FENOBUCARB	Insecticide	-1,461	0,004
FENOPROP	Herbicide	1,404	0,004
FENPROPATHRINE	Insecticide	-0,112	0,004
FENPROPIDINE	Fongicide	-0,112	0,006
FENPROPIMORPHE	Fongicide	-1,180	0,006
FENSULFOTHION	Insecticide	-0,169	0,021
FENTINE ACETATE	Fongicide	-0,843	0,004
FENVALERATE	Insecticide	0,501	0,003
FIPRONIL	Insecticide	-0,955	0,005
FLORASULAM	Herbicide	0,169	0,009
FLUAZIFOP-BUTYL	Herbicide	1,180	0,007
FLUCHLORALINE	Herbicide	-0,277	0,005
FLUCYTHRINATE	Insecticide	-0,056	0,005
FLUFENACET	Herbicide	-0,056	0,003
FLUMORPH	Fongicide	0,843	0,007
FLUOROACETAMIDE	Insecticide	0,618	0,005
FLUQUINCONAZOLE	Fongicide	0,674	0,004
FLUSILAZOLE	Fongicide	-0,334	0,004
FLUVALINATE	Insecticide	0,562	0,003
FOMESAFEN	Herbicide	-0,420	0,004
FONOFOS	Insecticide	0,056	0,012
FORMETANATE	Insecticide	-0,225	0,008

Pesticide	Type	di	hii
FORMOTHION	Insecticide	0,562	0,003
FOSPIRATE	Insecticide	-0,618	0,004
FOSTHIAZATE	Insecticide	-1,124	0,005
FURATHIOCARBE	Insecticide	1,180	0,006
FURFURAL	Fongicide	1,292	0,003
GAMMA-CYHALOTHRINE	Insecticide	0,843	0,006
HALFENPROX	Insecticide	-0,281	0,004
HALOSULFURON-METHYL	Herbicide	-0,787	0,010
HEPTENOPHOS	Insecticide	0,169	0,004
HEXACONAZOLE	Fongicide	0,225	0,006
HEXAZINONE	Herbicide	-0,112	0,006
HYMEXAZOL	Fongicide	1,236	0,010
ICARIDIN	Insecticide	0,449	0,008
IMIPROTHRIN	Insecticide	-0,906	0,004
IOXYNIL	Herbicide	-0,899	0,004
IPROBENFOS	Fongicide	1,236	0,004
ISOCARBOPHOS	Insecticide	1,011	0,005
ISOPROCARBE	Insecticide	0,506	0,004
ISOPROTHIOLANE	Fongicide	2,809	0,005
ISOPROTURON	Herbicide	-1,236	0,007
ISOXATHION	Insecticide	-0,618	0,004
KELEVAN	Insecticide	0,112	0,004
LAMBDA-CYHALOTHRINE	Insecticide	-0,674	0,006
LINDANE	Insecticide	0,618	0,003
LINURON	Herbicide	0,730	0,005
MALATHION	Insecticide	-1,629	0,006
MCPA-THIOETHYL	Herbicide	0,393	0,003
MCPB	Herbicide	0,225	0,010
MECARBAM	Insecticide	0,000	0,006
MEPIQUAT	Herbicide	0,618	0,009
METALAXYL	Fongicide	-0,393	0,004
METAMITRONE	Herbicide	-1,011	0,006
METHOMYL	Insecticide	-1,404	0,005
METOMINOSTROBIN	Fongicide	-0,899	0,004
METSULFOVAX	Fongicide	1,124	0,010
MEVINPHOS	Insecticide	-0,393	0,016
MONOCROTOPHOS	Insecticide	-0,337	0,008
MORPHOTHION	Insecticide	-0,843	0,003
NALED	Insecticide	-0,664	0,004
NAPTALAME	Herbicide	0,506	0,006
NITHIAZINE	Insecticide	-1,236	0,003

Pesticide	Type	di	hii
NITRAPYRINE	Bactéricide	-1,683	0,004
NITROFENE	Herbicide	0,899	0,007
OCTHILINONE	Fongicide	-0,337	0,004
OFURACE	Fongicide	0,506	0,007
OXYCARBOXINE	Fongicide	-3,933	0,006
OXYDEMETON-METHYLE	Insecticide	0,562	0,005
PARAQUAT	Herbicide	-0,506	0,003
PARATHION	Insecticide	-0,674	0,021
PARATHION METHYL	Insecticide	0,562	0,018
PEBULATE	Herbicide	0,787	0,006
PETHOXAMID	Herbicide	-1,685	0,004
PHENKAPTON	Insecticide	2,715	0,006
PHENTHOATE	Insecticide	0,169	0,003
PHOSALONE	Insecticide	-0,225	0,004
PICLORAME	Herbicide	-0,225	0,010
PIPEROPHOS	Herbicide	0,674	0,003
PIRIMICARB	Insecticide	-1,348	0,003
PLIFENATE	Insecticide	-0,337	0,013
PRALLETHRIN	Insecticide	-0,674	0,003
PRETILACHLORE	Herbicide	-0,112	0,011
PROMETONE	Herbicide	-0,337	0,006
PROPANIL	Herbicide	0,225	0,005
PROPARGITE	Insecticide	-0,393	0,007
PROPICONAZOL	Fongicide	-1,107	0,004
PROPOXUR	Insecticide	1,461	0,004
PROSULFURON	Herbicide	0,843	0,003
PROTHIOFOS	Insecticide	1,292	0,004
PYMETROZINE	Insecticide	0,169	0,012
PYRAZOPHOS	Fongicide	3,221	0,003
PYRAZOXYFENE	Herbicide	0,899	0,005
PYRIDABENE	Insecticide	-0,056	0,003
PYRIDAFENTHION	Insecticide	-0,056	0,004
PYRIFENOX	Fongicide	0,449	0,008
PYRIMETHANIL	Fongicide	0,112	0,011
PYROQUILONE	Fongicide	1,404	0,003
QUINALPHOS	Insecticide	1,180	0,004
QUINCLORAC	Herbicide	-0,169	0,008
SETHOXYDIME	Herbicide	0,112	0,008
SIMETRYNE	Herbicide	1,124	0,004
SULFOTEP	Insecticide	0,169	0,015
SULFOXAFLOR	Insecticide	-1,236	0,004

Pesticide	Type	di	hii
SULPROFOS	Insecticide	-0,449	0,003
TEBUCONAZOLE	Fongicide	0,843	0,005
TECLOFTALAM	Fongicide	0,393	0,007
TECNAZENE	Fongicide	-0,281	0,004
TEFLUTHRINE	Insecticide	0,225	0,009
THIOCARBOXIME	Insecticide	1,461	0,008
THIODICARBE	Insecticide	-0,730	0,004
THIOFANOX	Insecticide	0,449	0,011
THIOMETON	Insecticide	0,337	0,005
TOLFENPYRAD	Insecticide	0,416	0,003
TRALKOXYDIME	Herbicide	-0,211	0,003
TRIALLATE	Herbicide	-0,547	0,004
TRIBUFOS	Herbicide	1,067	0,003
TRICHLORFON	Insecticide	0,393	0,003
TRICHLORONATE	Insecticide	0,812	0,007
TRICYCLAZOLE	Fongicide	0,843	0,003
TRIDIPHANE	Herbicide	0,169	0,007
TRIETAZINE	Herbicide	0,562	0,003
TRIFLUMIZOLE	Fongicide	1,180	0,004
TRIMETHACARBE	Insecticide	-0,116	0,003
VAMIDOTHION	Insecticide	-2,079	0,004
Ensemble de validation			
2,4-D	Herbicide	0,337	0,004
ALDOXYCARBE	Insecticide	0,281	0,006
ALLETHRINE	Insecticide	-1,180	0,004
ALPHACYPERMETHRINE	Insecticide	0,000	0,006
AZINPHOS-ETHYL	Insecticide	0,225	0,011
BARBANE	Herbicide	0,000	0,004
BENSULIDE	Herbicide	-0,225	0,003
BENSULTAP	Insecticide	-1,404	0,004
BETACYPERMETHRINE	Insecticide	0,730	0,004
CHLORBROMURON	Herbicide	-0,056	0,007
CHLORBUFAME	Herbicide	-0,843	0,008
CHLORPROPHAME	Herbicide	1,629	0,011
CLOSANTEL	Insecticide	1,573	0,003
CRIMIDINE	Rodonticide	-0,225	0,021
DEMETON-S-METHYL	Insecticide	-1,629	0,005
DICHLORPROP-P	Herbicide	2,135	0,004
DIMETHOATE	Insecticide	0,000	0,003
DINOCAP	Fongicide	0,112	0,004
DIOXABENZOPHOS	Insecticide	-0,112	0,003

Pesticide	Type	di	hii
DITALIMPHOS	Fongicide	-0,393	0,010
DNOC	Herbicide	0,169	0,006
ENDOSULFAN	Insecticide	1,798	0,007
ETACONAZOLE	Fongicide	-0,506	0,005
ETHIOFENCARBE	Insecticide	0,562	0,003
ETHIPROLE	Insecticide	0,000	0,010
FENARIMOL	Fongicide	1,292	0,007
FENAZAQUIN	Acaricide	-0,112	0,003
FENITROTHION	Insecticide	0,000	0,003
FLONICAMIDE	Insecticide	-0,674	0,005
FLUAZIFOP-P-BUTYL	Herbicide	-1,685	0,006
FLUOROGLYCOPHENE	Herbicide	-0,674	0,005
FURALAXYL	Fongicide	1,404	0,004
FURMECYCLOX	Fongicide	1,011	0,009
GLUFOSINATE	Herbicide	-0,225	0,007
GLUTARALDEHYDE	Fongicide	-1,011	0,003
HALOFENOZIDE	Insecticide	0,787	0,007
IMAZALIL	Fongicide	-1,292	0,003
INDOXACARBE	Insecticide	0,562	0,003
ISOPHENPHOS-METHYLE	Insecticide	-1,011	0,008
LEPTOPHOS	Insecticide	1,236	0,007
MCPA	Herbicide	0,618	0,005
MECOPROP	Herbicide	2,022	0,006
METAZACHLORE	Herbicide	-0,843	0,009
METCONAZOLE	Fongicide	-1,180	0,003
METHAZOLE	Herbicide	0,787	0,004
METHIDATHION	Insecticide	0,281	0,007
METOLACHLOR	Herbicide	0,000	0,005
METRIBUZINE	Herbicide	1,742	0,005
MOLINATE	Herbicide	-1,292	0,004
MONOLINURON	Herbicide	-2,022	0,008
NITENPYRAM	Insecticide	0,506	0,006
OXADIXYL	Fongicide	1,685	0,006
OXAMYL	Insecticide	-0,562	0,018
PENDIMETHALINE	Herbicide	1,236	0,008
PHOSMET	Insecticide	2,303	0,004
PROFENOPHOS	Insecticide	0,169	0,003
PROMECARBE	Insecticide	1,067	0,005
PROPAZINE	Herbicide	1,517	0,010
PROSULFOCARBE	Herbicide	-0,337	0,007
PROTHOATE	Insecticide	2,303	0,012
TEBUTAME	Herbicide	0,449	0,012
TEBUTHIURON	Herbicide	-0,506	0,003

Pesticide	Type	di	hii
TEPRALOXYDIM	Herbicide	0,674	0,011
TERBUPHOS	Insecticide	0,618	0,023
TETRACONAZOLE	Fongicide	1,854	0,004
THIACLOPRID	Insecticide	0,337	0,003
THIOBENCARBE	Herbicide	1,236	0,004
TRALOMETHRINE	Insecticide	-0,449	0,004
TRIAZAMATE	Insecticide	1,236	0,005
TRIDEMORPHE	Fongicide	0,674	0,003
VERNOLATE	Herbicide	-3,371	0,006

ANNEXE C

Descripteurs sélectionnés pour le modèle QSAR des pesticides

Descripteurs	Signification	Catégorie
MATS2p	Moran autocorrelation - lag 2 / weighted by atomic polarizabilities	2D Autocorrelations indices
MATS1m	Moran autocorrelation - lag 1 / weighted by atomic masses	
N-072	RCO-N< / >N-X=X	Atom-Centred fragments
H-046	H^a attached to $C^0(sp^3)$ no X attached to next C	
PJI3	3D Petijean shape index	Geometrical descriptors
H6m	H autocorrelation of lag 6 / weighted by atomic masses	Getaway descriptors
HATSe	Leverage-weighted total index / weighted by atomic Sanderson electronegativities	
HATS0m	Leverage-weighted autocorrelation of lag 0 / weighted by atomic masses	
RDF030e	Radial Distribution Function - 3.0 / weighted by atomic Sanderson electronegativities	RDF descriptor
RDF020e	Radial Distribution Function - 2.0 / weighted by atomic Sanderson electronegativities	
Mor15m	3D-MoRSE - signal 15 / weighted by atomic masses	3D-Morse descriptor
Mor23u	3D-MoRSE - signal 23 / unweighted	
Mor26u	3D-MoRSE - signal 26 / unweighted	
Du	D total accessibility index / unweighted	Whim descriptors
E1u	1st component accessibility directional WHIM index / unweighted	
nArX	Number of X on aromatic ring	Functional group counts
nS	Number of Sulfur atoms	Constitutional descriptors