



MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR
ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITÉ «Abbès LAGHROUR» DE KHENCHELA
FACULTÉ DES SCIENCES ET TECHNOLOGIE



Département des Sciences de la Matière

N° de série :.....

Mémoire de fin d'études

Pour l'obtention du diplôme de Master (L.M.D)

Filière : Chimie.

Spécialité : Chimie Analytique et Environnement

Intitulé :

*Contrôle de la pollution des eaux par les
pesticides: modèle QSPR.*

Réalisé par :

-HADIDI NIAMA

-MEFARDJI FAYZA

Dirigé par : M^{elle} BOUAKKADIA A

Membres de jury :

HAKKAR Farida

MAA

UA L Khenchela

Présidente

BOUAKKADIA Amel

MCB

UA L Khenchela

Encadreur

KERTIOU Nour Eddine

MAA

UA L Khenchela

Examinateur

Promotion: 2017/2018

Résumé	
Abstract	
Liste des tableaux	
Liste des figures	
Symboles et abréviation	
Sommaire	
Introduction Générale	2
Etude bibliographique	
I- Généralités sur les pesticides :	4
I- 1- Introduction.....	4
I- 2- Historique.....	4
I- 3- Définition des pesticides	7
I- 4- Classification des pesticides.....	7
I. 4- 1- Premier système de classification	8
I- 4- 2- Deuxième système de classification	8
I- 5- Modes d'action des pesticides.....	9
I. 6- Conception des pesticides:	10
I- 7- l'utilisation des pesticides:.....	10
I- 8- Toxicité des pesticides :	10
I- 8- Le marché des pesticides.....	11
I- 8- 1- dans le Monde:.....	11
I- 9- Les pesticides et l'environnement:	13
I- 9-1- Contamination des milieux par les pesticides:.....	13
a- Contamination des eaux superficielles et souterraines :	14
b- Contamination des sols :.....	14
c- Contamination de L'air	14
II- Modélisation QSPR	16
II- 1- Généralités sur la modélisation QSPR/QSAR	16
II- 2- Principe des méthodes QSPR/QSAR	17
II- 3- Méthodologie générale d'une étude QSPR/QSAR	17
II- 4- Applications des méthodes QSPR/QSAR.....	18
II- 5- Les Méthodes Mathématiques Utilisés par le Model QSPR :	19
II- 5- 1- Régression Linéaire Multiple (MLR) :	19
II- 5- 2- Régression en Composantes Principales (RCP) :	22

II- 6- Calcule Des Descripteurs Moléculaire :.....	22
II- 7- Les types d'apprentissages d'un modèle QSPR :.....	23
II- 7- 1- Sélection d'un sous-ensemble de variables par algorithme génétique (GA-VSS) :.....	23
Application	
I- Calcul du modèle :	31
I- 1- Choix du modèle :	31
I- 2- Choix d'une famille de descripteurs :	31
II- Développement du modèle :	33
II- 1- Matrice de corrélation :	33
II- 2- Equation et analyse de régression :	34
II- 3- Diagramme de Williams	35
II- 4- Qualité de l'ajustement	36
II- 5- Test de randomisation	36
Conclusion générale	38
Références bibliographiques	39

LISTE DES TABLEAUX

Tableau I : Développement des substances actives herbicides, fongicides et insecticides durant le XXème siècle [Rapport d'office parlementaire, 2003].	6
Tableau II : les composés étudiés et leurs solubilités	25
Tableau III : Descripteurs moléculaires intervenant dans la modélisation de la solubilité....	32
Tableau IV : Valeurs de log S, MATS6V, BELP3, Mor26p	32
Tableau V : Matrice de corrélation	33
Tableau VI : Analyse de régression.....	34
Tableau VII : Les paramètres statistiques pour l'ensemble de calibration (19 composés) et celui de validation (7 composés),	34

Liste Des Figures

Figure 1 : Evolution du marché mondial des pesticides sur la période 1990-2003 (ORP, 12	12
Figure 2 : Processus et transfert des pesticides dans les sols (Barruiso et al., 1996) 13	13
Figure 3 : Diagramme de Williams 35	35
Figure 4 : Qualité de l'ajustement 36	36
Figure 5 : Test de randomisation 37	37

SYMBOLES ET ABREVIATION

AM1	Austin Model 1.
EQMC	Erreur quadratique moyenne sur l'ensemble de calibration.
EQMP	Erreur quadratique moyenne sur l'ensemble de prédiction.
EQMP _{ext.} :	Erreur quadratique moyenne sur l'ensemble de prédiction externe.
e_i	Résidu : différence entre les valeurs observées et estimées.
$e_{i \text{ std.}}$	Résidu de prédiction standardisé.
F	Statistique de Fisher.
H	Matrice de projection, ou matrice chapeau.
h_{ii}	Eléments diagonaux de la matrice chapeau.
N	Dimension de la population.
n-p	Nombre de degrés de liberté de la somme des carrés des résidus.
P	Nombre de descripteurs en comptant la constante (nombre de paramètres).
PRESS	Somme des carrés des erreurs de prédiction.
QSAR	Quantitative Structure/ Properties - Relationships.
R ²	Coefficient de détermination.
S	Erreur standard.
SCE	Somme des carrés des écarts.
SCT	Somme des carrés totaux.
TEPP	Tétra-éthylpyrophosphate
Q ²	Coefficient de prédiction.
X	Matrice des valeurs observées des variables explicatives.

X'	Matrice transposée de X
VIF:	Facteur d'inflation de la variance.
GA:	Algorithme génétique
LMO:	Validation croisée par omission d'un ensemble d'observations.
OCDE	Organisation de Coopération et de Développement Economiques
QSPR :	Quantitative Structure/ Property Relationships.
RLM:	Régression linéaire multiple.
ACP	d'analyse en composantes principales
LUMO	la plus basse orbitale moléculaire inoccupée
HOMO	l'énergie de la plus haute orbitale moléculaire occupée

Résumés

La solubilité d'un ensemble homogène de quelques composés polluants ont été corrélées linéairement avec des descripteurs moléculaires théoriques sélectionnés par algorithmes génétiques, en utilisant une approche QSPR

La série des données a été divisé aléatoirement en deux sous ensembles, un ensemble de 19 composés pour la construction du modèle et un ensemble de 7 composés pour la validation.

Les différentes statistiques établies pour les ensembles de calibration et de validation (coefficient de détermination et de prédiction : R^2 ; Q^2 et écart type s) seront confirmées.

Le domaine d'application a été discuté à l'aide de diagramme de Williams.

Abstract

Solubility of a continuum of some polluting compounds was linearly correlated with theoretical molecular descriptors selected by genetic algorithms, using a QSPR approach.

The series of the data was divided randomly into two sets, a set of 19 compounds for the model and a set of 7 compounds for the validation

The applicability domain has been discussed with Williams's diagram.

الذوبانية لسلسلة متجانسة متصلة مع بعض المركبات الملوثة مرتبطة خطيا مع الواصفات الجزيئية لنظري
بواسطة الخوارزميات الجينية، وذلك باستعمال مقارنة QSPR

ة الثانية تشمل

26

سلسلة المعطيات مقسمة عشوائيا الى مجموعتين:
7 مركبات لتحقيقه.

تطبيق المناقش بمساعدة تمثيل ويليام.

Introduction Générale

Actuellement, le terme « pesticide » est une appellation générique désignant toutes les substances naturelles ou synthétiques utilisées pour la prévention, le contrôle ou l'élimination d'organismes (microorganismes, animaux ou végétaux) jugés indésirables ou nuisibles pour l'agriculture, mais également pour d'autres applications (hygiène et santé publiques, soins vétérinaires, traitements de surfaces non-agricoles...) (Institut Pasteur, 1999 ; Aubertot *et al.*, 2005; ORP, 2008).

On distingue les deux expressions suivantes :

- Produits phytopharmaceutiques : pesticides à usages agricole et non agricole (parcs et jardins, entretien des abords d'axes de transport...) pour la protection des plantes;
- Produits biocides: pesticides destinés à tous les usages non agricoles hormis ceux voués à la protection des végétaux (désinfectants, insecticides ménagers, traitement des charpentes, produits antialissure...).

Il existe, parallèlement, de nombreuses autres appellations telles que produits phytosanitaires, produits agro pharmaceutiques, produits antiparasitaires ou encore produits de protection des plantes.

Notons que toutes ces expressions peuvent aussi bien désigner les substances actives (molécules) responsables de l'action visée que les formulations commerciales. Celles-ci sont constituées d'un mélange de deux ou de plusieurs substances dont au moins une substance active et un adjuvant (support ou solvant pouvant présenter les mêmes risques de toxicité que les matières actives).

Les pesticides peuvent être groupés de différentes manières: par famille chimique, par le type de formulation (liquide, granulé...), etc. Toutefois, la classification la plus courante est la classification de ces produits selon la nature de l'espèce nuisible sur laquelle ils doivent agir (Sanusi, 1996; Rapport d'office parlementaire, 2003). On retrouve ainsi les termes: d'herbicides (contre les " mauvaises herbes "), d'insecticides (contre les insectes, leurs larves et leurs oeufs), de fongicides (contre les champignons, virus, et bactéries), de molluscicides (contre les limaces et escargots), de rodenticides (contre les taupes et rongeurs), de corvicides (contre les oiseaux ravageurs), de nématocides (contre les vers), d'acaricides (contre les arachnides).

Ainsi, on désigne par pesticide toute substance destinée à repousser, détruire ou combattre les ravageurs et les espèces indésirables de plantes ou d'animaux. Sont également inclus les régulateurs de croissance des plantes, les défoliants (pour faire tomber les feuilles), les

dessicants (pour absorber l'humidité), les agents qui réduisent le nombre de fruits ou évitent leur chute précoce ainsi que les substances appliquées avant ou après récolte pour conserver les produits pendant leur stockage et leur transport (rapport d'information sur les pesticides – Assemblée Nationale, 02/06/2009).

Ainsi, l'objectif principal de ce travail est d'étudier le contrôle des pollutions d'eau par les pesticides.

Ce mémoire est constitué de deux parties:

- ✓ A la suite d'une introduction générale, la première partie est une étude bibliographique qui consiste en une synthèse des connaissances sur les pesticides et aussi une étude consacrée à différentes méthodes de traitement QSPR.
- ✓ La deuxième partie présente la partie application

Le mémoire est enfin complet par une conclusion générale.

I- Généralités sur les pesticides :

I- 1- Introduction

Les pesticides ont été appliqués de manière préventive afin de repousser ou d'atténuer les effets des organismes nuisibles. Bien que la plupart d'entre eux aient été interdits dans de nombreux pays en raison d'effets mutagènes et cancérigènes, les pesticides et leurs métabolites sont toujours présents dans l'environnement, en particulier dans les sols et les sédiments, en raison de leur persistance et leur propriétés lipophiles (A TOR *et al.*, 2006).

La quantité de pesticides en contact direct avec les microorganismes ciblés est extrêmement faible par rapport à la quantité appliquée. Des effets secondaires indésirables peuvent alors se produire sur certaines espèces, sur les communautés ou sur l'écosystème (A HAYO, 1996).

Le lessivage des pesticides dans le sol et le transport des matières actives par le vent dans les zones parfois même inaccessibles en sont des exemples concrets et très inquiétants (présence de DDT au Pôle Nord). Leur degré de concentration reste fonction des conditions physico-chimiques et topographiques du terrain, ainsi que d'autres considérations telles que la nature et la composition du sol et le climat (S SEBIH, 2008). Dans une étude portant sur l'évaluation du degré de pollution de 23 échantillons des eaux superficielles et souterraines, on soulignait que 50% des eaux analysées contenaient de l'atrazine et indiquait la présence d'autres pesticides (lindane, alachlore, dinoterbe...) à des fréquences moins importantes (M O S OULD KANKOU, 2004).

I- 2- Historique

En devenant agriculteur, l'Homme a rassemblé sur un lieu donné certaines espèces végétales en favorisant par la même occasion la concentration de leurs déprédateurs. Pendant des siècles, les malheurs survenus aux cultures ont été attribués à une punition divine ou à un sortilège, il n'y avait d'autre remède que les prières ou la magie. Au fil du temps, des moyens artificiels ont été mis en place pour contrôler la prolifération des espèces nuisibles envers les cultures (J Fournier, 1988) Les premières descriptions de ces dispositifs remontent à 1000 ans avant Jésus-Christ (J -C) avec l'utilisation du soufre. Ainsi, pour empêcher la pyrale de naître

dans les vignes, Pline (naturaliste romain né en 23 après (J-C) proposait la préparation suivante :

«Faites cuire deux congés (mesure de capacité chez les Romains valant 3,24 L) de marc d'olives jusqu'à consistance de miel, puis faites cuire le résidu avec un tiers de bitume et un quart de soufre, en plein air, car à l'intérieur on aurait à craindre le feu. Oignez la vigne avec ce mélange à la cime et aux aisselles; de cette façon il n'y aura pas de pyrales.» (Histoire Naturelle de Pline l'Ancien, livre 17) L'arsenic et la nicotine sont connus dès le XVI^{ème} siècle en Chine (Institut Pasteur, 1999). Les propriétés insecticides du tabac sont utilisées en 1763 par des arboriculteurs de Montreuil afin d'éliminer les pucerons qui ont envahis leurs pêchers (J Fournier, 1988) Le chlorure mercurique est proposé au XVIII^{ème} siècle pour protéger les bois.

La lutte chimique s'est réellement développée à partir du XIX^{ème} siècle avec des produits d'origine naturelle comme la roténone. Extraite des racines de *Derris* (une légumineuse), elle a pour effet de paralyser le système nerveux des insectes et des animaux à sang froid. Le pyrèthre, mélange d'esters contenus dans les fleurs de certains chrysanthèmes, est introduit comme insecticide (J Fournier, 1988; ORP, 2008). A la même époque, l'utilisation plus généralisée des pesticides a également suivi les progrès de la chimie minérale. Les traitements fongiques à base de sulfate de cuivre (dont la célèbre bouillie bordelaise, mélange de sulfate de cuivre et de chaux) se répandent, notamment pour protéger la vigne du Mildiou.

Les sels de mercures sont utilisés au début du XX^{ème} siècle pour le traitement des semences.

Autour des années 1920, on s'aperçoit que les fruits et les légumes traités avec des produits arsenicaux recèlent des doses d'insecticide pouvant être mortelles pour les consommateurs.

On se met alors activement à chercher de nouveaux produits moins dangereux pour la santé humaine (J Fournier, 1988)

Le début de l'ère des pesticides organiques se situe autour des années 1930, parallèlement au développement de la chimie organique de synthèse. Les recherches militaires avaient déjà perfectionné des gaz de combat qui, à défaut d'être utilisés pendant les hostilités, le furent contre les insectes (ORP, 2008). En 1874, Zeidler synthétise le DDT (dichlorodiphényltrichloroéthane), dont Müller, en 1939, établit les propriétés insecticides (Fournier J, 1988). Le DDT, commercialisé dès 1943, ouvre la voie à la famille des organochlorés et domine le marché des insecticides jusqu'au début des années 1970. Il est utilisé en grandes quantités en médecine préventive (destruction du moustique responsable de

la malaria) et en agriculture (élimination du doryphore) (ORP, 2008). En 1944, l'herbicide 2,4-D, copié d'une hormone de croissance de plante et encore fréquemment employé de nos jours, est synthétisé. Depuis, la chimie n'a cessé de proposer de nouvelles molécules.

En 1950-55, apparaissent aux Etats-Unis, les herbicides de la famille des urées substituées (linuron, diuron), suivi peu après par les triazines. Les fongicides du type imidazoliques et triazoliques, datant des années 70, représentent actuellement le plus gros marché des fongicides. Les insecticides pyréthrinoïdes, mis au point dans les années 1970-80, dominent, pour leur part, le marché des insecticides. Peu à peu les applications se diversifient et des biocides sont élaborés pour l'industrie textile et du bois, pour les usages domestiques (aérosols tue-mouches,...), pour l'entretien des routes et pour une utilisation en médecine.

Initialement, la recherche des matières actives se faisait au hasard en soumettant de nombreux produits à des tests biologiques. Désormais, l'accent est mis sur la compréhension des modes d'action et la recherche de cibles nouvelles telles que des cibles physiologiques de l'animal ou du végétal (J Fournier, 1988)

Tableau I : Développement des substances actives herbicides, fongicides et insecticides durant le XX^{ème} siècle [Rapport d'office parlementaire, 2003].

	Herbicides	Fongicides	Insecticides
Avant 1900	Sulfate de cuivre, Sulfate de fer	Soufre, Sels de cuivre	Nicotine
1900-1920	Acide sulfurique	-	Sels d'arsenic
1920-1940	Colorants niters	-	-
1940-1950	Phytohormones...	-	Organochlorés, Organophosphorés
1950-1960	Triazines, Urées substituées, Carbamates	Dithiocarbamates, Phtalimides	Carbamates
1960-1970	Dipyridyles, Toluidines...	Benzimidazoles	-
1970-1980	Amino- phosphonates, Propionates...	Triazoles, Dicarboximides, Amides, Phosphites, Morholines	Pyréthrinoïdes, Benzoyl-urées

	Herbicides	Fongicides	Insecticides
1980-1990	Sulfonyl urées...	-	-
1990-2000	-	Phénylpyrroles, Strobilurines	-

I- 3- Définition des pesticides

Le mot pesticide composé de deux parties: le suffixe «-cide» qui a pour origine le verbe latin «caedo, cadere» qui signifie «tuer». On lui a adjoint la racine anglaise «pest» qui signifie animal ou plantes nuisibles à la culture. (B López *et al.*, 2005)

L'article 2 de la loi algérienne du journal officiel N° 87-17 du 1 Août 1987 relative à la protection phytosanitaire désigne par pesticide:

« Toute substance ou mélange de substances destiné à repousser, détruire ou combattre les organismes nuisibles, en vue de la protection ou de l'amélioration de la production végétale». (Loi algérienne N° 87-17, 1987).

Un pesticide est une substance sensée prévenir, détruire, repousser ou contrôler :

- tout ravageur animal
- toute maladie causée par des microorganismes
- les mauvaises herbes.

Leur action peut se faire par :

- le contact
- Ingestion
- Autres sortes d'expositions effectives pendant les phases de croissance. (J BOLAND *et al.*, 2004).

I- 4- Classification des pesticides

En général les substances actives sont classées (R CALVET, 2005; E BARRIUSSO *et al.*, 2005) en fonction de:

- la nature de l'espèce à combattre (premier système de classification).
- la nature chimique de la principale substance active (deuxième système de classification).

I. 4- 1- Premier système de classification

Il repose sur le type de parasites à contrôler. Il existe principalement trois grandes familles d'activités (K EL MRABET, 2006):

-Les Herbicides:

Ce sont les plus utilisés dans le monde en tonnage et en surface; ils permettent d'éliminer les mauvaises herbes des cultures.

-Les Insecticides:

Ce sont les premiers pesticides utilisés et les plus utilisés en Algérie. Ils sont destinés à détruire les insectes nuisibles.

-Les Fongicides:

Ils permettent de lutter contre les maladies cryptogamiques qui causent de graves dommages aux végétaux cultivés. Ils combattent la prolifération des champignons pathogènes (C G B MARGOUM, 2010).Outre, ces trois grandes familles, d'autres peuvent être citées en exemple:

- ✓ les taupicides contre les taupes
- ✓ les acaricides contre les acariens
- ✓ les rodenticides contre les rongeurs
- ✓ les nématocides contre les nématodes et les vers
- ✓ les molluscicides contre les mollusques, limaces et escargots
- ✓ les corvicides contre les corbeaux et tous les oiseaux ravageurs de cultures.

I- 4- 2- Deuxième système de classification

Le classement se fait en fonction de la nature chimique de la substance active.

On distingue:

- Les Pesticides Organiques

- ✓ organochlorés
- ✓ organophosphorés
- ✓ carbamates
- ✓ triazines

- ✓ urées substituées
- ✓ pyréthrénoïdes

- Les Pesticides Inorganiques:

En général ce sont des éléments chimiques qui ne se dégradent pas. Leur utilisation entraîne souvent de graves effets toxicologiques sur l'environnement par accumulation dans les sols.

Le plomb, l'arsenic et le mercure sont fort toxiques (J BOLAND *et al.*, 2004).

- Les Biopesticides:

Ce sont des substances dérivées de plantes ou d'animaux. Elles peuvent être constituées d'organismes tels que les:

- moisissures
- bactéries
- virus
- nématodes
- composés chimiques dérivés de plantes
- phéromones d'insectes.

La présence de certains groupements fonctionnels et/ou atomes confère aux pesticides certaines propriétés physico-chimiques (ionisabilité, hydrophobie, solubilité, persistance).

Par exemple:

- le groupement donneur ou accepteur de proton d'un pesticide est susceptible de s'ioniser.
- Un pesticide comprenant des atomes de chlore est généralement récalcitrant à la dégradation.

Toutefois, il est important de souligner que la connaissance de la famille chimique à laquelle un pesticide appartient ne suffit pas à elle seule à la définition de ses propriétés ni à la prédiction de son comportement dans l'environnement (A Mokhtari, 2012)

I- 5- Modes d'action des pesticides

- **Les Herbicides** peuvent agir sur les adventices se trouvant en concurrence avec une culture donnée en inhibant la photosynthèse ou les réactions enzymatiques impliquées dans la synthèse des lipides et des acides aminés chez les mauvaises herbes.

- **Les Insecticides**, leurs effets toxiques s'exercent sur les fonctions vitales de l'insecte telles que la transmission de l'influx nerveux et la respiration. Les insecticides agissent par contact, par inhalation ou par ingestion des molécules par l'insecte.
- **Les Fongicides** peuvent contrôler les champignons en affectant leur respiration ou leur division cellulaire ou en inhibant la biosynthèse des acides aminés et des stérols (A Mokhtari, 2012)

I. 6- Conception des pesticides:

Un pesticide est composé d'un ensemble de molécules comprenant:

- une ou plusieurs matières actives à laquelle est du tout ou en partie l'effet toxique.
- Un diluant qui est une matière solide ou liquide (solvant) incorporé à une préparation et destiné à en abaisser la concentration en matière active. Ce sont le plus souvent des huiles végétales dans le cas des liquides, de l'argile ou du talc des solides.
- des adjuvants qui sont des substances dépourvues d'activité biologique mais susceptibles de modifier les qualités du pesticide et d'en faciliter l'utilisation (A Mokhtari, 2012)

I- 7- l'utilisation des pesticides:

- dans l'agriculture: les pesticides sont utilisés pour lutter contre les insectes, les parasites, les champignons et les herbes estimés nuisibles à la production et à la conservation de cultures et produits agricoles ainsi que pour le traitement des locaux.
- dans l'industrie: en vue de la conservation de produits en cours de fabrication (textiles, papiers) vis-à-vis des moisissures dans les circuits de refroidissement, vis-à-vis des algues et pour la désinfection des locaux.
- dans les constructions, pour protéger le bois et les matériaux.
- en médecine: paludisme, malaria, typhus, et autres épidémies (B ZEBOUJJI, 2005)

I- 8- Toxicité des pesticides :

Les pesticides présentent des risques et des dangers pour la santé humaine et l'environnement car ils provoquent la plupart des effets nocifs pour ces derniers. La contamination de l'homme par les pesticides peut se faire par différentes voies. Il peut les absorber via les aliments, l'eau, par contact avec la peau ou encore par inhalation.

En 1975, l'OMS a établi une classification des pesticides en fonction de leur toxicité avec comme critère la dose létale 50 (DL50). Le degré de toxicité des pesticides est étudié sur des rats et animaux de laboratoire au moyen de la DL 50 (dose létale 50). La DL 50 est une caractéristique de la toxicité aiguë; c'est la quantité de pesticide ingérée nécessaire pour provoquer la mort de 50% des rats participant à une expérience en laboratoire. Cette dose létale 50 est exprimée en ppm. (Partie par million, mg par kg).

On distingue différents degrés de toxicité:

- ✓ toxicité aiguë: mort de l'animal.
- ✓ toxicité subaiguë: trouble du métabolisme, du comportement, de la fécondité.
- ✓ toxicité chronique (ou indirecte): résulte de l'absorption répétée de petites doses de produits.

La dose létale 50 ne permet pas de mesurer le danger réel que présentent les pesticides, les expériences sont faites sur des animaux de laboratoire, loin des conditions réelles auxquelles l'homme peut être soumis. La toxicité peut se manifester à long terme, il faut distinguer la toxicité des pesticides pour une dose absorbée en une fois et celle qui résulte d'absorptions répétées.

Suite aux nombreuses constatations faites sur le terrain après les épandages insouciantes de pesticides toxiques effectués ces 20 dernières années, les scientifiques ont mis en évidence trois types d'effets de ces produits sur la faune et l'homme, à savoir:

- ✓ Des effets cancérogènes: provoquant des tumeurs.
- ✓ Des effets mutagènes: entraînant des modifications du matériel génétique de la cellule.
- ✓ Des effets tératogènes: entraînant des malformations de l'embryon.

I- 8- Le marché des pesticides

I- 8- 1- dans le Monde:

Il existe dans le Monde près de 100 000 spécialités commerciales autorisées à la vente. Elles sont composées à partir de 900 matières actives différentes. On enregistre 15 à 20 nouvelles matières actives qui s'y rajoutent chaque année.

Le marché mondial (environ 40 milliards de dollars) est globalement stable depuis quelques années (2000).

Il faut noter que certains événements climatiques récents (chaleur et sécheresse en Europe,

pluie en Océanie) influencent fortement ces chiffres, en Europe et en Amérique du Nord. Les herbicides représentent 70 à 80% des produits utilisés (notamment à cause de la forte augmentation des cultures de maïs) tandis que sous les Tropiques, 50% des produits appliqués sont des insecticides.

La diversification des cultures, avec l'amélioration du niveau de vie dans certains pays, modifie également cet équilibre. Ainsi la Chine a converti l'équivalent de la surface de l'Angleterre de rizières en cultures maraîchères, entraînant une diversification des produits mis en œuvre. (Union des Industries de la Protection des Plantes).

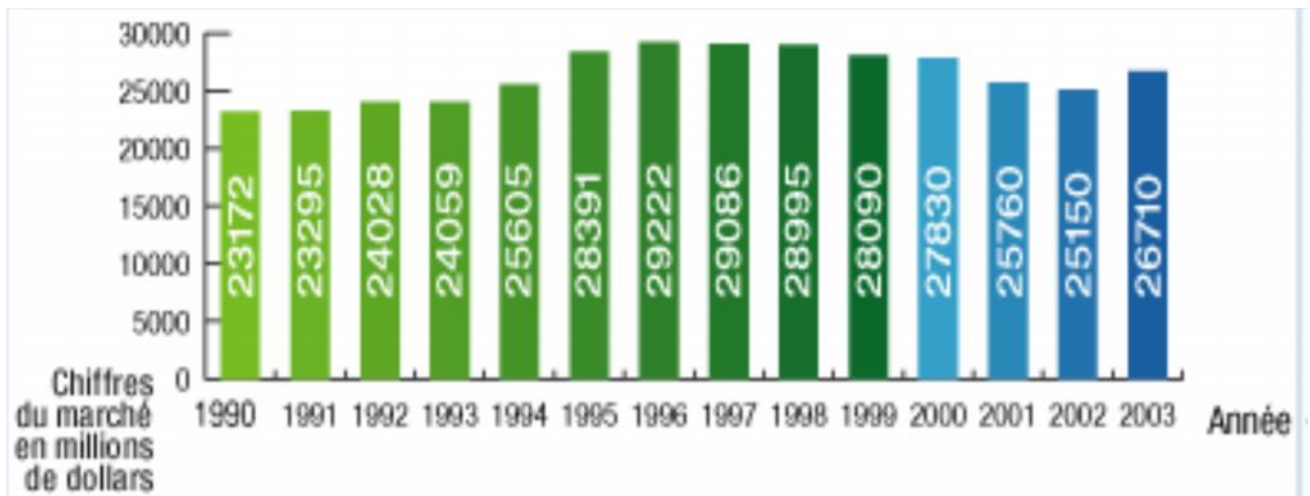


Figure 1: Evolution du marché mondial des pesticides sur la période 1990-2003 (ORP,2008)

I- 8- 2- En Algérie:

Une enquête réalisée par nos soins auprès des fellahs de la chambre d'agriculture d'ORAN et de l'institut de protection des végétaux de la wilaya d'ORAN nous à montrée que les pyrèthrinoïdes, les organophosphorés et les carbamates sont les pesticides les plus utilisés en Algérie.

Selon l'institut nationale de protection des végétaux, la plus grande quantité d'insecticides est utilisée contre la lutte anati acridienne le marché algérien en pesticides ne cesse d'augmenter; en 2009 l'Algérie à importé 67 millions USD de pesticides et en 2008, 77 million USD contre 49,4 millions USD en 2007 (FAOSTAT)-(CENTRE ANTI POISON ALGER, 2011).

I- 9- Les pesticides et l'environnement: (D Devault, 2007)

La contamination des eaux par les pesticides est impliquée dans près d'un quart des cas de non conformité des eaux destinées à l'alimentation humaine (IFEN, 2002). L'IFEN, en 2004, avançait que 80% des eaux de surface et 57% des eaux souterraines analysées présentaient des concentrations en phytosanitaires supérieures au seuil de potabilisation et induisaient la perturbation du milieu aquatique. Parmi ceux-ci, les plus abondamment usités dans les pays développés sont les herbicides. Ces molécules ont un vaste spectre d'usage les rendant ubiquistes puisqu'elles sont employées tant pour l'acquisition de denrées que pour le confort.

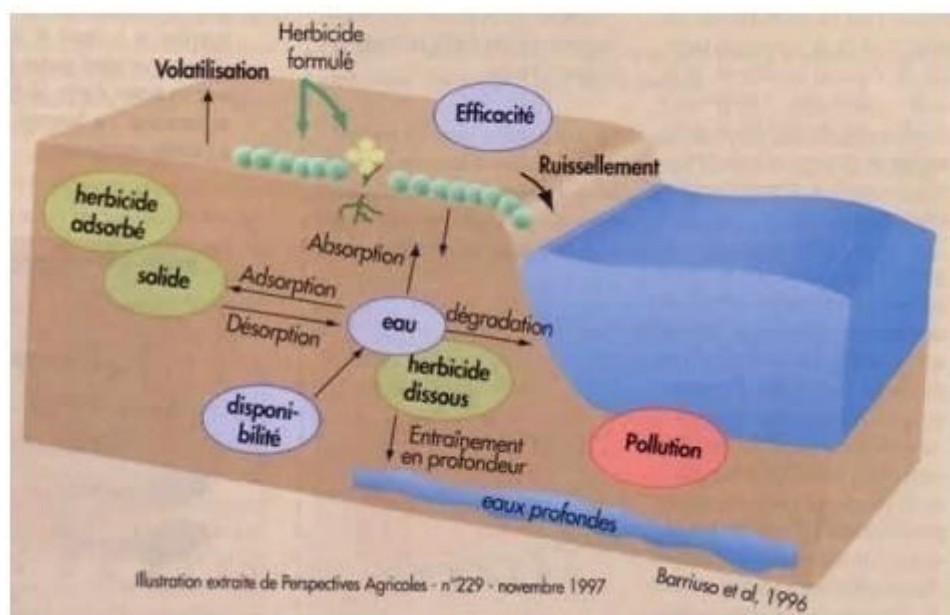


Figure 2 : Processus et transfert des pesticides dans les sols (Barriuso *et al.*, 1996)

I- 9-1- Contamination des milieux par les pesticides:

La "contamination" est définie comme la présence anormale de substances, micro organismes... dans un compartiment de l'environnement. Pour tous les pesticides de synthèse, on peut donc parler formellement de contamination y compris pour les sols agricoles, même si la présence de pesticides y est attendue et volontaire (ce qui n'est pas le cas pour les milieux aquatiques, par exemple). Le terme de "pollution" désigne la présence de substances au-delà d'un seuil pour lequel des effets négatifs sont susceptibles de se produire.

a- Contamination des eaux superficielles et souterraines :

Les eaux continentales sont les milieux pour lesquels les données sont les plus nombreuses, et font l'objet d'une compilation annuelle par l'IFEN. Ces données mettent en évidence une contamination quasi-généralisée des eaux de surface et des eaux souterraines par les pesticides, et la prépondérance des herbicides parmi les molécules les plus fréquemment détectées (du moins à l'échelle des "masses d'eau" au sens de la DCE). Une contamination significative peut être générée par des pertes en pesticides très faible : une fuite de moins de 1/1000e de la masse d'herbicide épanchée sur une parcelle peut suffire par exemple pour contaminer l'eau qui s'en écoule au dessus du seuil de potabilité.

Les données collectées ne permettent toutefois pas de quantifier avec précision les niveaux de contamination, ou de calculer l'exposition des organismes. Elles sont en effet souvent :

- très hétérogènes et difficilement comparables, car produites par des réseaux de suivi divers (ayant des objectifs différents: surveillance de l'eau potable, suivi de bassin versant...), avec des méthodes analytiques et des listes de molécules recherchées différentes... ;
- non représentatives, car les prélèvements sont peu fréquents, et ne permettent notamment pas la détection des pics de pollution notamment dans le cadre de projets soutenus par l'ADEME, mais les données restent fragmentaires (campagnes de suivi ponctuelles dans le temps et dans l'espace) et les listes de molécules suivies sont limitées (enquête SCEES, 2001)

b- Contamination des sols :

Il n'existe pas de dispositif équivalent à ceux relatifs à l'eau et à l'air pour la caractérisation de la contamination des sols par les pesticides. La pollution chronique par certaines substances minérales (cuivre) et l'existence éventuelle de "résidus liés" (c'est-à-dire non extractibles par les méthodes classiques d'analyse) pose la question du risque environnemental à long terme, notamment dans le cas d'une réallocation des terres agricoles à d'autres usages.

Ce risque est illustré par le cas du chlordécone, utilisé de 1972 à 1993 pour la lutte contre le charançon de la banane, en Guadeloupe et Martinique: resté stocké dans les sols, il pollue actuellement (et sans doute encore pour des décennies) les eaux et peut contaminer les productions dans certaines zones (J N Aubertot, 2005)

c- Contamination de L'air

L'air est un compartiment assez mal connu. Ce sont pourtant quelques 15 m³ qui transitent dans nos poumons chaque jour, sans oublier la respiration cutanée. Les mesures

réalisées en France dans l'atmosphère depuis 2000 indiquent que les pesticides sont présents aussi bien dans les zones agricoles qu'en milieu urbain. Des biocides sont également détectés en zone urbaine.

En l'absence de réglementation relative à la présence de pesticides dans l'air que nous respirons, des limites indicatives de concentration sont définies. Ces valeurs limites sont établies à partir d'informations fournies par les tests toxicologiques des substances. Il existe deux types de valeurs en France :

- Ñ la VLCT, valeur limite d'exposition à court terme: Ce sont les valeurs mesurées sur une durée maximale de 15 minutes. Elles sont établies pour prévenir les risques toxiques immédiats ou à court terme ;
- Ñ la VME ou valeur limite moyenne d'exposition. Mesurée ou estimée sur la durée moyenne d'un poste de travail de 8 heures, elles sont indicatives pour la protection des travailleurs à moyen ou long terme.

Le groupe AIR'PHYT CORPEN¹⁵ du a réalisé une synthèse intéressante des connaissances et études les plus récentes sur le sujet : « Les Produits phytosanitaires dans l'air : origine, surveillance et recommandations pratiques en agriculture » (2007). Nous en reprenons ici quelques éléments. Les phénomènes qui expliquent la présence ou la persistance des substances dans l'air sont complexes. Les contaminations ne sont pas toujours liées directement aux applications. Elles peuvent provenir de la volatilisation du produit après application, soit à partir du végétal, soit à partir du sol, soit aussi par érosion éolienne de particules de sol contaminées. Les conditions météorologiques (brouillards, vents, transports par les nuages...) sont également responsables des phénomènes de dérive qui entraînent les substances loin de leur lieu d'application. De plus, une bonne application des produits sur la plante se fait avec un brouillard fin de gouttelettes de produits, ce qui permet une meilleure efficacité des traitements. En revanche, s'il y a moins de produit au sol qu'avec un brouillard lourd, l'atmosphère est davantage chargée (CORPEN, 2007)

L'air extérieur est imprégné de pesticides, à des doses très variables allant de la limite de la détection à de forts pics de concentration comme en zones agricoles ou viticoles lors des périodes d'application. En air intérieur, les résultats surprennent: la présence de nombreuses molécules, essentiellement des biocides, prouve que les substances biocides utilisées et celles venant de l'extérieur restent longtemps présentes en atmosphère confinée. Les substances sont plus ou moins persistantes dans l'air, tant intérieur qu'extérieur: des molécules très

rémanentes comme le lindane, l'alachlore ou la trifluraline sont détectées longtemps après leur utilisation, d'autres sur quelques jours seulement après leur application.

Les quelques études qui existent sur la qualité de l'air au regard des pesticides interrogent et rendent inquiétant le manque de connaissances et de suivis sur ce milieu. Mais que signifient les doses mesurées pour la santé de l'individu, via la respiration? Les contaminations observées en air intérieur sont à mettre en relation avec un usage trop souvent banalisé des biocides. Une meilleure connaissance de la qualité de l'air permettrait de prendre conscience de l'imprégnation potentielle de l'air, et de nous renseigner sur notre propre imprégnation... avant d'en savoir plus sur les risques réels pour notre santé (CORPEN, 2007).

II- Modélisation QSPR

La connaissance des propriétés et des activités est d'une importance capitale pour pouvoir classer et utiliser les composés chimiques. La caractérisation expérimentale complète est difficile, voire impossible, pour des raisons de temps, de coût, de dangerosité de certains essais ou d'éthique (limitations des essais sur les animaux).

L'utilisation des méthodes alternatives à l'expérience est devenue plus qu'indispensable. Parmi ces méthodes, on trouve les méthodes de modélisation moléculaire qui permettent de justifier les données expérimentales disponibles et prédire les propriétés/activités pour des composés nouveaux ou des composés pour lesquels les données expérimentales ne sont pas disponibles. Parmi ces méthodes de modélisation les plus utilisées, on peut citer les méthodes QSPR (Quantitative Structure-Property Relationships) et QSAR (Quantitative Structure Activity Relationships). Ces méthodes s'appuient sur le principe que les propriétés physico chimiques et les activités biologiques des molécules dépendent fortement de leurs structures chimiques.

II- 1- Généralités sur la modélisation QSPR/QSAR

Une relation QSPR/QSAR est un modèle ou une formule mathématique qui permet de relier, d'une manière quantitative, la structure d'une molécule à une propriété ou à une activité donnée. Les méthodes QSPR/QSAR sont de plus en plus utilisées, du fait de la croissance des moyens de calculs (machines, logiciels,...). Récemment, on assiste à la mise en place d'un nouveau règlement REACH (N Margossian, 2007) (Registration, Evaluation, Authorisation and Restriction of Chemicals) qui recommande l'utilisation des méthodes alternatives pour limiter le recours à l'expérimentation.

En fait, les premiers travaux QSPR/QSAR remontent au 19^{ème} siècle. En effet, dès 1868, Crum-Brown et Frasers (A Crum Brown, 1868) ont postulé l'existence de relations entre les activités physiologiques et les structures chimiques en reliant les variations de l'activité biologique à des modifications structurales. Cependant, à cette époque, les structures moléculaires n'étaient pas encore connues.

II- 2- Principe des méthodes QSPR/QSAR

Le principe des méthodes QSPR/QSAR est d'établir une relation mathématique reliant de manière quantitative des propriétés moléculaires, appelées descripteurs, avec une observable macroscopique (activité biologique, toxicité, propriété physico-chimique, etc.), pour une série de composés chimiques similaires à l'aide de méthodes d'analyses de données. La forme générale d'un tel modèle est la suivante :

$$\text{Propriété/ Activité} = f(D1, D2, \dots, Dn, \dots)$$

D1, D2, ...Dn sont des descripteurs des structures moléculaires.

L'objectif d'une telle méthode est d'analyser les données structurales afin de détecter les facteurs déterminants pour la propriété /activité mesurée. Pour ce faire, différents types d'outils statistiques peuvent être employés :

- Régressions linéaires simples et multiples (J Ghasemi *et al.*, 2007)
- Régressions aux moindres carrées partielles (PLS) (P Geladi, B R Kowalski, 1986)
- Arbres de décision (A J Myles *et al.*, 2004)
- Réseaux de neurones (A F Duprat *et al.*, 1998), J Gasteiger J Zupan, 1993)
- Algorithmes génétiques (R Leardi, 2001)
- Vecteurs Machines. (J Gasteiger, J Zupan, 1993)

Une fois cette relation est établie et validée, elle peut alors être employée pour la prédiction de la propriété/activité de nouvelles molécules, pour lesquelles les valeurs expérimentales ne sont pas disponibles. De tels modèles peuvent être également utilisés pour mieux comprendre les mécanismes et les modes d'action.

II- 3- Méthodologie générale d'une étude QSPR/QSAR

La méthodologie générale d'une étude QSAR/QSPR est la suivante:

- a- Constituer une base de données à partir des mesures expérimentales fiables de la propriété ou de l'activité de chaque composé.
- b- Sélectionner les descripteurs en relation avec la propriété ou l'activité étudiée.
- c- Diviser cette base de données, aléatoirement, en une série d'apprentissage (training set) qui contient généralement les 2/3 de la base de données et une série de test (test set) constituée par le 1/3 restant.
- d- Etablir des modèles mathématiques en utilisant la série d'apprentissage.
- e- Caractériser les modèles élaborés par leurs indices de validation internes et vérifier leur robustesse par un test de hasardisation (Randomization) de la variable dépendante Y (réponse).
- f- Valider les modèles élaborés en utilisant la série de test et calculer leurs paramètres statistiques de validation externe.
- g- Elaborer le domaine d'applicabilité du modèle retenu.
- h- Explorer et exploiter les modèles validés pour comprendre les mécanismes et les modes d'action.

II- 4- Applications des méthodes QSPR/QSAR

Les applications des méthodes QSPR/QSAR sont très nombreuses, elles touchent tous les domaines où la structure chimique intervient, entre autre on peut citer:

- **Propriétés physico-chimiques:** point d'ébullition, Point de fusion, densité, indice de réfraction, température critique, viscosité, solubilité, pression de vapeur, tension superficielle, Coefficients de partition: eau/octanol, air/eau, huile/air, lait/plasma. (A R Katritzky *et al.*, 2010).

Ñ **Autres propriétés/activités:**

- Prédiction de la toxicité aquatique des composés chimiques vis-à-vis des espèces environnementales.
- Toxicité des nanoparticules
- Toxicité des pesticides et des colorants
- Propriétés inhibitrices de corrosion
- Concentration micellaire critique
- Prédiction de plusieurs propriétés dangereuses telle que l'explosibilité et l'inflammabilité de certaines familles de molécules chimiques.

- Conception des médicaments et de nombreux autres produits tels que les agents tensio-actifs, parfums, les colorants et les produits de la chimie fine.

II- 5- Les Méthodes Mathématiques Utilisés par le Model QSPR :

II- 5- 1- Régression Linéaire Multiple (MLR) :

Un modèle de régression linéaire multiple entre une variable expliquée Y et p variables explicatives X_1, \dots, X_p , s'écrit pour tout $i=1, \dots, n$:

$$y_i = \beta_0 + \sum_{j=1}^p \beta_j X_{ij} + \epsilon_i \quad (1)$$

Où les $y_i, X_{i1}, X_{i2}, \dots, X_{ip}$ sont des données respectivement relatives aux variables Y, X_1, \dots, X_p .

Les estimateurs β_j sont calculés en utilisant la méthode des moindres carrés ordinaires. Les variables aléatoires ϵ_i représentent les termes d'erreur non observables du modèle.

On peut estimer ces erreurs par les résidus ordinaires e_i , différence entre les valeurs observées y_i et les valeurs estimées \hat{y}_i .

Pour construire le modèle et admettre que les coefficients de la régression sont sans biais et convergents, on montre qu'il faut poser comme hypothèses :

a) Les résidus (e) ont une espérance mathématique nulle :

$$E(\mathbf{e}) = \mathbf{0} \quad (2)$$

b) Le modèle choisi est correct (aucune variable explicative n'a été omise).

c) Les résidus sont indépendants entre eux :

$$E(e_i, e_j) = 0 \quad \text{si } i \neq j \quad (3)$$

leurs covariances sont nulles.

d) Les résidus ont tous même variance σ^2 (propriété d'homoscédasticité).

Par ailleurs, l'emploi de tests statistiques pour analyser la variation expliquée par la régression conduit à admettre que:

e) Les résidus suivent une distribution normale (de Laplace- Gauss).

L'analyse des résidus présente un intérêt à plusieurs égards. Elle permet en effet de vérifier, a posteriori, la validité du modèle utilisé, en ce qui concerne, d'une part la forme de celui-ci (linéarité ou non linéarité de la relation, par exemple) et d'autre part, certaines hypothèses plus spécifiques, telles que l'égalité des variances résiduelles, la normalité des résidus ou l'absence d'auto- corrélation.

Pour minimiser l'influence des erreurs de détermination des valeurs explicatives (ou régresseurs) sur la précision des résultats de la régression 4 à 5 données (variables dépendantes, ou encore observations) doivent, à la limite, être associées à chaque variable explicative. Le nombre de degré de liberté final (n-p-1) doit être (Loi sur l'Air – Convention 18/99, 2000) tel que :

$$n - p - 1 \geq 10 \tag{4}$$

n étant la dimension de l'échantillon, et p le nombre de variables explicatives entrant dans la construction du modèle.

Pour les modèles à plus de deux descripteurs, de faibles coefficients de corrélation croisés n'assurent pas forcément l'orthogonalité des descripteurs. Une indépendance globale acceptable des descripteurs sera vérifiée lorsque les facteurs d'inflation de la variance (FIV) calculés pour chacun d'eux obéissent (R Todeschini *et al.*, 2005) à la condition $FIV < 5$.

Deux paramètres statistiques sont couramment utilisés pour l'évaluation de la qualité du modèle :

- Le coefficient de détermination multiple :

$$R^2 = 1 - \frac{SCE}{SCT} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{5}$$

Où \bar{y} est la valeur moyenne des valeurs observées.

- La racine de l'écart quadratique moyen de prédiction :

$$s_N = EQMP = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_{(i)})^2} = \sqrt{\frac{PRESS}{n}} \quad (6)$$

Il est intéressant de considérer, également, la racine de l'écart quadratique moyen calculé sur les ensembles de calibration (EQMC), et sur l'ensemble de validation externe (EQMP_{ext}) :

$$EQMC = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (7)$$

$$EQMP_{ext} = \sqrt{\frac{\sum_{i=1}^{n_{ext}} (y_i - \hat{y}_{(i)})^2}{n_{ext}}} \quad (8)$$

La validation croisée par « leave – one - out » (LOO) (R Wehrens *et al.*, 2000) consiste à recalculer le modèle sur (n-1) observations, et à utiliser le modèle ainsi obtenu pour calculer la grandeur d'intérêt du composé écarté, notée $\hat{y}_{(i)}$. On répète le procédé pour chacune des grandeurs d'intérêt. La somme des carrés des erreurs de prédiction, désignée par le symbole PRESS (eq. (6)), est une mesure de la dispersion des estimations. On l'utilise pour définir le coefficient de prédiction (R Wehrens *et al.*, 2000):

$$Q_{LOO}^2 = \frac{SCT - PRESS}{SCT} \quad (9)$$

Contrairement à R^2 qui augmente avec le nombre de paramètres du modèle, le facteur Q_{LOO}^2 affiche une courbe avec maximum (ou avec palier) obtenu pour un certain nombre de descripteurs, puis décroît de façon monotone. Ce fait confère une grande importance au coefficient Q_{LOO}^2 . Une valeur de $Q_{LOO}^2 > 0,5$ est considérée comme satisfaisante, et une valeur supérieure à 0,9 est excellente (L Eriksson *et al.*, 2003)

Si de petites valeurs de Q_{LOO}^2 indiquent des modèles peu robustes, caractérisés par de faibles capacités prédictives internes, le contraire n'est pas nécessairement vrai. En fait, si une forte valeur de Q_{LOO}^2 est une condition nécessaire de robustesse et d'une possible capacité

prédictive élevée d'un modèle, cette condition seule n'est pas suffisante, et peut conduire à une surestimation de la capacité prédictive du modèle lorsqu'il est appliqué à des composés réellement externes.

Evidemment, on peut être amené à écarter 2, 3 ou un plus grand nombre d'éléments à la fois, ce qui conduit aux procédures LMO (leave – many- out). Cependant, ces procédures ne sont que rarement rapportées avec les résultats QSAR courants, et n'ont pas été pleinement exploitées dans le présent travail.

II- 5- 2- Régression en Composantes Principales (RCP) :

Dans la régression en composantes principales on soumet d'abord les variables à une ACP, puis l'analyse de régression est opérée sur les premières composantes principales en nombre limité. Lorsqu'on réalise une régression en composantes principales par, disons, sélection progressive alors l'équation résultante ne s'exprimera pas nécessairement en

fonction des composantes principales les plus basses. Ceci est dû au fait que l'ordre des composantes principales correspond à leur capacité à expliquer la variance des variables indépendantes, alors que l'analyse de régression concerne l'explication de la variable dépendante. En règle générale seules les composantes principales dont les valeurs propres sont inférieures à 1 seront insérées dans les régressions en composantes principales. Lorsqu'une valeur propre est inférieure à 1, alors une des variables originales de l'ensemble est mieux à même d'expliquer la variance que la composante principale. Néanmoins, et c'est souvent le cas à la limite, les 2 premières composantes conduisent à la meilleure corrélation avec la variable dépendante. Un autre fait à souligner en RCP est que, lorsqu'on incorpore de plus en plus de composantes principales, les coefficients des régresseurs déjà présents ne changent plus. Ceci est dû à l'orthogonalité des composantes principales, et parce que le rôle de chaque nouvelle composante principale est d'expliquer la variance non encore couverte.

II- 6- Calcule Des Descripteurs Moléculaire :

La représentation numérique de la structure chimique (descripteurs moléculaires) est une étape importante de l'investigation QSPR. Les performances du modèle élaboré et la précision des résultats sont étroitement liées au mode de détermination de ces descripteurs.

Nous avons utilisé le logiciel de modélisation moléculaire Hyperchem 6.03 (Hyperchem™ Release 6.03 for windows, 2000) pour représenter les molécules puis, à

l'aide de la méthode semi-empirique AM1, obtenir les géométries finales. Tous les calculs ont été menés dans le cadre du formalisme RHF sans interaction de configuration. Les structures moléculaires ont été optimisées à l'aide de l'algorithme Polak-Ribiere avec pour critère une racine du carré moyen du gradient égale à 0,01 kcal/mol (R Todeschini *et al.*, 2005). Les géométries ainsi optimisées ont été transférées dans le logiciel informatique DRAGON pour le calcul de plus de 1600 descripteurs (si l'on tient compte de ceux calculés à l'aide du logiciel Hyperchem) appartenant à 20 classes différentes.

En utilisant les options correspondantes du logiciel DRAGON, nous avons d'abord éliminé les descripteurs à valeurs constantes (écarts types inférieurs à 0,001) qui n'apportent aucune information, ensuite ceux qui sont hautement corrélés ($R > 0,9$) et qui véhiculent une information redondante. Pour chaque paire de descripteurs corrélés, est éliminé automatiquement celui qui présente les plus hautes corrélations croisées avec les autres descripteurs.

II- 7- Les types d'apprentissages d'un modèle QSPR :

Les techniques d'apprentissages ont été développées en informatique ces 15 dernières années. Il y'a deux types d'apprentissages :

- Non-supervisé : les activités et propriétés ne sont pas utilisés (seul les descripteurs), ce type utilisé pour grouper de composé chimique.
- Supervisé: établi une relation entre données entrées/input data (composés chimiques), et données produites output data (propriétés).

II- 7- 1- Sélection d'un sous-ensemble de variables par algorithme génétique (GA-VSS) :

Parmi les stratégies de sélection des variables explicatives on peut citer : les méthodes de pas à pas, ainsi que les algorithmes évolutionnaires et génétiques, la comparaison se fait souvent à l'avantage des algorithmes génétiques.

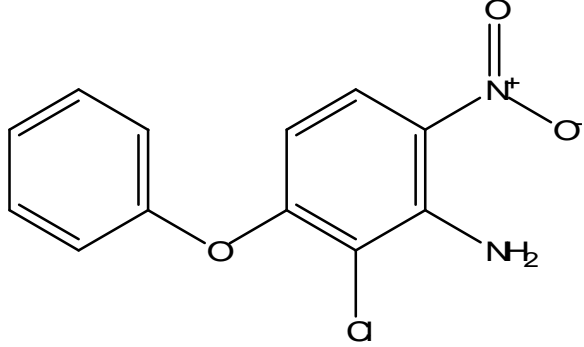
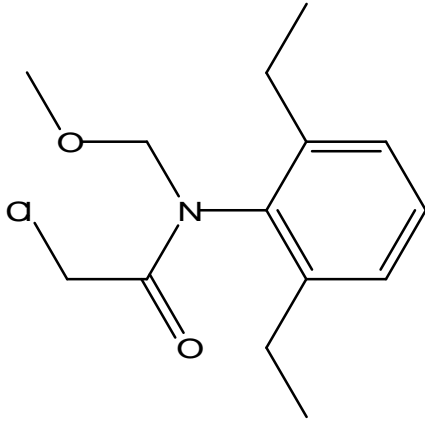
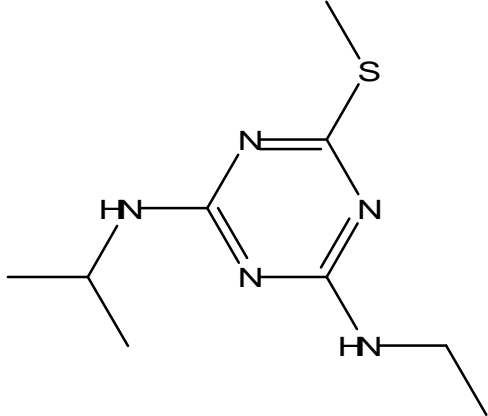
Dans un algorithme génétique adapté à l'optimisation, une solution potentielle est considérée comme un individu dans une population. La valeur de la fonction de coût associée à une solution mesure «l'adaptation» de l'individu associé à son environnement. Un algorithme génétique simule l'évolution, sur plusieurs générations, d'une population initiale 1mutation. Après un certain nombre de générations, la population est constituée d'individus bien adaptés, autrement dit des solutions supposées «bonnes» au problème d'optimisation.

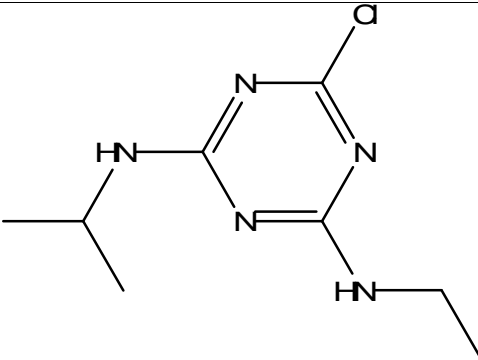
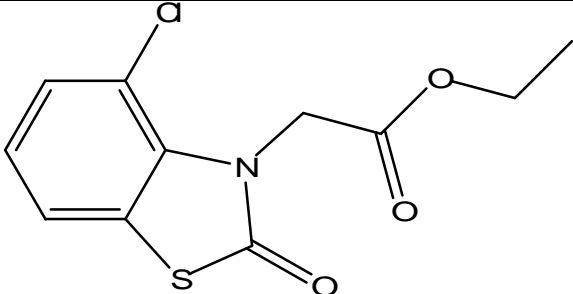
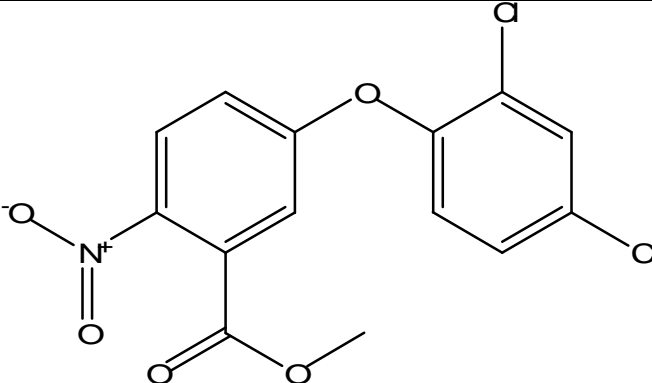
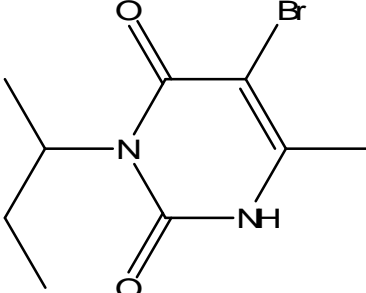
Dans le présent travail, la sélection des descripteurs a été réalisée par algorithme génétique, dans la version MOBY DIGS de Todeschini, en maximisant Q^2_{LOO} .

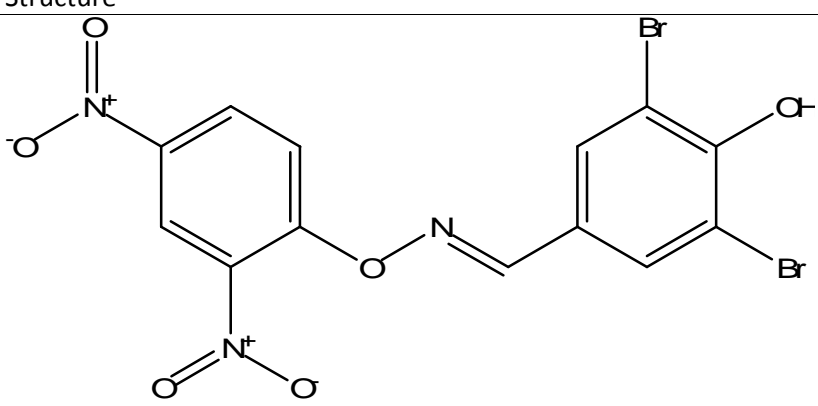
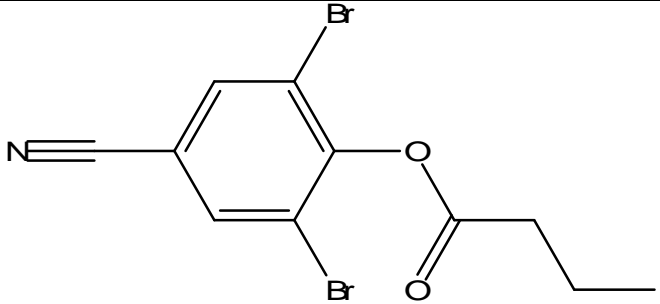
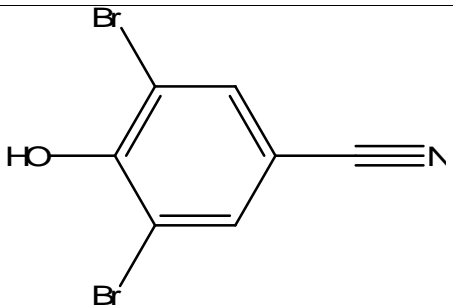
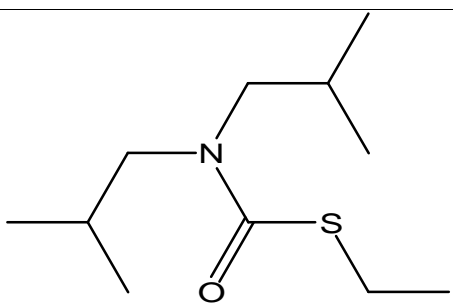
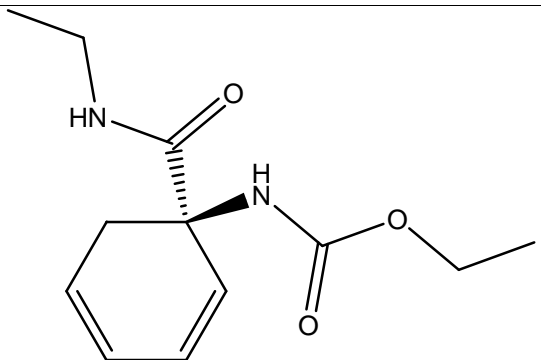
Dans cette partie expérimentale, nous allons déterminer la solubilité d'une série de pesticides par modélisation d'un ensemble d'estimation. Le Meilleur modèle obtenue sera validé par l'approche hybride algorithme génétique et régression multilinéaire GA/RLM.

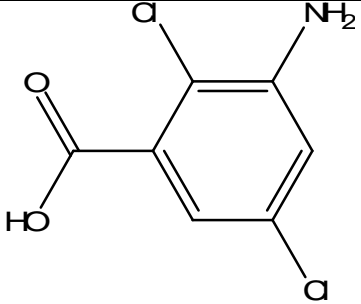
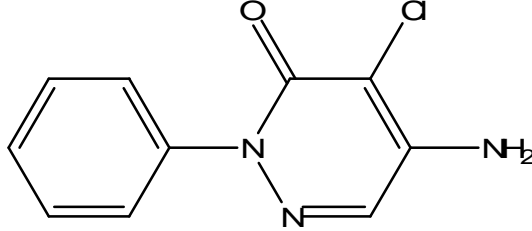
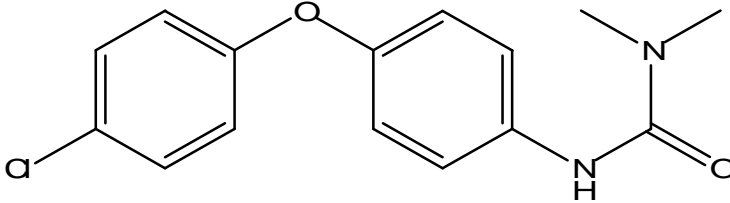
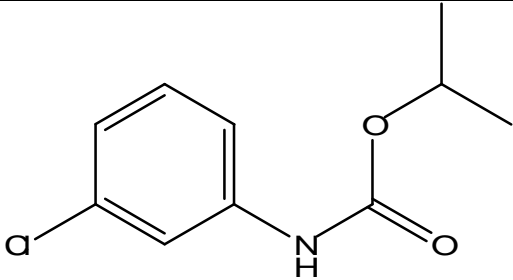
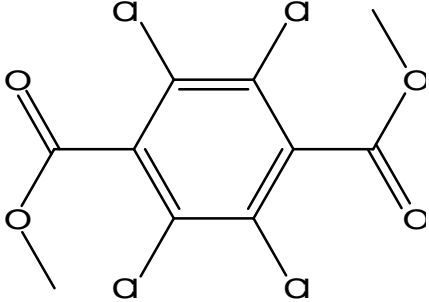
La liste des molécules étudiées dans notre travail présente dans le Tableau II, cette base des données contient 26 composé, avec la propriété à étudier.

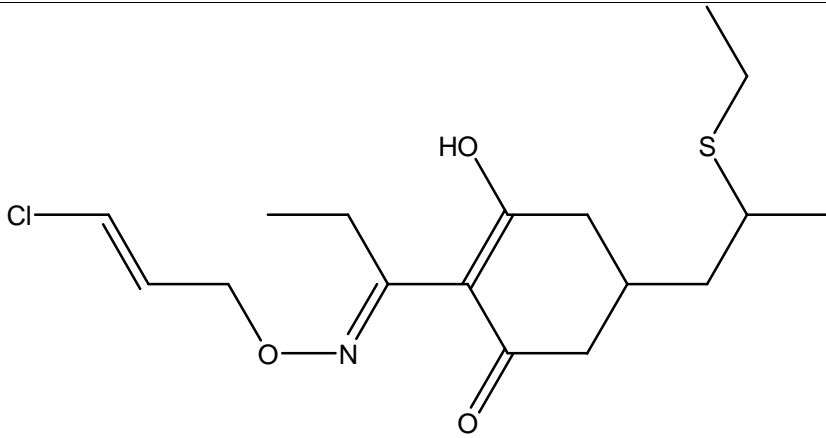
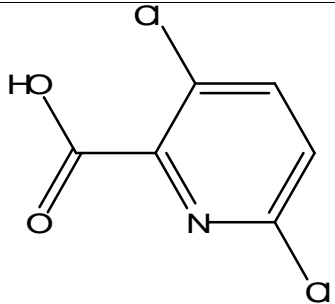
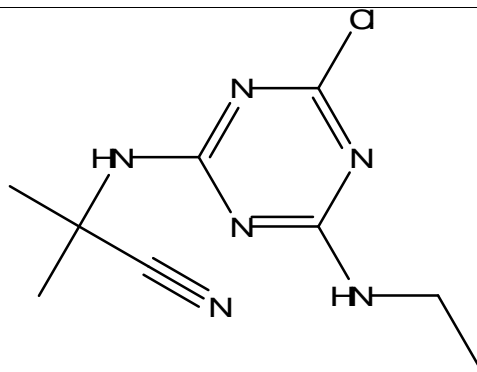
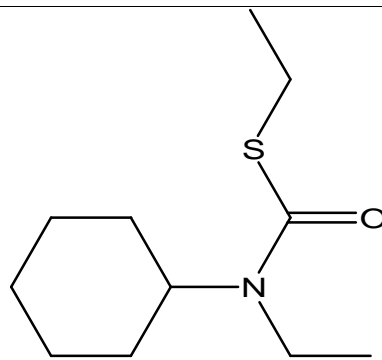
Tableau II : les composés étudiés et leurs solubilités

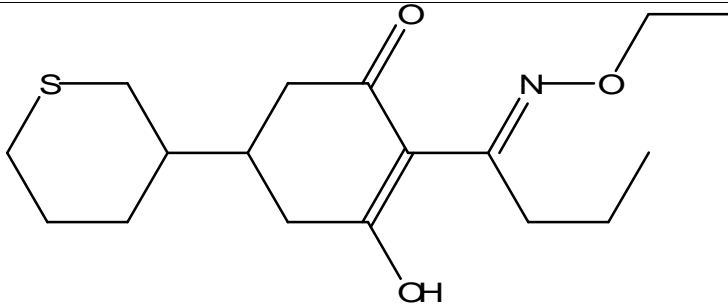
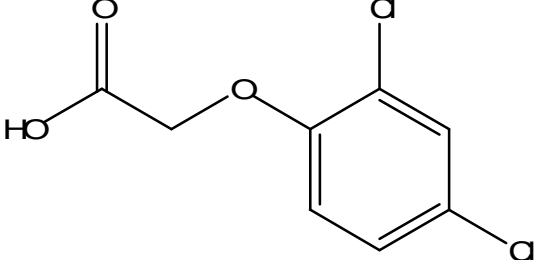
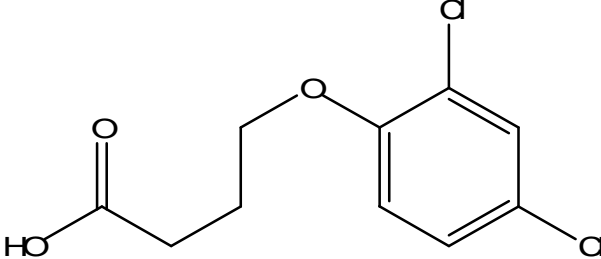
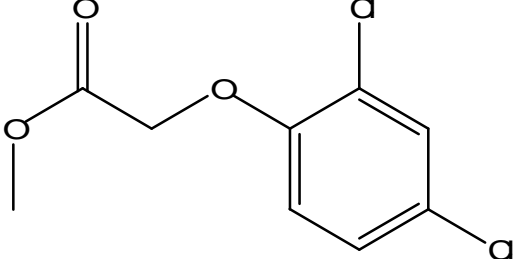
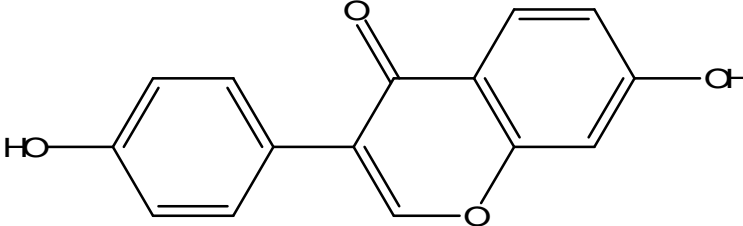
N°	Composé	Structure	Log S
1	Aclonifen		2.38202
2	Alachlor		2.26717
3	Ametryn		1.51851

N°	Composé	Structure	Log S
4	Atrazine		0.14613
5	Benazolin*		2.69897
6	Bifenox		-0.4559
7	Bromacil*		2.8451

N°	Composé	Structure	Log S
8	Bromofenoxim		0.41497
9	bromoxynil butyrate		1.43136
10	Bromoxynil*		2.11394
11	Butylate		1.64345
12	Carbetamide		3.54407

N°	Composé	Structure	Log S
13	Chloramben		2.8451
14	Chloridazon*		2.53148
15	Chloroxuron		0.39794
16	Chlorpropham		1.94939
17	Chlorthal dimethyl		-0.301

N°	Composé	Structure	Log S
18	Clethodim		3.73239
19	Clopyralid		5.15534
20	Cyanazine*		2.233
21	Cycloate*		1.97772

N°	Composé	Structure	Log S
22	Cycloxydim*		1.60206
23	2,4-D		2.94939
24	2,4-DB		1.66276
25	2,4-D-methyl		2
26	Qq		2.17609

*Composés de validation

I- Calcul du modèle :

I- 1- Choix du modèle :

Les molécules sont dessinées par le logiciel ChemDraw (ChemDraw ultra 7.0) puis elles sont optimisées en utilisant le logiciel HyperChem (Hyperchem™ Release 6.03 for Windows, 2000). Les descripteurs moléculaires ont été calculés à l'aide du logiciel informatique Dragon (R Todeschini *et al.*, 2005), plus de 1600 descripteurs sont calculés.

L'ensemble des données a été décomposé en deux sous ensembles aléatoirement, 70% de la totalité des composés pour la construction du modèle et 30% pour la validation externe.

D'après l'algorithme génétique dans la version MobyDigs (B Ballabio *et al.*, 2004), plusieurs modèles sont obtenus pour chaque jeu de groupe; le choix a été opté pour le modèle qui conduit aux meilleurs statistiques des 100 modèles générés par algorithmes génétiques. Sa validation est effectuée par 7 composés, alors que sa calibration par 19.

I- 2- Choix d'une famille de descripteurs :

Les descripteurs entrant dans le modèle obtenus par algorithme génétique, leurs classes et de brèves définitions sont donnés dans le tableau III

Tableau III : Descripteurs moléculaires intervenant dans la modélisation de la solubilité (log S).

N°	Descripteur	Classe	Signification
1	MATS6v	List of 2D autocorrelation indices	Moran autocorrelation - lag 6 / weighted by atomic van der Waals volumes
2	BELp3	Burden eigenvalue descriptors	lowest eigenvalue n. 3 of Burden matrix / weighted by atomic polarizabilities
3	Mor26p	3D-MoRSE descriptors	3D-MoRSE - signal 26 / weighted by atomic polarizabilities

Les valeurs du logarithme de la solubilité et les valeurs des descripteurs sélectionnés par algorithme génétiques sont présentées dans le tableau IV au dessous:

Tableau IV: Valeurs de log S, MATS6V, BELP3, Mor26p

N°	Composé	Log s	MATS6v	BELp3	Mor26p
1	Benazolin*	2.69897	0.18	1.503	-0.219
2	Bromacil*	2.8451	-0.331	1.327	-0.008
3	Bromoxynil*	2.11394	-0.429	1.335	-0.329
4	Chloridazon*	2.53148	-0.225	1.209	-0.282
5	Cyanazine*	2.233	-0.414	1.55	-0.005
6	Cycloate*	1.97772	-0.195	1.508	-0.087
7	Cycloxydim*	1.60206	-0.081	1.65	-0.005

N°	Composé	Log s	MATS6v	BELp3	Mor26p
8	Acclonifen	0.1461	-0.116	1.536	-0.4
9	Alachlor	2.38202	-0.217	1.56	-0.083
10	Ametryn	2.26717	-0.37	1.547	-0.013
11	Atrazine	1.51851	-0.495	1.542	-0.07
12	Bifenox	-0.4559	-0.098	1.565	-0.38
13	Bromofenoxim	0.41497	0.032	1.576	-0.333
14	Bromoxynil butyrate	1.43136	-0.429	1.335	-0.329
15	Butylate	1.64345	-0.096	1.531	-0.147
16	Carbetamide	3.54407	-0.16	1.647	0.033
17	Chloramben	2.8451	0.594	1.116	-0.381
18	Chloroxuron	0.39794	-0.285	1.701	-0.377
19	Chlorpropham	1.94939	0.044	1.375	-0.217
20	Chlorthal dimethyl	-0.301	-0.784	1.404	-0.467
21	Clethodim	3.73239	0.066	1.609	0.076
22	Clopyralid	5.15534	0.492	0.78	-0.234
23	2,4-D	2.94939	-0.039	0.989	-0.326
24	2,4-DB	1.66276	-0.01	1.343	-0.341
25	2,4-D-methy	2	-0.024	1.271	-0.293
26	Qq	2.17609	-0.097	1.496	-0.364

*Composés de validation

II- Développement du modèle :

II- 1- Matrice de corrélation :

La matrice de corrélation entre la variable dépendante et les variables explicatives est présentée dans le tableau V

Tableau V: Matrice de corrélation

	Log S	MATS6v	BELp3
MATS6v	0.565 0.012		
BELp3	-0.512 0.025	-0.533 0.019	
Mor26p	0.562 0.012	-0.022 0.927	0.293 0.224

Les trois descripteurs MATS6v, BELp3, Mor26p (TableauV) sont corrélés avec le logarithme de la solubilité, donc les trois descripteurs porte des informations complémentaires au log S. le Mor26p il n'est pas corrélé avec les deux autres descripteurs, le BELp3 et le MATS6v sont corrélés mais la suspension d'un descripteur d'entre eux entraine une dégradation des capacités prédictives du modèle, ils portent des informations importantes au log S.

II- 2- Equation et analyse de régression :

L'équation de régression du modèle calculé est la suivante :

$$\log S = 8.63(\pm 1.085) + 1.24(\pm 0.5190)\text{MATS6v} - 3.55 (\pm 0.7156)\text{BELp3} + 6.56(\pm 0.8863)\text{Mor26p} \quad (10)$$

Tableau VI: Analyse de régression

Régresseur	Coef	SE Coef	T	P	VIF
Constante	8.632	1.085	7.96	0.000	
MATS6V	1.2398	0.5190	2.39	0.030	1.435
BELP3	-3.5510	0.7156	-4.96	0.000	1.569
MOR26	6.5592	0.8863	7.40	0.000	1.124

Les valeurs de T du descripteur Mor26p est importante se qui confirme que la grande contribution de se dernier, les deux autres descripteurs ont une contribution proche et de même pourcentage presque.

Les valeurs des VIF inférieure à 5, ce qui traduit l'indépendance des trois descripteurs intervenant dans le modèle.

Les paramètres statistiques sont présentés dans le tableau suivant

Tableau VII: Les paramètres statistiques pour l'ensemble de calibration (19 composés) et celui de validation (7 composés),

Les paramètres statistiques	
F	32.59
R ²	86.7
R ² _{adj}	84.04
Q ² _{Loo}	80.86
Q ² _{ext}	77.37

Les paramètres statistiques	
S	0.577
EQMC	0.512
EQMP	0.615
EQMP _{ext}	0.643

Les paramètres statistiques reproduit dans le tableau ci-dessus montrent une bonne explication de la variabilité log S, par les descripteurs choisis, de l'ordre de 81 pour le coefficient de détermination, et une robustesse du modèle due à la valeur de Q_{LOO}^2 , alors que les valeurs de l'erreur quadratique moyenne de prédiction et de calcul sont petits et proches; en outre ce modèle est significatif avec une valeur du paramètre de Fisher: (F=35,57), Les statistiques montrent aussi que le modèle à une très grande capacité prédictive externe, ceci traduit par la grande valeur de Q_{ext}^2 et la faible valeur de EQMP_{ext}.

II- 3- Diagramme de Williams

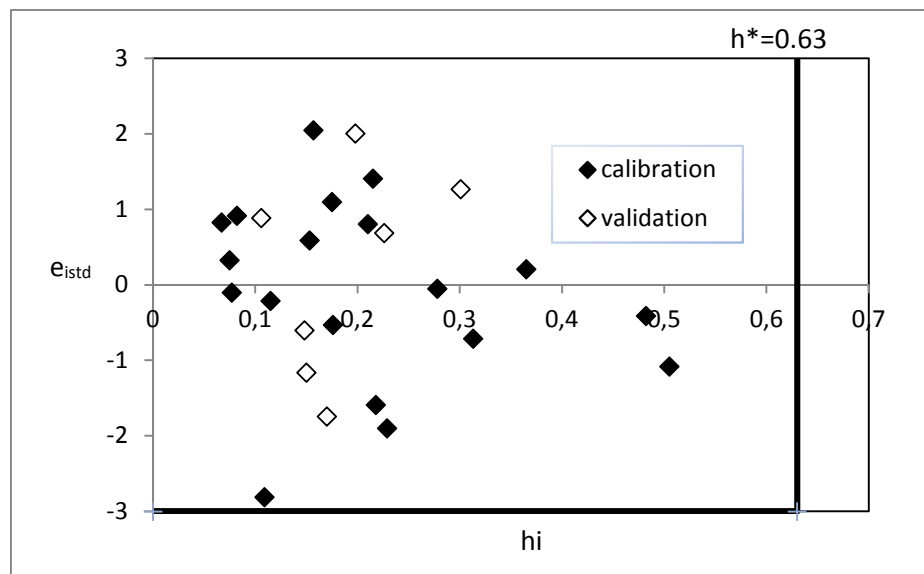


Figure 3 : Diagramme de Williams

D'après le graphe on remarque que tous les erreurs standardisé sont comprise entre les limites ± 3 , donc une absence des points aberrant, on remarque aussi une absence des points influents c'est-à-dire tous les points ont une valeur de levier inférieur à la valeur critique $h^*=0.63$.

II- 4- Qualité de l'ajustement

La qualité de l'ajustement a été vérifiée par le graphe des valeurs calculées de l'ensemble de calibration et les valeurs prédites de l'ensemble de validation en fonction de celles expérimentales du logarithme de la solubilité.

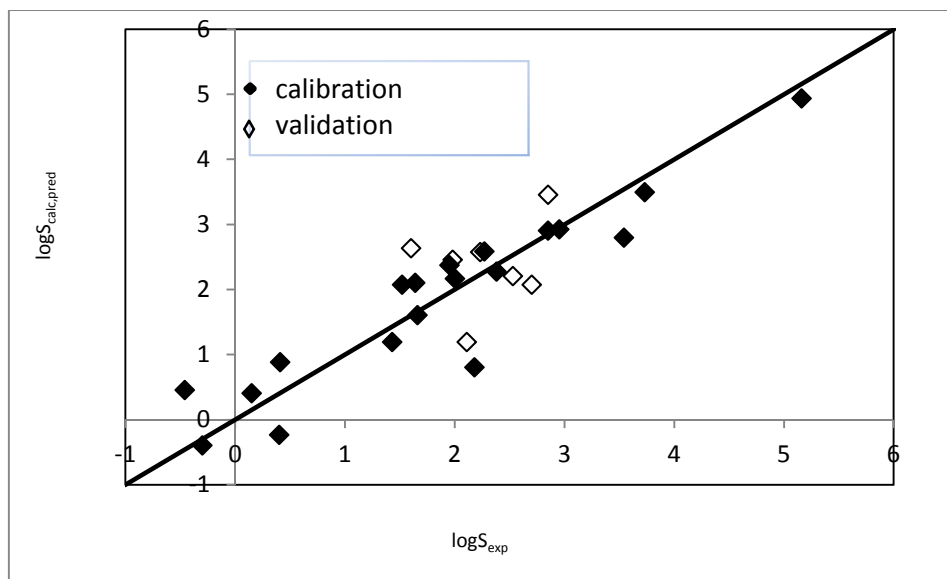


Figure 4 : Qualité de l'ajustement

D'après le graphe on remarque une certaine dispersion autour de la première bissectrice des deux ensembles de calibration et de validation. Ceci est confirmé par la valeur de Q^2_{LOO} , qu'elle n'est pas trop élevée.

II- 5- Test de randomisation

Dans le but de montrer que le modèle obtenu n'est pas dû au hasard nous avons appliqué le test de randomisation de y (Figure 05), dans cette figure on a présenté le graphe des coefficients statistiques Q^2 et R^2 .

Il est clair que les statistiques obtenues pour les log S modifiés sont plus petites que celles du modèle réel, les cercles vides regroupés dans la région des valeurs négatives de Q^2 ont des valeurs petites de R^2 , seul le cercle noirci a des valeurs élevés et proches pour ces deux statistiques, il représente notre modèle calculé, donc il n'est pas dû au hasard.

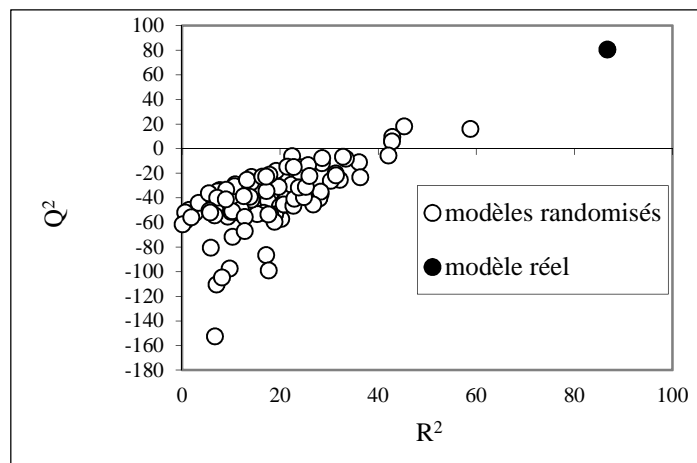


Figure 5: Test de randomisation

Conclusion générale

Dans notre travail nous avons utilisé la méthodologie QSPR pour prédire la solubilité d'une série de 26 produits phytopharmaceutiques pour cela nous avons relié cette propriété à des descripteurs moléculaires théoriques,

Pour cette étude nous avons appliqué la méthode de régression linéaire multiple, les descripteurs intervenant dans le modèle sont choisis par algorithmes génétiques,

Le modèle choisi conduit aux bons résultats: qualité de l'ajustement, robustesse interne et externe, capacité prédictive... ce qui prouve que la relation entre la solubilité et les variables explicatives (MATS6v, BELP3, Mor26p) elle est fondamentalement linéaire.

Références bibliographiques

- A. Crum Brown, T.R. Fraser.1868. On the connection between chemical constitution and physiological action. Part 1. On the physiological action of salts of the ammonium bases, derived from strychnia, brucia, the bia, codeia, morphia and nicotia. Trans. Roy. Soc. Edinburgh, 25, 151-203.
- A. J . Myles , R.N. Feudale , Y. Liu, N.A. Woody, S.D. Brown.2004. An introduction to decision tree modeling, J. Chemom. 18, 275-285.
- A. Mokhtari Nahida . 2012 . Identification et dosage des Pesticides dans l'Agriculture et les problèmes d'Environnement liés,Laboratoire de syntèse organique appliquee (LSOA).
- A. R. Katritzky, M. Kuanar, S. Slavov, and C. D. Hall. 2010. Quantitative Correlation of Physical and Chemical Properties with Chemical Structure: Utility for Prediction, Chem. Rev. 110, 5714–5789.
- A.Hayo,M.G. van der werf, J. 1996.Ecosystems and Environment 60 81-96.
- A.Tor , M.Emin Aydin, S.Ozcan.2006. Analytica Chimica Acta 559. 173–18.
- B. Ballabio, R. Todeschini, V. Consonni, A. M. Pavan.2004.Mobydigs Software for Multilinear Regression Analysis and Variable Subest Selection by Genetic Algorithm. Release1.0 for Regression Analysis and Variable Subest Selection by Genetic Algorithm. Release1.0 for Windows, Milano.
- B. C . López, Á. S. Gómez, G. M .Rey, G .B .Cancho, G. J .Simal. 2005. Determination of carbamates and organophosphorus pesticides by SDME–GC in natural water, Analytical and bioanalytical chemistry. vol 383 -4-: 557-561.
- B.Zeboudji .2005 mémoire de magister Ecole Nationale Polytechnique Alger.
- C.G.B. Margoum.2010 Thèse de doctorat Reims Champagne-Ardenne.
Centre anti poison Alger .2011.
- D. Devault. Approche spatio-temporelle de la contamination par les herbicides.
- E. Barriusso, C. Bedos, P. Benoit, M.P. Charnay Y. Coquet.
- E. Capkin Altioik I.,Karahhan S.2006. Water quality and fish size affect toxicity of endosulfan, an organochlorine pesticide, to rainbow trout. Chemosphere vol 64:1793–1800.
- E. M .Arias, P. E.Lopez, C. E.Martinez, G. J. Simal, J. C. Merut, R. L. Garcia .2008.The

Références bibliographiques

- mobility and degradation of pesticides in soils and the pollution of groundwater resources. *Agriculture, Ecosystems and Environment* vol 123:247-260.
- Enquête SCEES.2001. sur blé, 85% en moyenne des agriculteurs déclarent enregistrer leurs pratiques phytosanitaires (95% dans les régions de grande culture).
- Faostat <http://faostat.fao.org/site/423/default.aspx#ancor>.
- http://www.nord-nature.org/environnement/pollutions/pesticides_pasteur.pdf.
- HyperchemTM Release 6.03 for windows.2000. Molecular Modeling System.
- Institut Pasteur de Lille, 1999. Produits phytosanitaires dans les eaux de pluie de la Région Nord – Pas-de-Calais.
- J .Fournier. 1988. Chimie des pesticides. Cultures et Techniques, Agence de Coopération, Culturelle et Technique.
- J. Boland, I.Koonen, J. Van Lidth, D.E. Jeude, J.Oudejans .2004. Les pesticides composition, utilisation et risques. Editions Agrodok .
- J. Ghasemi, S. Saaidpour, S.D. Brown.2007.QSPR study for estimation of acidity constants of some aromatic acids derivatives using multiple linear regression (MLR) analysis. *J. Mol. Struct. (Theochem)*, 805, 27-32.
- J.Gasteiger, J.Zupan, A.F. Duprat, T. Huynh, G. Dreyfus.1993. Toward a principled methodology for neural network design and performance evaluation in QSAR. Application to the prediction of Log P. *J. Chem. Inf. Comput. Sci.* 1998, 38, 586-594- J. Gasteiger, J. Zupan, Neural Networks in Chemistry, *Angew. Chem. Int. Ed. Engl*, 32, 503-527.
- J.N Aubertot, J.M .2005.Barbier Expertise scientifique collective INRA – Cemagref, Alain Carpentier.
- K.El.Mrabet.2006.thèse de doctorat , Paris.
- L. Eriksson, J. Jaworska, A. P. Worth, D. Cronin, R. M. Mc Dowell, P.2003. Gramatica Methods for Reliability and Uncertainty Assesement and for Applicability Evaluation of Classification and Regression Based QSARs. *Environmental Health Perspectives* ,111, 1361-1375.
- Les Produits phytosanitaires dans l'air .2007.origine, surveillance recommandations et pratiques en agriculture , corpen .
- Loi sur l'Air – Convention18/99. 2000. http://www.memoireonline.com/11/12/6459/m_Etude-sur-les-pesticides.html ineris drc-00–23449– aire –569a-CDu-fmrLaboratoire Central de Surveillanc de la Qualité de l'Air; Unité Qualité de l'air – Direction des Risques

Références bibliographiques

Chroniques; Loi sur l'Air – Convention 18/99.

Loi algérienne N° 87-17 relative à la protection phytosanitaire. 1987. Journal officiel de la république algérienne N°32 du 05-08.

M.O.S.Ould Kankou .2004.thèse de doctorat Limoges .

N. Margossian. 2007.Le règlement Reach - La réglementation européenne sur les produits chimiques, Dunod - L'usine Nouvelle, Paris.

ORP. 2008.Observatoire des Résidus de Pesticides, site internet consulté en septembre : <http://www.observatoire-pesticides.gouv.fr>

P. Geladi, B.R. Kowalski.1986.Partial Least Squares Regression: a Tutorial, Anal. Chim. Acta 185, 1-17.

R. Calvet.2005.Les pesticides dans le sol. edition France Agricole .

R. Leardi.2001. Genetic algorithms in chemometrics and chemistry: a review J. Chemometr, 15, 559-569.

R. Todeschini, V. Consonni, M. Pavan.2005. DRAGON, Software for the Calculation of Molecular Descriptors. Release 5.3 for windows, Milano.

R. Wehrens, H. Putter, L. M. C. Buvidens. 2000.Chemom. Int. Lab. Syst (p,35-52,54).

S.Sebih .2008. memoire de magister Ecole Nationale Polytechnique Alger .

Union des Industries de la Protection des Plantes <http://www.uipp.org/Services-pro/Chiffres-cles/Reperes-monde-et-Europe>.

Références bibliographiques