



MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR
ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITÉ «ABBÈS LAGHROUR» DE KHENCHÉLA
FACULTÉ DES SCIENCES ET DE LA 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)

Spécialité : Chimie analytique .

Option : Chimie analytique et environnement.

Modélisation Du coefficient de partage octanol/eau d'une série de quelques composés polluants

*Réalisé par : TRAD ASMA
HOGGAS WISSAME*

Dirigé par : Mm BOUAKKADIA Amel

Membres de jury :
Kertiou Noureddine MAA
HAKKAR Farida MAA

Présenté le 01/07/2017

Année universitaire

2016-2017

Remerciements :

Avant tout, nous remercions

Allah, pour son guide et sa protection, et d'avoir nous donner l'envie et la force pour mener à terme ce travail.

Au début de cette mémoire, nous tenons tout particulièrement à exprimer nos sincères remerciements à l'enseignante encadreur

Mlle **Bouakkadia Amel** pour son encadrement et sa gentillesse, ainsi que pour son aide et ses précieux conseils lors de la rédaction de ce mémoire. Sa patience et sa persévérance dans le suivi. Nous lui exprimons notre sincère reconnaissance.

Nous remercions par ailleurs l'ensemble des membres de jurés Qui nous ont honoré par leurs assistance pour juger notre modeste travail.

Nos remerciements s'étendent également à tous les professeurs et docteurs qui nous ont enseigné durant les années des études.

Nous sincères remerciement à tout enseignants, chercheurs et personnels de l'équipe du département de SMdul'Université **Abesse Laghrour Khanchela** qui par leurs gentillesse, leurs modestie, leurs aide, leurs conseils et leurs amitié et les bons moments passés ensemble, je cite parmi eux : **Mr Bekha, Mr Kertiou, Mr Badis, Mm Golea, Mlle Amel ...**

Merci à tous ceux qui ont contribué de près ou de loin à la réalisation de ce travail, pour leurs conseils, leurs encouragements et leurs soutient.

Dédicace :

Toutes les lettres ne sauraient trouver les mots qu'il faut...

Tous les mots ne sauraient exprimer la gratitude, l'amour, le respect, la reconnaissance...

Ainsi, c'est tout simplement que je dédie ce projet de fin d'étude :

A mes chers parents :

Dont leurs mérites, leurs sacrifices, leurs qualités humaines m'ont permis de vivre ce jour. Les mots me manquent pour exprimer toute la reconnaissance, la fierté et le profond amour que je vous porte pour les sacrifices qu'ils ont consenti pour ma réussite, qu'ils trouvent ici le témoignage de mon attachement ma reconnaissance, gratitude et respect, que dieu leur préservent bonne santé et longue vie. Tous mes sentiments pour vous.

A mes adorables sœurs et mes chers frères :

Merci d'être toujours à mes côtés.

A mes chers grands parents :

Que ce modeste travail, soit l'expression des vœux que vous n'avez cessé de formuler dans vos prières. Que dieu vous préserve santé et longue vie.

A tous mes amis :

Que dieu nous garde si tendres et aimants les un envers les autres .

Wissem

Dédicace :

Dieu tout puissant merci d'être toujours au près de moi.

Je dédie ce travail aux être les plus chers à mon cœur :

Mes chers parents : Hammadiet Hnia

Autant des phrases et d'expressions aussi éloquentes soient-elles ne sauraient exprimer ma gratitude et ma reconnaissance. Vous avez su m'inculquer le sens de la responsabilité, de l'optimisme et de la confiance en soi face aux difficultés de la vie. Vos conseils ont toujours guidé mes pas vers la réussite. Votre patience sans fin, votre compréhension et votre encouragement sont pour moi le soutien indispensable que vous avez toujours su m'apporter. Je vous dois ce que je suis aujourd'hui et ce que je serai demain et je ferai toujours de mon mieux pour rester votre fierté et ne jamais vous décevoir. Que dieu, le tout puissant, vous préserve, vous accorde santé, bonheur, qui étudie de l'esprit et vous protège de tout mal.

A ma sœur :Sara

J'espère attient le seuil de tes espérances. Que ce travail soit l'expression de ma profonde affection je te remercie pour le soutien moral et l'encouragement que tu m'as accordés. Je te souhaite tout le bonheur que tu mérites.

A mes frères :

Merci d'être toujours à mes côtés. Par votre présence, par votre amour dévoué et votre tendresse, pour donner du goût et du sens à ma vie. Je prie dieu, le tout puissant, pour qu'il vous donne bonheur et prospérité.

A ma grand-mère : Yamina

A toute ma famille et surtout ma tante Rokia et ses enfants Ilyas, Israe et Yasser.

A ma sœur Jalila et ses enfants Mehdi et Ayoub.

A mes chers cousines : Maissa, Fatiha, Houda et Kaouthar.

A tous mes chers collègues et amis :

A tout ceux qui ont su m'apporter aide et soutien aux moments propices, je dédie ce travail, reconnaissant et remerciant chaleureusement.

Asma

Sommaire

Introduction générale	1
Etude bibliographique	
Chapitre I	
I- Généralités sur les pesticides :	2
I- 1- Définition des pesticides :	2
I- 2- Historique des pesticides :	2
I- 3- Les classes des pesticides :	3
I- 3- 4- Les insecticides organophosphorés :	4
I- 4- L'effet des pesticides sur l'environnement :	5
I- 4- 1- L'impact sur l'air :	6
I- 4- 2- L'impact sur le sol :	7
I- 4- 3- L'impact sur l'eau :	8
I- 4- 4- L'effet des pesticides sur les villes :	8
I- 4- 5- L'impact sur Les maisons :	9
I- 5- Toxicologie des pesticides :	9
I- 6- Le marché des pesticides :	10
Chapitre II	
II- Phénomène de partage :	12
II- 1- Définition :	12
II- 2- Historique de la constante de partage :	12
II- 3- Définition de coefficient de partage Octanol-Eau (Kow) :	13
II- 4- les paramètres influençant sur la constante de partage :	14
Chapitre III	
III- Les modèles QSAR/QSPR :	16
III- 1- Définition de QSPR :	16
III- 2- Principe de QSPR :	17
III- 3- Les méthodes Mathématiques Utilisés par le modèle QSPR :	17
III- 3- 1- Régression linéaire multiple (MLR) :	17

III-	3- 2- Régression en composantes principales (RCP) :.....	20
III-	3- 3- régression moindre carré (PLS) :	21
III-	4- Calcule des descripteurs moléculaires :.....	21
III-	5- Les types d'apprentissages d'un modèle QSPR :	22
III-	6- Sélection d'un sous-ensemble de variables par algorithme génétique (GA-VSS) :	22
Application		
I-	Calcul du modèle :.....	28
I-	1- Choix du modèle :.....	28
I-	2- Choix d'une famille de descripteurs :.....	28
II-	Développement du modèle :	29
II-	1- Matrice de corrélation :	29
II-	2- Equation et analyse de régression :.....	30
II-	3- Diagramme de Williams.....	33
II-	4- Qualité de l'ajustement :.....	33
II-	5- Test de randomisation :	34
Conclusion générale.....		36
Références bibliographiques		37

Liste des figures

Figure 1 : Schéma montre le voyage des pesticides dans l'environnement	6
Figure 2: Caractérisation expérimentale de la dégradation des pesticides dans les sols	8
Figure 3: Diagramme de Williams.....	34
Figure 4: Graphe des valeurs calculées, prédites en fonction des valeurs observées du log Kow.	35
Figure 5: Test de randomisation.....	36

Liste des tableaux

Tableau I : Structure des composés étudiés.....	24
Tableau II: Descripteurs moléculaires intervenant dans la modélisation de la constante de partage octanol -eau (log kow).	29
Tableau III: Matrice de corrélation	30
Tableau IV: Analyse de régression	31
Tableau V: Les paramètres statistiques pour l'ensemble de calibration (33 composés) et celui de validation (10 composés),.....	32

Symboles et abréviations

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_{istd} :	Résidu de prédiction standardisé.
F	Statistique de Fisher.
H	Matrice de projection, ou matrice chapeau.
h_{ii}	Eléments diagonaux de la matrice chapeau.
Log P :	Coefficient de partage octanol /eau.
LOO	Validation croisée par omission d'une observation.
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/ Proprieties - 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 prediction.
X	Matrice des valeurs observées des variables explicatives.
X'	Matrice transposée de X
VIF:	Facteur d'inflation de la variance.
AG:	Algorithme génétique
LMO:	Validation croisée par omission d'un ensemble d'observations.
OCDE	Organisation de Coopération et de Développement Economiques
QSPR :	Quantitative Structure/ PropertyRelationships.
RLM:	Régression linéaire multiple.
ACP	d'analyse en composantes principales
ϵ_{LUMO}	la plus basse orbitale moléculaire inoccupée
ϵ_{HOMO}	l'énergie de la plus haute orbitale moléculaire occupée

Introduction générale

Depuis les années 1950, le développement de l'agriculture et la volonté d'augmenter les rendements ont conduit à une utilisation croissante des pesticides, cette utilisation s'est fortement développée durant les dernières décennies, cette évolution et suite à une utilisation massives des énergies pour augmenter la production agricole ont causé une contamination de l'environnement et en particulier une pollution diffuse d'un grand nombre d'aquifères sur l'ensemble du territoire. Cependant, les pesticides montrent leurs limites avec stagnation des rendements et l'apparition de résistances de plus en plus nombreuses et précoces aux nouvelles molécules de synthèse. L'Algérie utilise environ 6000 à 10000 tonnes de pesticides par an [1], on est ainsi passé de l'usage occasionnel à l'usage permanent, du curatif au préventif. Si l'intérêt s'est porté en premier temps sur la contamination du réseau de surface en raison de forte concentration détectée précocement.

Les sources ponctuelles ont fait l'objet de nombreux travaux et recherches sur le terrain et en laboratoire, par contre, les sources diffusées et particulièrement la contamination par les pesticides (herbicides, fongicides, insecticides, ...) en zones agricoles n'ont attiré l'attention des scientifiques et gouvernement que depuis la fin des années 1970 où des analyses d'eau souterraine, de surface et de drainage révélaient la présence des herbicides et d'autres produits phytosanitaires agricoles.

Ainsi, l'objectif principal de ce travail est d'étudier la prédiction du coefficient de partage octanol/eau des pesticides en utilisant les méthodes QSPR.

Les insecticides étudiés dans ce travail sont des organophosphorés. Ils sont fortement utilisés dans l'agriculture en Algérie[2].

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 consacré aux différentes méthodes de traitement QSPR.
- La deuxième partie présente la partie Application .

Le mémoire est enfin complété par une conclusion générale récapitulant les résultats obtenus.

I- Généralités sur les pesticides :

I- 1- Définition des pesticides :

Les pesticides sont, en terme générique utilisé pour désigner toutes les substances naturelles ou des synthèses capables de contrôler, d'attirer, de repousser, de détruire ou de s'opposer au développement des organismes vivants (microbes, animaux ou végétaux) considérés comme indésirables pour l'agriculture, l'hygiène public (par exemple les cafards dans les habitations), la santé publique (les insectes parasites (poux, puces)) ou vecteurs de maladies telles que le paludisme et les bactéries pathogènes de l'eau détruites par la chloration, la santé vétérinaire, ou les surfaces non-agricoles (routes, aéroports, voies ferrés, réseaux électriques...), ils sont l'un des rares substances qui sont à la fois toxiques et délibérément rejetés dans l'environnement. Les pesticides sont présents dans presque tous les milieux de l'environnement y compris les eaux de surface, les eaux souterraines, l'air ambiant, la poussière, le sol, le brouillard, la pluie, et la glace.

Les pesticides sont diffusés partout, car les résidus de pesticides ont été trouvés dans plus de 70%des fruits et légumes, plus de 60% des échantillons de blé et de 99% de lait, ils sont également présents dans les oranges de presque tous les adultes et les enfants.

Le terme «pesticide» couvre un champ plus vaste et général que les expressions «produit phytosanitaire» ou «produit phytopharmaceutique» car il englobe tous les produits destinés à lutter contre tous les dits nuisibles, ou indésirables (ex : les champignons qui pourraient attaquer une charpente) et les médicaments vétérinaires destinés à protéger les animaux domestiques, gibiers ou de compagnie (par exemple, le collier antipuces pour les chiens)[3].

I- 2- Historique des pesticides :

Les pesticides ont été reconnus depuis longtemps :

Etude bibliographique/chapitre I

Ø Dès avant 2500 BCE, les humains ont utilisé des pesticides pour protéger leur récoltes. Le premier pesticide utilisé est par l'époussetage du soufre élémentaire utilisé dans la Sumeria environ 4500 ans.

Ø Par le 15^{ème} siècle, les produits chimiques toxiques comme l'arsenic, le mercure et le plomb ont été appliquées à des cultures pour tuer les parasites.

Ø Au 17^{ème} siècle, le sulfate de nicotine a été extrait de feuilles de tabac pour l'utilisation d'un insecticide.

Ø Le 19^{ème} siècle a vu l'introduction de deux autres pesticides naturels, pyrèthre, qui est dérivé de chrysanthèmes, la roténone et qui est dérivé de la racine des légumes tropicaux.

Ø En 1939, Paul Müller à découvert que le DDT est un insecticides très efficace. Il est rapidement devenu le plus largement utilisé des pesticides dans le monde.

Ø Dans les années 1940, les fabricants ont commencé à produire de grandes quantités de pesticides de synthèse et leur utilisation s'est généralisée.

Ø Certaines sources estiment les années 1940 et 1950 pour le début de l'ère des pesticides.

Ø L'usage des pesticides a augmenté de 50 fois depuis 1950 et 2,3 millions de tonnes (2,5 millions de tonnes impériales) de pesticides industriels sont maintenant utilisés chaque année.

Ø 65% de tous les pesticides dans le monde sont utilisés les pays développés, mais l'utilisation dans les pays en développement est de plus en plus élevée[3].

I- 3- Les classes des pesticides :

Il existe 3 grandes familles de pesticides :

Etude bibliographique/chapitre I

I- 3- 1-Les herbicides : destinés à lutter contre certains végétaux entrant en concurrence avec les plantes cultivées. Leur mode d'épandage est différent puisqu'ils sont déposés directement sur le sol, tandis que les autres produits sont plutôt pulvérisés sur la plante en croissance. Les herbicides constituent aujourd'hui la famille la plus importante en nombre de molécules et la plus utilisée[4].

I- 3- 2-Les fongicides: destinés à éliminer les moisissures et parasites fongiques (champignons) des plantes[4], on distingue 2 sortes de fongicides :

- Les produits préventifs empêchant le développement des spores à la surface de la plante.
- Les produits curatifs qui stoppent le développement du champignon déjà installé dans la plante.

Le mode d'action des fongicides est multiple. Il peut bloquer la respiration du champignon, affecter les défenses immunitaires, ou empêcher la synthèse des protéines.

I- 3- 3-Les insecticides : destinés à la lutte contre les insectes. Ils interviennent en tuant ou en empêchant leur reproduction. Ce sont souvent les pesticides les plus toxiques et c'est dans cette famille que l'on trouve la plupart des polluants organiques persistants, dont le DDT et la plus utilisée[4].

Les familles les plus rencontrées sont les organophosphorés (malathion), les carbamates insecticides (carbaxyle), les pyréthrinoïdes (deltaméthrine) et les organochlorés (en dosulfan).

I- 3- 4- Les insecticides organophosphorés :

Les pesticides organophosphorés sont des composés organiques contenant du phosphore. Classés parmi les plus couramment utilisés aujourd'hui, ils sont utilisés dans l'agriculture, à la maison, dans les jardins, et dans la pratique vétérinaire[5].

L'exposition aux mêmes organophosphorés par de multiples voies ou à plusieurs organophosphorés par de multiples voies peuvent conduire à de graves toxicités additives[5].

Etude bibliographique/chapitre I

Il est important de comprendre, cependant, qu'il existe un large éventail de la toxicité de ces agents et la grande variation est par l'absorption cutanée, ce qui implique des identifications spécifiques et de gestion très important[5].

On a plusieurs classes des insecticides organophosphorés :

- Organophosphorés aliphatique : acéphate, déméton, dichlorvos, dicrotophos, diméthoate, éthion, formotion, mévinphos, monocrotophos, nage, ométhoate, phorate, phosphamidon, trichlorfon, ils sont généralement hautement toxiques et peu stables[6].
- Organophosphorés à cycle phényl : bromophos, chlorfenvinphos, fénitrothion, fenthion, fonofos, parathion, parathionéthyl, parathion méthyl, phosalone, profénofos, protiofos, ils sont plus stables que le groupe précédent[6].
- Organophosphorés à hétérocycle : chlorpyrifos, diazinon, étrimfos, isoxation, quinalphos, méhidation, phosmet, des produits issus de ces 3 groupes sont regroupés ci-dessous selon leur mode d'action :
 - Produits de contact : bromophos, diazinon, fénitrothion, malathion, dichlorvos, fonofos, parathion, phosmet, profénofos, téméphos, hepténophos, mevinphos, trichlorophon, phosalone[6].
 - Produit systémiques : diméthoate, formotion, isofenphos, triazophos, monocrotophos, ométhoate, phosphamidon, thiométon, vamidothion[6].

I- 4- L'effet des pesticides sur l'environnement :

Les pesticides sont des substances qui possèdent une activité biologique qui agit sur diverses fonctions physiologiques et entraînent le plus souvent la mort de l'organisme vivant qui a pu l'absorber. Ce résultat est recherché pour lutter contre les ennemis des plantes et les parasites des bâtiments, des animaux et de l'homme mais leur usage peut aussi s'accompagner des inconvénients. En effet, l'utilisation des pesticides présente un certain nombre de risques à l'égard de la composition chimique de l'air, des eaux et des sols qui se traduisent par des pollutions dont les conséquences toxicologiques et écotoxicologiques peuvent être préjudiciables à la qualité de l'environnement[7].

Etude bibliographique/chapitre I

De manière générale les pesticides ont un effet néfaste sur l'environnement [8]:

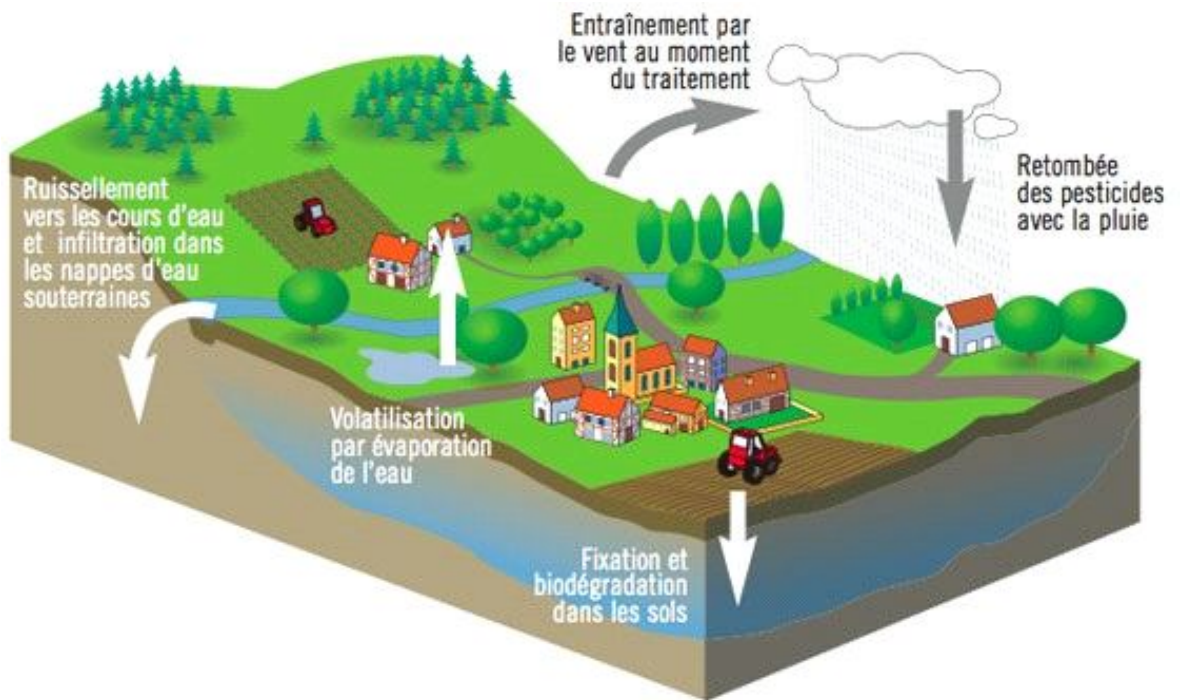


Figure 1 :Schéma montre le voyage des pesticides dans l'environnement[8]

I- 4- 1- L'impact sur l'air :

Dans l'air, l'INRA a installé des stations de mesures de pesticides dans les eaux de pluies 1955 et 1966. Les résultats furent apparemment accablants. En effet presque tous les échantillons contenaient des pesticides et 60% d'entre eux dépassaient la CMA $0.1 \mu\text{g}$ [8].

Ils existent plusieurs phénomènes qui sont à l'origine de la présence des pesticides dans l'air :

- Phénomène de dérivé : Se produit quand des fines gouttelettes restent dans l'atmosphère et voyagent sur de longues distances hors de la zone traitée. Ils se répandent sur terre grâce à l'eau de pluie[8].
- Phénomène d'érosion : Des sols traités se produisent par un transfert de pesticides par le vent sous forme de particules du sol et de poussières contaminées[8].

Etude bibliographique/chapitre I

- La volatilisation : se produit par un transfert de pesticides sous forme de gaz à partir des plantes ou du sol traités[8].

À travers ses différents phénomènes les pesticides entraînent des effets irréversibles sur l'environnement car ils pourraient avoir des impacts nocifs sur la faune et la flore non ciblé par le produit phytosanitaire, sur terre et dans l'eau[8].

I- 4- 2- L'impact sur le sol :

Lors d'un traitement aux pesticides, 90% des quantités de pesticides utilisés n'atteignent pas le ravageur visé. Ainsi l'essentielle des produits phytosanitaires aboutissent dans le sol où ils subissent les actions simultanés des phénomènes de transfert, d'immobilisation et de dégradation[8].

- **Phénomène de transfert :**

Il concerne 5% des produits phytosanitaires appliqués. Ils sont à l'origine de la pollution des eaux de surface lorsqu'ils sont entraînés, soit à l'état dissous par les eaux de pluies ou retenu sur des particules de terre elles-mêmes entraînées[8].

- **Phénomène d'immobilisation :**

Les pesticides sont adsorbés par les matières humiques du sol (situé dans la partie supérieure du sol.) Comme celles-ci sont adsorbées ils ne sont plus en solutions avec les phases liquides et gazeuses. Leurs effets biologiques sont supprimés. Ils ne sont plus dégradés par les microorganismes du sol, et plus entraînés par l'eau empêchant la pollution de celle-ci. La désorption (libération des molécules dans le sol, l'inverse de l'adsorption) leur rend toutes leurs capacités biotoxiques[8].

- **Phénomène de dégradation :**

Il est assuré par les Organismes biologique du sol comme les bactéries. Le sol est un filtre actif mais sélectif car il est capable de retenir certains produits comme l'oxychlorure de cuivre (bouillies bordelaises). Qui s'accumule dans le sol et qui a entraîné la stérilisation de 50.000 ha de bananeraies au Costa Rica[8].

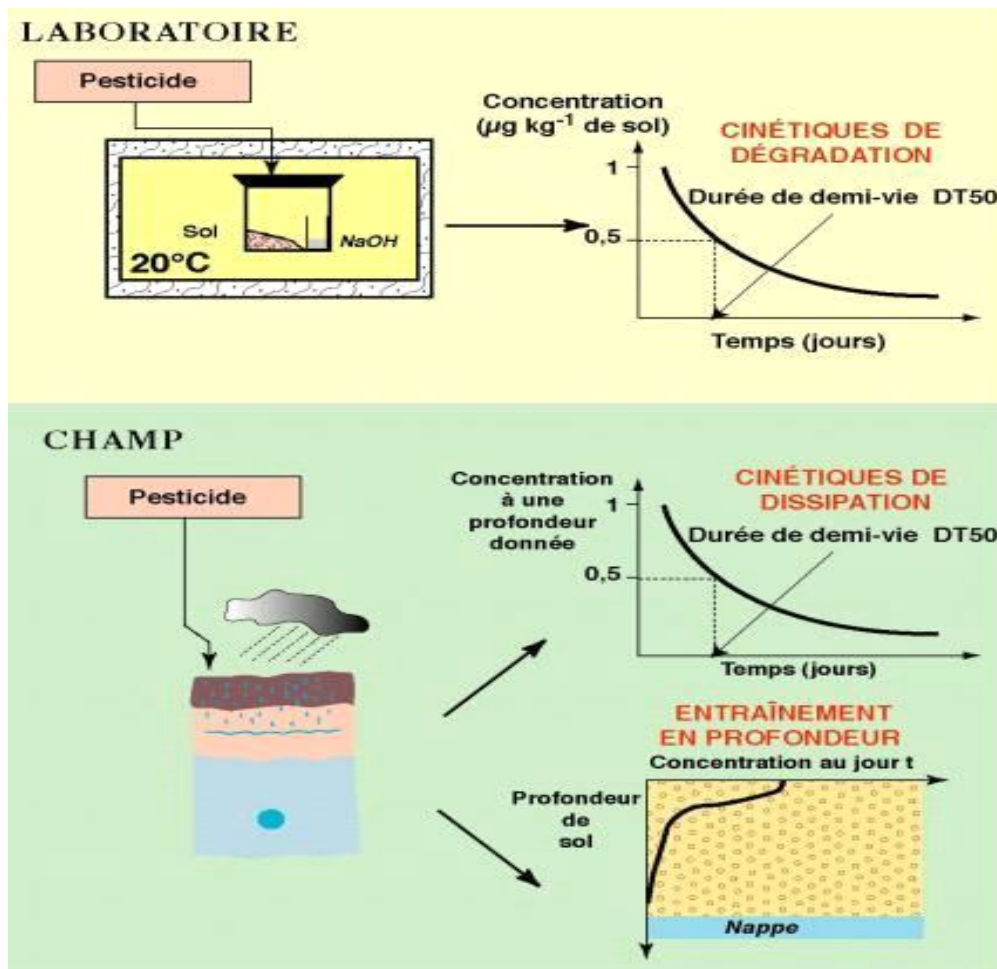


Figure 2:Caractérisation expérimentale de la dégradation des pesticides dans les sols[9]

I- 4- 3- L'impact sur l'eau :

Nous pouvons constater suites à plusieurs études réalisées par l'Institut français de l'environnement que les pesticides sont retrouvés dans 96% des eaux superficielles et 61% des eaux souterraines de France. 100% des cours d'eau sont contaminés par les pesticides. Les herbicides étant les résidus les plus retrouvés dans les eaux[8].

I- 4- 4- L'effet des pesticides sur les villes :

On pense souvent que les pesticides ne se trouvent que dans les zones rurales mais, c'est faux. Selon une étude réalisée par AirParif, des pesticides sont présents dans l'air au cœur de notre capitale jusqu'aux villes qui l'entourent. Dans 80 pesticides recherchés, une vingtaine a été

Etude bibliographique/chapitre I

retrouvée dans l'agglomération francilienne, certains d'entre eux peuvent avoir un effet néfaste sur l'environnement. La trifluraline est un herbicide, retrouvé en forte concentration en agglomération, c'est un polluant de l'eau ayant des propriétés cancérigènes possible[8].

I- 4- 5- L'impact sur Les maisons :

Les pesticides se retrouvent aussi dans nos habitations. Une étude américaine, réalisé par Neil Immerman en janvier 1990, à mis en évidence la présence d'au moins un pesticide dans l'air des maisons étudiées. Une autre, réalisé par Lewis en 1994, mis en évidence la présence de 8 à 18 pesticides différents par habitation. Il semblerait aussi que l'air présent dans nos habitation semblent plus contaminés par les pesticides que l'air extérieur, selon une étude réalisé par Lewis le 20 août 1989[8].

I- 5- Toxicologie des pesticides :

C'est le manque de sélectivité des pesticides vis-à-vis de leur cible qui provoque la plupart des effets nocifs pour l'environnement. Les animaux absorbent les pesticides via la nourriture ou l'eau d'alimentation, via l'air respiré ou au travers de leur peau ou de leur cuticule. Ayant franchi diverses barrières, le toxique atteint les sites du métabolisme ou est stocké. On utilise habituellement pour estimer la toxicité d'une substance chimique la dose (ou la concentration) qui provoque un effet particulier chez la moitié statistiquement de la population soumise au toxique (DE50 ou CE50). Si cet effet est la mort, on parle de dose (ou de concentration) létale 50 (DL50 ou CL50). La dose (ou la concentration) maximale sans effet (DMSE) est la dose immédiatement inférieure à celle qui provoque le moindre effet dans la même épreuve expérimentale (Severn et Ballard, 1990). Dans les cas, où les vitesses d'excrétion ou de métabolisation de la molécule sont faibles, où elle est liposoluble, où elle est fortement liée à d'autres constituants de l'organisme, sa concentration finale dans l'organisme sera plus élevée que dans le milieu de cet organisme (Madhun et Freed, 1990).

Pour ce qui est des substances liposoluble, la bioaccumulation dépend du coefficient de partage octanol/eau K_{ow} , si celui-ci est élevé et la vitesse de dégradation faible, la substance s'accumulera à des concentrations croissantes dans les organismes se succédant le long de la chaîne trophique. C'est ce qu'on appelle la bioaccumulation (biomagnification, cooper, 1991). Et

Etude bibliographique/chapitre I

il est évident qu'un pesticides qui suit ce processus est, à exposition et toxicité égale, plus dangereux pour l'environnement qu'un autre produit qui se s'accumule pas.

La toxicité, et notamment la toxicité chronique, se manifeste par les effets très divers. Pour ce qui est de l'impact sur l'homme, on doit prendre en compte, outre la toxicité proprement dite, les effets carcinogènes, immunodépresseurs, mutagènes, neurotoxiques et tératogènes (Hayes, 1991).

On a montré récemment que les pesticides étaient capables d'endommager le système immunitaire (Culliney, 1991) ou de perturber les régulations hormonales, tant chez l'homme que chez l'animal, provoquant des symptômes variés (Leblanc, 1995). parmi les problèmes de santé humaine, on a soupçonné un lien entre la présence de produits chimiques «perturbateurs endocriniens» et un taux accru de cancers du sein, de la prostate, du testicule, d'endométriose, de malformations congénitales de l'appareil reproducteur masculin et de réductions du nombre de spermatozoïdes (Hileman, 1994 ; Davis et Bradlow, 1995 ; Kelce et al., 1995)[10].

I- 6- Le marché des pesticides :

- Dans le monde :

Le marché mondial représente environ 40 milliards de dollars et à peu stable depuis les années 2000.

Les Etats-Unis sont le premier consommateur des pesticides dans le monde, survient l'Inde, la France, le premier consommateur en Europe, puis l'Allemagne. Malgré les recommandations du grenelle de l'environnement, la France est toujours championne de l'utilisation des produits phytosanitaires en Europe. On trouve des insecticides, des herbicides, fongicides, partout, y compris dans notre environnement.

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 diversifications des cultures, avec l'amélioration du niveau de vie dans certains pays, modifié également cet équilibre, ainsi la chine

Etude bibliographique/chapitre I

a converti l'équivalent de la surface de l'Angleterre de rizières en cultures maraîchères, entraînant une diversification des produits mise en œuvre.

Le Japon utilise 12 kg et est le premier consommateur de pesticides à l'hectare, l'Europe 3 kg, les Etats-Unis 2.5 kg, l'Inde gros producteur 0.5 kg/ha.

La consommation mondiale des pesticides est en augmentation constante depuis les années 40, passant de 0.49 kg/ha en 1961 à 2 kg/ha en 2004. 20% de la surface totale des Etats-Unis, 35% de celle de la France, sont soumis à des traitements[11].

- **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 a montrée que les Pyréthrinoïdes, 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 antiacridienne.

Le marché algérien en pesticides ne cesse d'augmenter, en 2009 l'Algérie a importé 67 millions USD de pesticides et en 2008, 77 millions USD contre 49.4 millions USD en 2007[12].

II- Phénomène de partage :

II- 1- Définition :

le partage d'une substance entre une phase aqueuse et une phase organique constitue une propriété physico-chimique très importante de cette substance, notamment en ce qui concerne sa nature lipophile ou hydrophile, la propriété de partage conditionne en partie les propriétés biologiques de la molécule considérée, telle que le transport, la diffusion à travers les membranes (une molécule qui présente un partage élevée sera très lipophile et diffusera donc beaucoup à travers la membrane et inversement), la biodisponibilité (distribution et accumulation), l'affinité pour un récepteur et la fixation par une protéine, l'activité pharmacologique ou encore la toxicité, s'agissant de contaminants, ce même partage conditionne leur devenir dans notre environnement en particulier leur accumulation dans les organismes aquatiques[13].

II- 2- historique de la constante de partage :

L'étude de la distribution d'une substance entre deux liquides non miscibles était l'objet d'études et de recherches expérimentales pendant plusieurs années. Les premières techniques de séparations étaient utilisées pour séparer des substances naturelles telles que les essences des fleurs[14,15].

Vers 1872 Berthelot et Jungfleisch étudièrent la distribution d'une substance entre deux liquides non miscibles, ils utilisèrent I₂ et Br dans le système eau-CS₂ et les acides HCl et H₂SO₄, entre l'eau et l'éthyl éther, leurs premières investigations étaient la variation du coefficient de partage K en fonction de la température[15].

en 1891, Nernst prouva que K n'est constant sauf que si la substance présente une forme moléculaire unique dans les deux phases, c'est-à-dire pas de dimérisation ou d'interaction entre les solvants[15].

en 1909, Hertz publia une formule reliant K au nombre d'extractions successives effectuées[13].

Etude bibliographique/chapitre II

$$X = X_0 \left(\frac{KV_s}{KV_e + V_e} \right)^i \quad (1)$$

Avec :

X_0 : Quantité de la substance

X : Quantité extraite

V_s : Volume de la phase organique

V_e : Volume de la phase aqueuse

i : Nombre d'extractions effectuées

Depuis les travaux de Collander à la fin des années 1950[15], puis ceux du groupe de HANSCH quelques années plus tard, le coefficient de partage K d'une molécule dans un système biphasique constitué de deux solvants non miscibles (le plus souvent le système octanol/eau), est reconnu pour sa faculté à mimer le passage de cette molécule à travers les membranes biologiques[13,15]. Pour des solutions diluées, ce coefficient de partage octanol/eau est le rapport de la concentration d'une molécule de soluté dans l'octanol sur sa concentration dans l'eau lorsque le système biphasique est en équilibre. Le système solvant octanol/eau est devenu un standard de mesure du coefficient de partage.

II- 3- Définition de coefficient de partage Octanol-Eau (Kow) :

Le Log Kow est une mesure de la solubilité différentielle de composés chimiques dans deux solvants, l'eau et l'octanol ; c'est le coefficient de partage octanol/eau : $\text{Log Kow} = \log (C_{\text{oct}}/C_{\text{eau}})$ où C_{oct} et C_{eau} représentent les concentrations du polluant respectivement dans l'octanol et l'eau. Cette valeur permet d'appréhender le caractère hydrophile ou hydrophobe (lipophile) d'une molécule. Le Log Kow permet d'évaluer le caractère polaire des molécules. En effet l'octanol est un solvant qui possède une polarité très similaire à celle des membranes lipidiques, c'est pourquoi il est utilisé pour évaluer le caractère lipophile d'une molécule. Si le Log Kow est positif et très élevé, cela exprime le fait que la molécule considérée est bien plus soluble dans l'octanol que dans l'eau, ce qui reflète son caractère lipophile, et inversement. Une valeur de Log

Etude bibliographique/chapitre II

$K_{ow} = 0$ signifie que la molécule se répartit de manière égale entre les deux phases et $C_{oct} = C_{eau}$. Ce K_{ow} rend compte de la tendance d'un polluant à se solubiliser et s'accumuler dans les membranes des organismes vivants.

Le LogP, aussi appelé $\log K_{ow}$, est une mesure de la solubilité différentielle de composés chimiques dans deux solvants (coefficient de partage octanol/eau).

Le LogP est égal au logarithme du rapport des concentrations de la substance étudiée dans l'octanol et dans l'eau. $\text{LogP} = \text{Log} (C_{oct}/C_{eau})$. Cette valeur permet d'appréhender le caractère hydrophile ou hydrophobe (lipophile) d'une molécule. En effet, si le LogP est positif et très élevé, cela exprime le fait que la molécule considérée est bien plus soluble dans l'octanol que dans l'eau, ce qui reflète son caractère lipophile, et inversement, si le LogP est négatif cela signifie que la molécule considérée est hydrophile. Un LogP nul signifie que la molécule est aussi soluble dans un solvant que dans l'autre. Le LogD est une mesure plus raffinée du LogP tenant compte de la modification de celui-ci en fonction du pH de la phase aqueuse et du PKa du composé étudié[16].

II- 4- les paramètres influençant sur la constante de partage :

Le coefficient de partage k est une constante thermodynamique qui dépend de la nature de la substance à extraire, et du système de solvants utilisés partant du principe « qui se ressemble s'assemble », un solvant dissout bien un composé qui lui ressemble, les solvants polaires et dissociants tels que l'eau dissolvent les composés ioniques et / ou polaires hydrophiles. Les molécules apolaires présentent souvent un $k > 10$ entre un solvant hydrophobe et l'eau, les solvants apolaires et peu dissociants dissolvent les molécules et les composés hydrophobes (solvants chlorés ou hydrocarbures).

Le solvant intervient aussi par son caractère protique ou aprotique, les solvants protiques tel que l'eau, ammoniac, alcools, phénols, acides et amides non substituées influent sur l'extraction à cause de l'existence d'un hydrogène labile.

D'autre part, la nature du soluté intervient aussi dans le phénomène de partage, en effet, la structure d'une molécule organique joue un rôle important sur son coefficient de partage, l'accroissement de la chaîne augmente la valeur de k d'environ 4 unités pour chaque groupement

Etude bibliographique/chapitre II

méthylénique (CH_2) incorporé dans la molécule. Un composé ramifié présente un k inférieure à celui de son isomère linéaire, il en va de même pour un composé non saturé par rapport au composé saturé correspondant. La présence d'hétéro atome (O, N) diminue parfois considérablement la valeur de k à cause des liaisons hydrogène avec l'eau[17]. L'introduction d'un halogène au contraire favorise le passage en phase organique donc l'augmentation de K [17].

L'inclusion d'une molécule simple dans un complexe chargé diminuera fortement le coefficient de partage, en effet la forme ionique ainsi obtenue est du fait de sa charge beaucoup plus fortement retenue par l'eau, solvant polaire. Une molécule possédant un moment dipolaire (μ) non nul, définissant en quelques sorte l'énergie de liaison van der Waals, la constante diélectrique (ϵ) elle, définit la polarisabilité moléculaire (ou polarisation moléculaire), elle permet de se rendre compte de l'affinité des solvants entre eux même[17].

En résumé, pour une substance donnée, k augmente avec :

- la longueur de chaîne.
- La linéarité de la chaîne.
- La saturation de la chaîne.
- Le contenu en chlorure.
- Une diminution du contenu en oxygène ou azote.

III- Les modèles QSAR/QSPR :

Au cours des décennies passés, les relations quantitatives structure- activité/propriété (QSAR/QSPR) sont devenue un puissant outil théorique, alternatif à la mécanique quantique, pour la description et la prédiction des propriétés des systèmes moléculaires complexes dans différents environnements. L'approche QSAR/QSPR procède de l'hypothèse d'une correspondance univoque entre n'importe quelle propriété physique, affinité chimique, ou activité biologique d'un composé chimique et sa structure moléculaire. Cette dernière peut être représentée par la composition chimique, la connectivité des atomes, la surface d'énergie potentielle, et la fonction d'onde électronique d'un composé. Différents descripteurs moléculaires physico-chimique reflétant la structure peuvent être déterminés empiriquement où en utilisant des méthodes théoriques et computationnelles de différentes complexités. Il est à souligner que la connaissance de la constitution chimique exacte et/ou de la structure moléculaire tridimensionnelle des composés chimiques étudiés est un prérequis à l'application de l'approche QSAR/QSPR.

Le succès de l'approche QSAR/QSPR dépend de façon critique de la définition précise et de l'utilisation appropriée des descripteurs moléculaires. On distingue, arbitrairement, les descripteurs moléculaires empiriques des descripteurs moléculaires théoriques[16].

III- 1- Définition de QSPR :

Le QSPR (quantitative structure-propertyrelationships) est le procédé par lequel des liens quantitatifs sont établis entre la structure moléculaire d'un ensemble de composés avec une propriété physico-chimique. Les grandes phases de développement d'un modèle QSPR peuvent être décrites comme suit :

- choisir des descripteurs adaptés au problème structure-propriété.
- Exploiter les valeurs des descripteurs comme variables, afin de définir une relation qui les corrèle à la propriété en question, à l'aide de machines d'apprentissage. C'est la fouille de données.

Etude bibliographique/chapitre III

- Etablir des critères de performance et de validation qui aideront au choix des meilleurs modèles pour le problème posé et estimer des incertitudes de prédiction[16].

III- 2- Principe de QSPR :

Le principe des méthodes QSPR est, comme leur nom l'indique, de mettre en place une relation mathématique reliant de manière quantitative des propriétés moléculaires aussi bien électroniques que géométriques, 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. Aussi la forme générale de modèle est :

$$\text{Propriété} = f(\text{descripteurs}) \quad (2)$$

L'objectif d'une telle méthode est donc d'analyser les données structurales afin de détecter les facteurs déterminants pour la propriété mesurée. Pour ce faire, différents types d'outils peuvent être employés : régressions multilinéaires (MLR), régressions aux moindres carrés partiels (PLS), arbres de décision, réseaux de neurones, algorithmes génétiques.

Une fois cette relation mise en place et validée sur un jeu validation, elle peut alors être employée pour la prédiction de la propriété de nouvelles molécules, pour lesquelles la valeur expérimentale n'est pas disponible, voir pour des molécules encore non synthétisées. De tels modèles peuvent également, dans certains cas, être utilisés pour mieux appréhender les phénomènes moléculaires mis en jeu dans la propriété d'intérêt[16].

III- 3- Les méthodes Mathématiques Utilisés par le modèle QSPR :

III- 3- 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 $i=1, \dots, n$:

$$y_i = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} + \xi(3)$$

Etude bibliographique/chapitre III

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 ξ_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 estimés \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 :

$$E(e) = 0$$

- b. Le modèle choisi est correct (aucune variable explicative n'a été omise).
- c. Les résidus sont indépendants entre eux :

$$E(e_i, e_j) = 0$$

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-gausse).

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 tel que [18]:

Etude bibliographique/chapitre III

$$n-p-1 \geq 10 \quad (4)$$

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* [18] à la condition $FIV < 5$.

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

- Le coefficient de détermination multiple :

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

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

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

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

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

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

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

La validation croisée par « leave-one-out » (LOO) [19] 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

Etude bibliographique/chapitre III

des carrés des erreurs de prédiction, désignée par le symbole PRESS, est une mesure de la dispersion des estimations. On l'utilise pour définir le coefficient de prédiction [19]:

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

Contrairement à R^2 qui augmente avec le nombre de paramètres du modèle, le facteur Q^2_{LOO} affiche une courbe avec maximum (ou avec palier) obtenue pour certain nombre de descripteurs, puis décroît de façon monotone. Ce fait confère une grande importance au coefficient Q^2_{LOO} . Une valeur de $Q^2_{\text{LOO}} > 0,5$ est considérée comme satisfaisante, et une valeur supérieure à 0,9 est excellente [20].

Si de petites valeurs de Q^2_{LOO} 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^2_{LOO} 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 – may – 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.

III- 3- 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.

Etude bibliographique/chapitre III

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.

III- 3- 3- régression moindre carré (PLS) :

Une alternative à la RCP consiste à utiliser la PLS[21,22]. La méthode PLS exprime une variable dépendante (y) en termes de combinaisons linéaires des variables indépendantes originales (x) selon :

$$y_1 = b_1t_1 + b_2t_2 + b_3t_3 + \dots + b_mt_m$$

ou : $t_1 = c_{11}x_1 + c_{12}x_2 + \dots + c_{1p}x_p$

$$t_2 = c_{21}x_1 + c_{22}x_2 + \dots + c_{2p}x_p$$

$$t_m = c_{m1}x_1 + c_{m2}x_2 + \dots + c_{mp}x_p$$

t_1 , t_2 , etc..., sont appelés variables latentes (ou composantes) et sont construits de telle façon à former un ensemble orthogonal. L'utilisation de combinaisons linéaires orthogonales des valeurs de x est similaire à l'ACP. La différence majeure est que les variables latentes en PLS sont construites pour expliquer à la fois la variabilité des variables indépendantes x , et la variabilité dans les observations y .

III- 4- Calcule des descripteurs moléculaires :

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[23] 6.03 pour représenter les molécules puis, à l'aide de la méthode semi-empirique AM1, obtenir les

Etude bibliographique/chapitre III

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-Ribierre avec pour critère une racine du carré moyen du gradient égale à 0,01 kcal/mol. Les géométries ainsi optimisées ont été transférées dans le logiciel informatique DRAGON[18] 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 ($R \geq 0,9$) 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.

III- 5- 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).

III- 6- 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[24,25], la comparaison se fait souvent à l'avantage des algorithmes génétiques[26].

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 dont les individus sont mal adaptés au moyen d'opérateurs génétiques de reproduction et de mutation.

Etude bibliographique/chapitre III

Après un certain nombre de générations, la population est constituée d'individus bien adaptés, autrement dit des solutions supposées «bonnes» au problème d'optimisation.

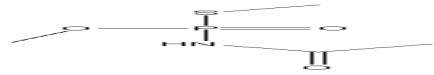
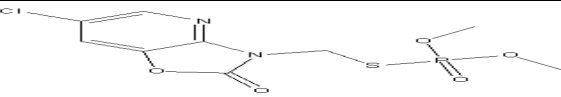

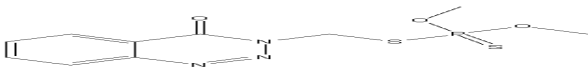

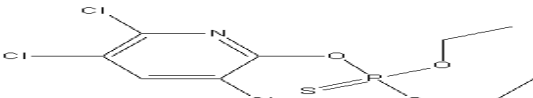
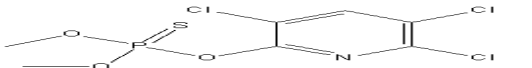


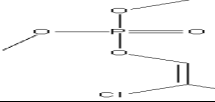
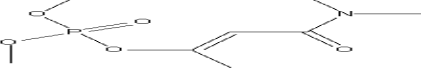

Dans le présent travail, la sélection des descripteurs a été réalisée par algorithme génétique, dans la version MOBY DIGS de Todeschini[27], en maximisant Q^2_{LOO} .

Application

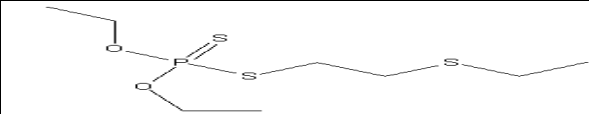
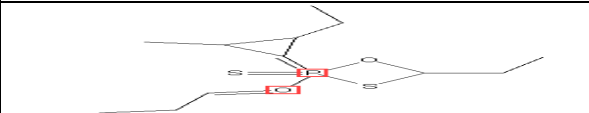
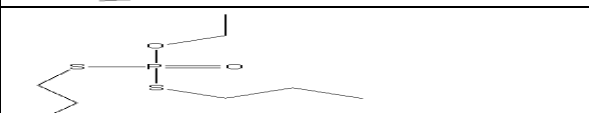
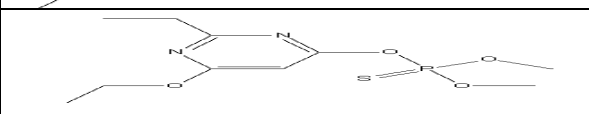
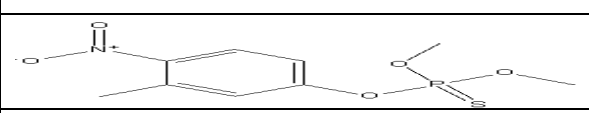
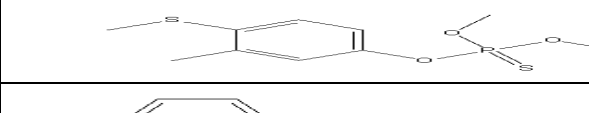
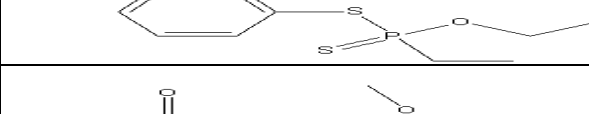
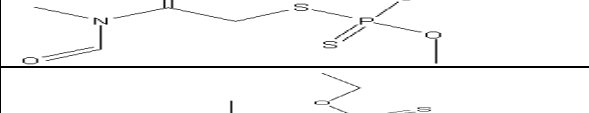
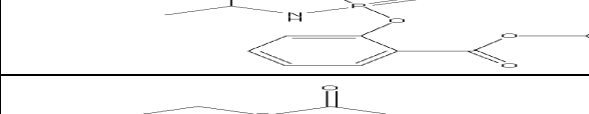
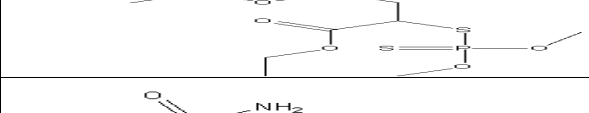
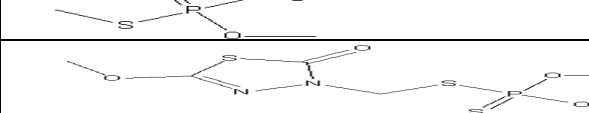
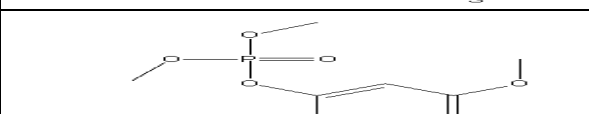
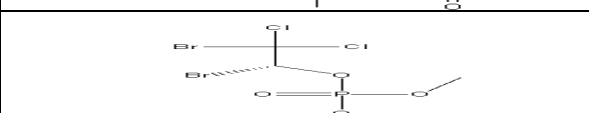
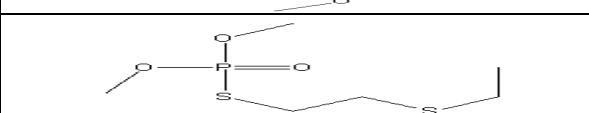
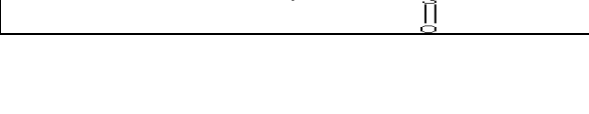
Dans cette partie expérimentale, nous allons déterminer la constante de partage octanol-eau d'une série de pesticides par modélisation d'un ensemble d'estimation. Le meilleur modèle obtenu sera validé par l'approche hybride algorithme génétique et régression multilinéaire GA/RLM.

La liste des molécules étudiées dans notre travail présente dans le Tableau 1, cette base de données contient 43 molécules [28], avec la propriété à étudier (la constante de partage n octanol-eau (Kow)) et les variables explicatives (descripteurs) sélectionnées par algorithme génétique.

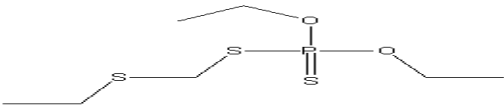
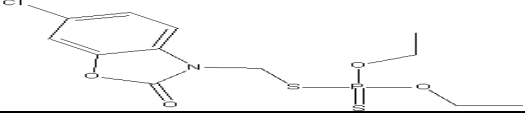
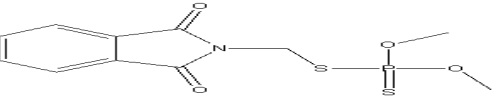
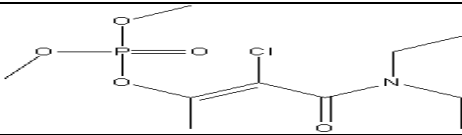
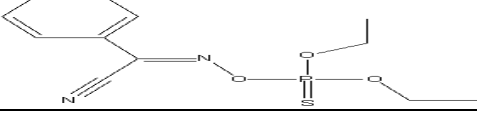
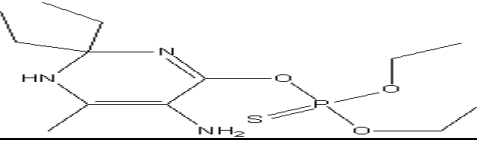
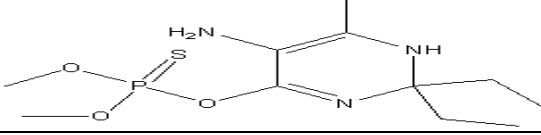
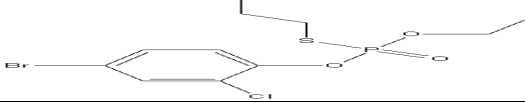

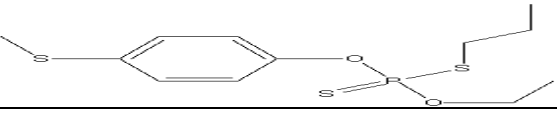
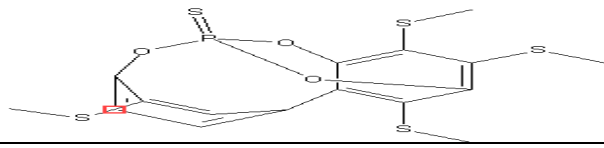
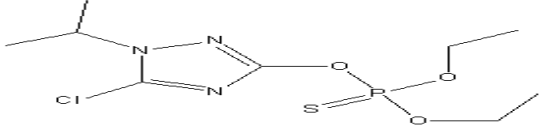
Tableau I : Structure des composés étudiés

N°	Composés	Structures	Log kow
1	acéphate		-0.89
2	azamethiphos		1.05
3	azinephos-eyhyl		3.18
4	azinephos-methyl		2.56
5	chlorfenvinphos		3.95
6	chlorpyrifos		4.7
7	Chlorpyrifosmethyl		4.24
8	cyanophos		2.65
9	diazinon		3.74
10	dichlorvos		1.16
11	dicrotophos		-0.49
12	dimethoate		3.46

Application

13	disulfoton		3.95
14	ethion		5.07
15	ethoprophos		3.59
16	etrimfos		3.3
17	fenitrothion		3.43
18	fenthion		4.09
19	fonofos		3.94
20	formothion		-0.56
21	isofenfos		4.12
22	malathion		2.75
23	methamidophos		-0.8
24	methidathion		2.2
25	mevinfos		0.13
26	naled		1.38
27	oxydometon-methyl		-0.74

Application

28	phorate		3.56
29	phosalone		4.3
30	phosmet		2.78
31	phosphamidon		0.79
32	phoxim		3.38
33	pirimiphos-ethyl		4.85
34	pirimiphos- methyl		4.2
35	profenofos		4.44
36	propetamphos		3
37	sulprofos		5.48
38	temephos		5.96
39	isazofos		3.82

I- Calcul du modèle :

I- 1- Choix du modèle :

Les molécules sont dessinées par le logiciel ChemDraw[29] puis elles sont optimisées en utilisant le logiciel HyperChem[23]. Les descripteurs moléculaires ont été calculés à l'aide du logiciel informatique Dragon [18], plus de 1600 descripteurs sont calculés.

L'ensemble des données a été décomposé en deux sous ensembles aléatoirement, 75% de la totalité des composés pour la construction du modèle et 25% pour la validation externe.

D'après l'algorithme génétique dans la version MobyDigs[27], 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 10 composés, alors que sa calibration par 33.

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 II.

Tableau II: Descripteurs moléculaires intervenant dans la modélisation de la constante de partage octanol -eau (log kow).

N°	Descripteur	Classe	Signification
1	Ms	Constitutional descriptors	meanelectrotopological state
2	MATS2m	Autocorrelation index	Broto-Moreau autocorrelation of a topological structure - lag 2 / weighted by atomic masses
3	BELv3	Burdeneigenvaluedescriptors	highest eigenvalue n. 3 of Burden matrix / weighted by atomic van der Waals volumes
4	R8p	GETAWAY descriptors	R autocorrelation of lag 8 / weighted by atomic polarizabilities
5	ALOGP	Molecularproperties	Ghose-Crippen octanol-water partition coeff. (logP)

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 III

Tableau III: Matrice de corrélation

	log K _{ow}	Ms	MATS2m	BELv3	R8p
Ms	-0,746 0,000				
MATS2m	0,438 0,011	-0,138 0,444			
BELv3	0,601 0,000	-0,654 0,000	0,413 0,017		
R8p	0,720 0,000	-0,590 0,000	0,332 0,059	0,776 0,000	
ALOGP	0,897 0,000	-0,624 0,000	0,360 0,039	0,537 0,001	0,602 0,000

Les cinq descripteurs (Ms;MATS2m; BELv3. R8p ; ALOGP) (tableau II) permettent de corréler log k_{ow}des 33composés des pesticides, la corrélation entre le log Kow et le MATS2m est faible et malgré sa porte une certaine information complémentaire au coefficient de partage octanol/ eau.

II- 2- Equation et analyse de régression :

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

$$\log K_{ow} = 10.6 - 1.89 Ms + 2.22 MATS2m - 3.43 BELv3 + 4.69 R8p + 0.642 ALOGP(9)$$

Tableau IV: Analyse de régression

Régresseur	Coef	Er-T	coefT	P	FIV
Constante	10,561	2,872	3,68	0,001	
Ms	-1,8895	0,4579	-4,13	0,000	2,4
MATS2m	2,2210	0,7693	2,89	0,008	1,4
BELv3	-3,432	1,446	-2,37	0,025	3,3
R8p	4,686	1,455	3,22	0,003	2,9
ALOGP	0,64179	0,09139	7,02	0,000	2,1

Les valeurs de T sont presque proche se qui confirme que la contribution des descripteurs sélectionnés par algorithme génétique, dans le modèle très proche et de même pourcentage.

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

Les paramètres statistiques sont présentés dans le tableau suivant (tableau V)

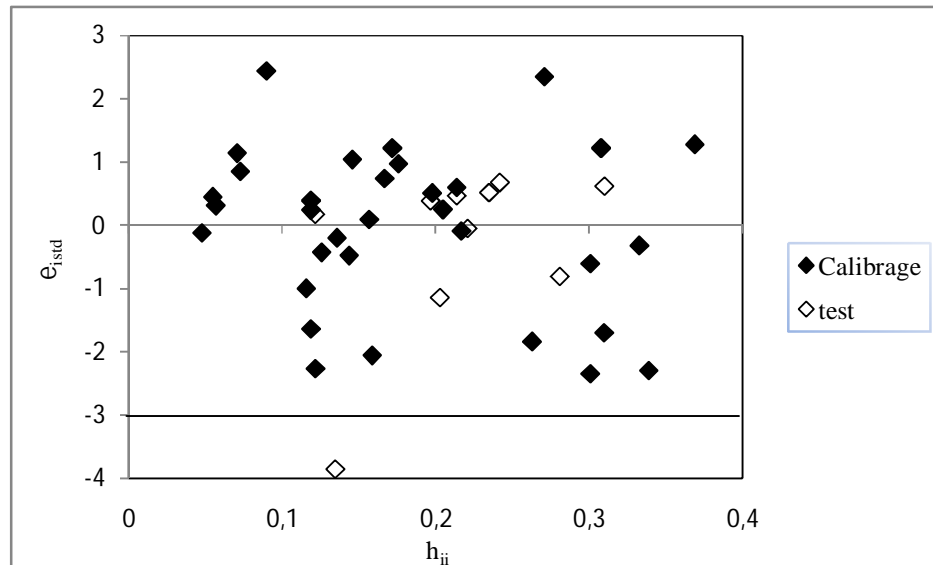
Tableau V: Les paramètres statistiques pour l'ensemble de calibration (33 composés) et celui de validation (10 composés),

Les paramètres statistiques	
F	60,16
R ²	91,76
R _{adj} ²	90,24
Q _{LOO} ²	87,26
Q _{BOOT} ²	83,89
Q _{ext} ²	84,36
S	0,578
EQMC	0,523
EQMP	0,65
EQMP _{ext}	0,72

Les paramètres statistiques reproduit dans le tableau ci-dessus montrent une bonne explication de la variabilité Kow, par les descripteurs choisis, de l'ordre de 90 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= 60.16), 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

Le diagramme de Williams est représenté dans la figure suivante, ou on a représenté sur la même figure : 03 pour les deux ensembles (calibration et validation) ; les valeurs des erreurs standardisés ($e_{i\text{std}}$) en fonction des valeurs de Leviers (h_i).



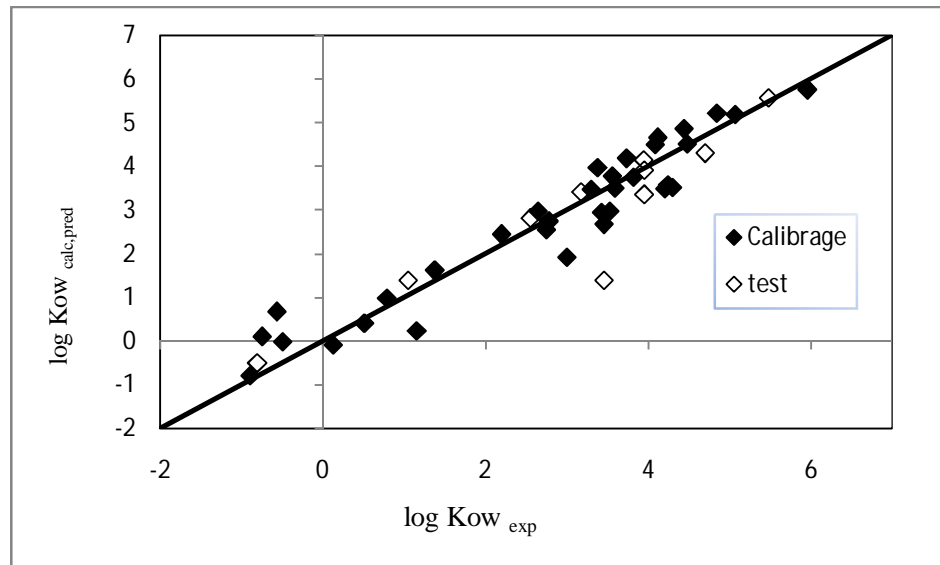


Figure 4: Graphe des valeurs calculées, prédites en fonction des valeurs observées du log Kow

D'après le graphe on remarque une faible dispersion autour de la première bissectrice. Ceci est confirmé par la valeur élevée de Q^2_{LOO} .

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 .

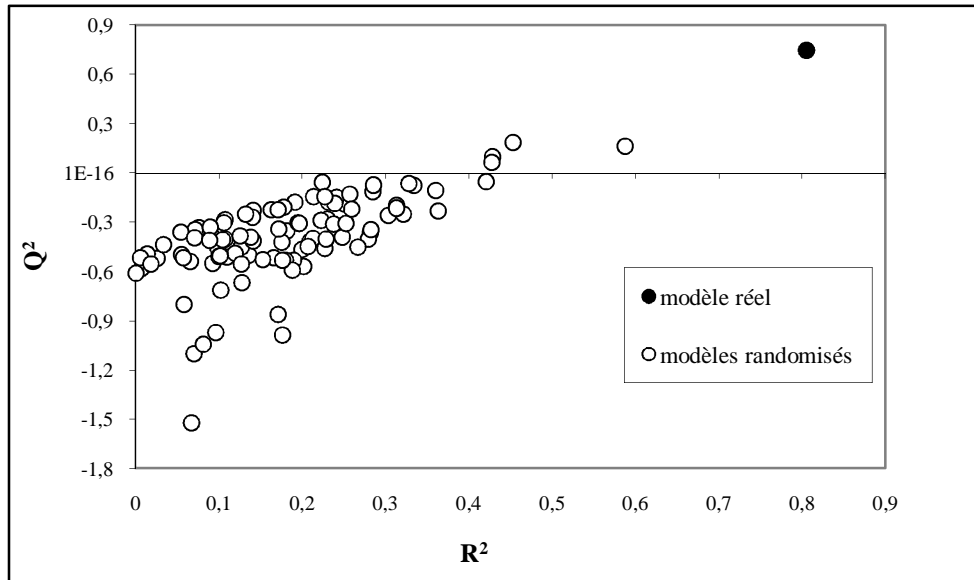


Figure 5: Test de randomisation

Il est clair que les statistiques obtenues pour les Kow 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.

Conclusion générale

Nous avons appliqué la méthodologie QSPR pour relier la propriété coefficient de partage octanol /eau de 33 pesticides, à des descripteurs moléculaires calculés à l'aide des logiciels spécialisés.

Dans le but de la recherche des corrélations linéaires entre la variable dépendante et les variables indépendantes sélectionnées par algorithme génétique, nous avons utilisé la méthode de la régression linéaire multiple (RLM).

Le modèle obtenu à cinq descripteurs, présente le meilleur modèle car les statistiques calculées permettent de juger sa bonne qualité de l'ajustement, sa robustesse interne et externe, aussi sa capacité prédictive...

Le test de randomisation montre, que le modèle obtenu n'est pas dû au hasard.

La relation entre le coefficient de partage octanol/ eau (K_{ow}) et les descripteurs moléculaires est fondamentalement linéaire.

Références bibliographiques

[1] K.M. Moussaoui, R. Boussahe, Y.Tchoulak, O. Haouchine, M. Benmami, N. Dalachi, utilisation évaluation et impacts des pesticides en algérie. Ecole nationale polytechnique. 2001.

(Consulté le 16-06-2017).

[2] J- Sabench (confédération paysanne, commission pesticides) MARDI 23 NOVEMBRE-MATINÉE(Consulté le 16-06-2017).

[3] http://www.memoireonline.com/11/12/6459/m_Etude-sur-les-pesticides6.html (Consulté le 16-06-2017).

[4] extrait d'un communiqué du Service du développement durable et environnement de la ville de Drummondville

<http://lolo200.unblog.fr/alimentation/> (Consulté le 16-06-2017).

[5] http://www.memoireonline.com/11/12/6459/m_Etude-sur-les-pesticides6.html (Consulté le 16-06-2017).

[6] http://www.memoireonline.com/11/12/6459/m_Etude-sur-les-pesticides0.html

INERIS DRC-00-23449- AIRE -569a-Cdu-FMr laboratoire central de surveillance de la qualité de l'air ; unité qualité de l'air- direction des risques chroniques ; loi sur l'air- convention 18/99, DECEMBRE 2000 (Consulté le 16-06-2017).

[7] AwatefBerrah Université de Tébessa Algérie - Master 2 en toxicologie appliquée 2011.

http://www.memoireonline.com/11/12/6459/m_Etude-sur-les-pesticides0.html (Consulté le 16-06-2017).

[8] <http://tpe-pesticides.over-blog.com/article-ii-effets-64479362.html> (Consulté le 16-06-2017).

[9] <http://pesticides.e-monsite.com/pages/ii-impact-sur-l-environnement/1-l-impact-sur-le-sol.html> (Consulté le 16-06-2017).

- [10] A –CALDERBANK , the occurrence and significance of bound pesticide residues in soil. *Reviews of environmental contamination and toxicology*, 108, 71-103, 1989. (Consulté le 16-06-2017).
- [11] Plan de surveillance des fruits et légumes 04 de la DGCCRF, Direction générale de l'alimentation, IFEN, Institut français de l'alimentation, Voir aussi Sciences et Vie 1072, Capital février 07, "Pesticides, agriculture et environnement" expertise Inra-Simagref 2005, "Les pesticides dans l'eau potable" Direction générale de la santé 2003, "Les pesticides dans les eaux", étude IFEN 2006) (Consulté le 16-06-2017).
- [12] W ;WANG ;B.MENG ;X.LU ;Y.LLIU ;S.TAO.J.analyticachimica acta 602(2007)211.222. (Consulté le 16-06-2017).
- [13] A. Capry ; Importance de la lipophilie en modélisation moléculaires, LPTC, UFR de chimie, Université de Bordeaux, France. (p 3-5. P 27). 1999. (Consulté le 16-06-2017).
- [14] A. Leo, C. Hansch, D. Elkins, Partition Coefficients and Their Uses, *Chemical Reviews*, Chimiques de Colorants Organiques, Université Joseph Fourier. Department of Chemistry, California. 71 (6). P 525-554. 2006. (Consulté le 16-06-2017).
- [15] A. Leo, C. Hansch, D. Elkins, Partition Coefficients and Their Uses, *Chemical Reviews*, Chimiques de Colorants Organiques, Université Joseph Fourier. Department of Chemistry, California. 71 (6). P 525-554. 2006. (Consulté le 16-06-2017).
- [16] D. Regadi, Développement de Modèle QSPR Pour La Prédiction des Propriétés Physiques des Quelques Composés Odorants. 2013 (Consulté le 16-06-2017).
- [18] R. Todeschini, V. Consonni, M. Pavan. Dragon, Software for The Calculation of Molecular Descriptors. Release 5.3 for Windows, MILANO. 2005 (Consulté le 16-06-2017).
- [19] R. Wehrens, H. Putter, L. M. C. Buvidens. *Chemom. Int. Lab. Syst.*, (p,35-52,54), 2000. (Consulté le 16-06-2017).
- [20] L. Eriksson, J. Jaworska, A. A. Worth, D. Cronin, R. M. McDowelle, P. Gramatica. *Methods for Reliability and Uncertainty Assessment and for Applicability Evaluation of Classification and*

RegressionBasedQSAR_s. EnvironmentalHealthPerspectives (111), 1361-1375. 2003. (Consulté le 16-06-2017).

[21] S. Woldchemometrics : Mathematics and StatisticsIn Chemistry. Reidel, Dordrecht, The Netherlands, 1984. (Consulté le 16-06-2017).

[22] F. Lindgren, P. Gelada, S. Wold. J. Chemometrics 8, 349, 1994. (Consulté le 16-06-2017).

[23] HyperchemTMRelease 6.03 for windows, MolecularModelingSystem, 2000 (Consulté le 16-06-2017).

[24] A. R. Leach. Modeling. Principales and Application, 2ndedPearson PrinticeHall, London 2001. (Consulté le 16-06-2017).

[25] A.R. Leach, V.L. Gillet An Introduction to Chemoinformatics : RevisedEdition ; Springer, Dordrecht, 2007. (Consulté le 16-06-2017).

[26] L. Chambres. Practical of Handbook of GenticAlgorithms. Lewis Publishing, 1995. (Consulté le 16-06-2017).

[27] R. Todeschini, B. Ballabio, V. Consonni, A. M. Pavan. MobydigsSoftware for MultilinearRegressionAnalysis and Variable SubestSelection by GeneticAlgorithm. Release 1.0 for Windows, Milano,2004. (Consulté le 16-06-2017).

[28] Ole. Christian Hansen, Quantitative StructureActivity Relation ShipsQSAR and Pesticides 2004. (Consulté le 16-06-2017).

[29] ChemDrew Ultra 7.0 (Consulté le 16-06-2017).