



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 :

**Etude de l'évolution des polluants dans
l'environnement par modélisation de la
constante de Henry**

Réalisé par :

-AMRANI Mohamed el Bachir .

- BELHABES Elayeche .

Dirigé par : M^{elle} BOUAKKADIA A

Membres de jury :

BENYAZA Nabil..... MAA

BOUAKKADIA Amel..... MCB

KERTIOU Nour Eddine.....MAA

UA L KhenchelaPrésidente

UA L Khenchela..... Encadreur

UAL Khenchela Examineur

Promotion: 2017/2018

Remerciement

Avant tout nous remercions le Dieu tout puissant qui nous a donné la force et la patience d'accomplir ce travail.

Nous tenons à remercier très chaleureusement notre plus profonde gratitude notre plus grand respect a M^{elle}/BOUAKKADIA Amel qui nous assistées et guidées, nous vous remercions d'avoir été toujours présent et nous avoir suivies régulièrement pour la réalisation de ce travail, pour sa gentillesse, ces conseils précieux et ces encouragement.

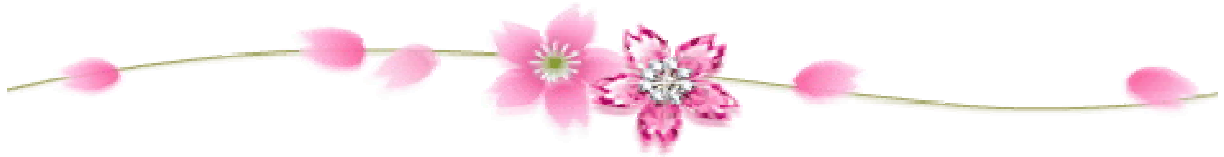
Un grand merci aux membres du jury :

Mr/BENYAZA NABIL

Mr/KERTIOU Nour eddine

.Un grand merci a tous.

Merci



Dédicaces

Je dédie ce travail à mes parents pour leur encouragement

Et leur morale Qui a toujours cru en moi

Et su mes redonner confiance lorsque la motivation n'était plus au rendez-vous.

A mes frères et ma sœur

A ma femme : Selma

A mon fils : Ouail

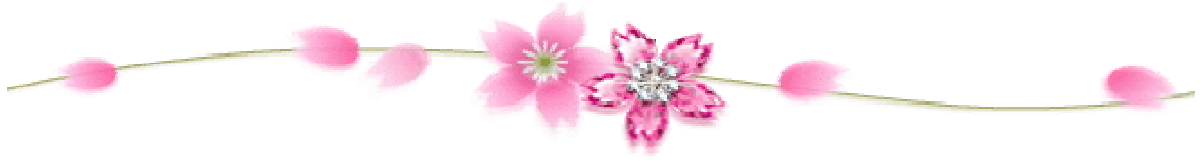
A vous tous

Je dédie ce travail

...

BACHIR





Dédicace

A ma mère et mon père qui m'a soutenu et encouragé dans tous les domaines et surtout pour

Réaliser ce mémoire

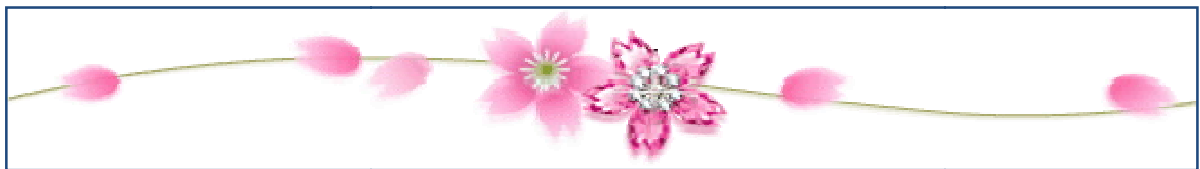
A mes chers frères et sœurs

A toute la famille

A tous mes amis

...

LAYECHE.



ت هذه
الخوازميات الجينية ، باستخدام نهج QSPR .
ض المركبات الملوثة مرتبطة خطياً بالصفات الجزيئية النظرية التي اختارتها

تم تقسيم سلسلة البيانات عشوائياً إلى مجموعتين فرعيتين
للتحقق من صحتها. 7 24

وسيتم تأكيد
والانحراف المعياري).
ي تم وضعها للمعايير
(معامل التحديد و Q^2 R^2 :)

التطبيق باستخدام مخطط وليام الذي كشف عن نقطة مؤثرة.

Résumé

La constante de Henry 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 24 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, qui fait ressortir un point influent de l'ensemble de validation.

Abstract

Henry's constant of a homogeneous set of some pollutant compounds was correlated linearly with theoretical molecular descriptors selected by genetic algorithms, using a QSPR approach.

The data series was randomly divided into two subsets, a set of 24 compounds for model construction and a set of 7 compounds for validation.

The different statistics established for the calibration and validation sets (coefficient of determination and prediction: R^2 , Q^2 and standard deviation s) will be confirmed.

The field of application was discussed using a William's diagram which revealed an influential point of the validation set.

Sommaire

Introduction générale.....	2
Etude bibliographique	
I. Généralités sur les pesticides :	4
I. 1. Définition des pesticides :	4
I. 2. Historique.....	4
I. 3. Catégories des pesticides.....	6
I. 3.1. Catégorie par origine	7
I. 3.2. Catégorie par niveaux de risque	7
I. 3.3 .Catégorie par usage.....	7
I. 4. Les Pesticides et l'Environnement :	8
I. 4.1. Contamination des milieux par les pesticides:.....	9
I. 5. Toxicologie des pesticides.....	12
I. 6. Le marché des pesticides :	13
II. Phénomène de Partage :	14
II. 1. Définition :.....	14
II. 2. Historique de la Constante de Partage :.....	14
II- 3. Paramètres Influençant sur la Constante de Partage :.....	15
III- Les Modèles QSAR/QSPR :	16
III. 1. Principe de QSPR :.....	17
III. 2. Importance des bases de données.....	18
III. 3. Base de données	18
III. 4. Régression linéaire multiple (MLR) :	19
III. 5. Régression en Composantes Principales (RCP) :	22
III. 6. Régression Moindre Carré (PLS) :	22
III. 7. Calcule Des Descripteurs Moléculaire :.....	23
III. 8. Les Types d'Apprentissages d'un Modèle QSPR :	23
III. 9. Sélection d'un Sous- Ensemble de Variables par Algorithme Génétique (GA-VSS) : 24	
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 :	33
II- 3- Diagramme de Williams	35
II- 4- Qualité de l'ajustement :	35
II- 5- Test de randomisation :	36
Conclusion générale :	38
Références bibliographiques.....	39

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}	Éléments 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

Liste des figures

Figure 1 : Type de pesticide	6
Figure 2: Processus et transfert des pesticides dans les sols (Barruiso et <i>al.</i> , 1996)	9
Figure 3 : Principe de la méthode QSPR.....	17
Figure 4: Diagramme de Williams	35
Figure 5 : Graphe des valeurs calculées, prédites en fonction des valeurs observées du	36
Figure 7 : Test de randomisation	36

Liste des tableaux

Tableau I : Structure des composés étudiés	25
Tableau II : Descripteurs moléculaires intervenant dans la modélisation de la constante de Henry (log cte H).....	31
Tableau III : valeurs de log cte H , AAC, RDF010u, Mor 13u.....	32
Tableau IV : Matrice de corrélation	33
Tableau V:Analyse de régression.....	33
Tableau VI : Les paramètres statistiques pour l'ensemble de calibration (17 composés) et celui de validation (7 composés).....	34

INTRODUCTION
GENERALE

Introduction générale

L'usage des insecticides, herbicides, fongicides, etc. regroupés sous le nom de pesticides ou encore produits phytosanitaires, a permis d'améliorer les rendements et la diversité des cultures afin de satisfaire la demande nutritionnelle liée à l'accroissement de la population mondiale. Cependant, cette utilisation a également provoqué des effets indirects et néfastes sur l'environnement. Ainsi des études ont montré la présence de résidus de pesticides dans les aliments ainsi que la contamination des eaux souterraines et superficielles. Malheureusement, une très grande partie des pesticides répandus n'atteint pas leur cible. Une partie importante des pesticides répandus est dispersée dans l'atmosphère, soit lors de leur application, soit par évaporation ou par envol à partir des plantes ou des sols sur lesquels ils ont été répandus. Disséminés par le vent et parfois loin de leur lieu d'épandage, ils retombent avec les pluies directement sur les plans d'eau et sur les sols d'où ils sont ensuite drainés jusqu'aux milieux aquatiques par les eaux de pluie (ruissellement et infiltration). (A. K. Abdessalem, 2008).

Les pesticides sont ainsi aujourd'hui à l'origine d'une pollution diffuse qui contamine toutes les eaux continentales : cours d'eau, eaux souterraines et zones littorales.

Si les pesticides sont d'abord apparus bénéfiques, leurs effets secondaires nocifs ont été rapidement mis en évidence. Leur toxicité, liée à leur structure moléculaire, ne se limite pas en effet aux seules espèces que l'on souhaite éliminer. Ils sont notamment toxiques pour l'homme. Estimer les effets sur les écosystèmes d'une pollution liée aux pesticides s'avère difficile, car il existe un millier de familles de pesticides, soit des dizaines de milliers de principes actifs. Ils sont en outre utilisés à faibles doses et leurs comportements sont très divers. Leur impact dépend à la fois de leur mode d'action (certains sont beaucoup plus toxiques que d'autres), de leur persistance dans le temps (certains se dégradent beaucoup plus rapidement que d'autres) et de leurs sous-produits de dégradation lesquels sont parfois plus toxiques et se dégradent moins vite que la molécule mère. Leurs effets sur les êtres vivants sont, eux aussi, encore très mal connus. (A. K. Abdessalem, 2008).

La première partie de ce mémoire sera consacré à une étude bibliographique sur les pesticides ainsi que leurs catégories, leur impact sur l'environnement, leur toxicologie et les marchés. (A. K. Abdessalem, 2008), décrira le phénomène de partage, et les paramètres influençant sur la constante de partage, et enfin étudiera les méthodes expérimentales et les différentes techniques analytiques utilisées au cours de cette étude.

Dans la deuxième partie on présentera les résultats obtenus au cours de cette étude et une discussion de ces résultats.

Le mémoire contient aussi une introduction et une conclusion générale.

I. Généralités sur les pesticides :

I. 1. Définition des pesticides :

Un pesticide est une substance chimique utilisée pour lutter contre des organismes considérés comme nuisibles. C'est un terme générique qui rassemble les insecticides, les fongicides, les herbicides, les parasitocides. Ils s'attaquent respectivement aux insectes ravageurs, aux champignons, aux « mauvaises herbes » et aux vers parasites.

Le terme pesticide comprend non seulement les substances « phytosanitaires » ou « phytopharmaceutiques » utilisées en agriculture, sylviculture et horticulture, mais aussi les produits zoo sanitaires, les produits de traitements conservateurs des bois, et de nombreux pesticides à usage domestique : shampoing anti poux, boules antimites, poudres anti-fourmis, bombes insecticides contre les mouches, mites ou moustiques, colliers antipuces, diffuseurs intérieurs de pesticides, etc. (<https://fr.wikipedia.org/wiki/Pesticide>).

I. 2. Historique

La lutte chimique existe depuis des millénaires : l'usage du soufre remonte à la Grèce antique (1 000 ans av. J.-C.) et l'arsenic est recommandé par Pline l'Ancien, naturaliste romain, en tant qu'insecticide. Des plantes connues pour leurs propriétés toxiques ont été utilisées comme pesticides (par exemple les aconits, au Moyen Âge, contre les rongeurs). Des traités sur ces plantes ont été rédigés (ex. : traité des poisons de Maïmonide en 1135). Les produits arsenicaux ou à base de plomb (Arséniate de plomb) étaient utilisés au XVI^e siècle en Chine et en Europe.

Les propriétés insecticides du tabac étaient connus dès 1690. En Inde, les jardiniers utilisaient les racines de *Derris* et *Lonchocarpus* (roténone) comme insecticide. Leur usage s'est répandu en Europe vers 1900.

En 1807 Isaac-Bénédict Prévost promeut l'usage du sulfate de cuivre dans le traitement de la carie du blé. Guère suivies en France, ses préconisations sont adoptées rapidement en Suisse, en Grande-Bretagne et aux Pays-Bas.

La chimie minérale s'est développée au XIX^e siècle, fournissant de nombreux pesticides minéraux à base de sels de cuivre (encore bien utile en agriculture biologique). Les fongicides à base de sulfate de cuivre se répandent, en particulier la fameuse bouillie bordelaise (mélange de sulfate de cuivre et de chaux) pour lutter contre les invasions fongiques de la vigne et de la pomme de terre, non sans séquelles de pollution sur les sols (cuivre non dégradable).

Des sels de mercure sont employés à partir du début du XX^e siècle pour le traitement des semences. En raison de la toxicité du mercure, ils sont interdits dans les pays de l'OCDE depuis 1991 et dès 1982 pour certains pays d'Europe de l'Ouest. Leur usage perdure dans d'autres pays.

Épandeur adapté à l'épandage sur les vignes.

L'ère des pesticides de synthèse débute vraiment dans les années 1930, profitant du développement de la chimie organique de synthèse et de la recherche sur les armes chimiques durant la Première Guerre mondiale.

En 1874, Zeidler synthétise le DDT (dichlorodiphényltrichloroéthane), dont Muller en 1939 établit les propriétés insecticides. Le DDT est commercialisé dès 1943 et ouvre la voie à la famille des organochlorés. Le DDT a dominé le marché des insecticides jusqu'au début des années 1970 où il a été interdit en UE, notamment.

En 1944, l'herbicide 2,4-D, copié sur une hormone de croissance des plantes et encore fortement employé de nos jours, est synthétisé.

Affiche publicitaire américaine diffusée durant la Seconde Guerre mondiale par l' Office for Emergency Management intitulée « Tirez pour tuer ; Protégez votre jardin de la victoire ». (Office of War Information. Domestic Operations Branch. Bureau of Special Services), utilisée du 03 septembre 1943 au 15 septembre 1945.

La Seconde Guerre mondiale a généré, à travers les recherches engagées pour la mise au point de gaz de combat, la famille des organophosphorés qui, depuis 1945, a vu un développement considérable encore de mise aujourd'hui pour certains de ces produits, tel le malathion.

En 1950-1955 se développe aux États-Unis les herbicides de la famille des urées substituées (linuron, diuron), suivis peu après par les herbicides du groupe ammonium quaternaire et triazines.

Les fongicides du type benzimidazole et pyrimides datent de 1966, suivi par les fongicides imidazoliques et triazoliques dits fongicides IBS (inhibiteurs de la synthèse des stérols) qui représentent actuellement le plus gros marché des fongicides.

Dans les années 1970-80 apparaît une nouvelle classe d'insecticides, les pyréthrinoïdes qui dominent pour leur part le marché des insecticides.

Auparavant, la recherche de matières actives se faisait au hasard en soumettant de nombreux produits à des tests biologiques. Lorsqu'un produit était retenu pour ses qualités biocides, on cherchait à en améliorer l'efficacité à travers la synthèse d'analogues. Cette procédure a permis de développer les techniques de synthèse qui sont de mise aujourd'hui.

Désormais, l'accent est mis sur la compréhension des modes d'action et la recherche de cibles nouvelles. Connaissant les cibles, on peut alors établir des relations structure-activité pour aboutir à l'obtention de matières actives. Ceci est possible grâce au développement de la recherche fondamentale dans les domaines de la biologie et de la chimie et aux nouveaux outils fournis par la chimie quantique, les mathématiques et l'informatique qui permettent la modélisation de ces futures molécules.

Actuellement, on assiste à une consolidation du marché au niveau des familles les plus récemment découvertes avec la recherche de nouvelles propriétés. Dans le même temps, de nouvelles cibles physiologiques de l'animal ou du végétal sont explorées dans le but de développer des produits à modes d'action originaux, des produits issus de la biotechnologie ou des médiateurs chimiques. (<https://fr.wikipedia.org/wiki/Pesticide>).

I. 3. Catégories des pesticides

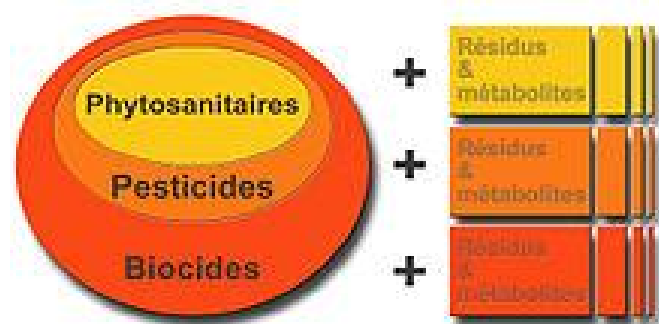


Figure 1 : Type de pesticide

Classification par grandes familles et sous-familles.

Les pesticides font partie des biocides.

Ils incluent les produits dits phytosanitaires ou phytopharmaceutiques (qui étymologiquement « soignent » les plantes : ce sont comme des médicaments pour les plantes en culture). En France, le ministère de l'Agriculture et de la Pêche et le ministère de l'Environnement (de l'Écologie, de l'Énergie, du Développement Durable et de l'Aménagement du Territoire) ont conjointement produit un document visant à mieux différencier les phytosanitaires au sein des biocides.

Ils incluent aussi des produits qui soignent les animaux ou l'homme (antiparasitaires externes ou internes par exemple). Ils peuvent désigner des molécules actives seules, ou des formulations associant plusieurs molécules ou des molécules actives et additifs (surfactants par exemple).

Chaque groupe chimique produit des métabolites au sein des organismes vivants ou des résidus en se dégradant spontanément. Ces résidus ou métabolites sont plus ou moins dégradables et susceptibles d'être retrouvés comme polluants de l'environnement ou contaminants de la nourriture ou de la boisson. (<https://fr.wikipedia.org/wiki/Pesticide>).

Les pesticides peuvent être regroupés selon différents axes : par type d'usage, par origine, par type d'activité, par groupe chimique, par mode d'action, etc. (<https://fr.wikipedia.org/wiki/Pesticide>).

I. 3. 1. Catégorie par origine

On distingue les pesticides organiques (contenant du carbone) et inorganiques (sans carbone autre que sous forme autre que carbonate ou cyanure). Parmi les pesticides organiques, une distinction est faite entre pesticides de synthèse (développés en laboratoire et produits en usine), pesticides naturels (d'origine animale, végétale ou microbienne) et micro-organismes. Les pesticides inorganiques sont des dérivés d'éléments minéraux (exemple : sulfate de cuivre).

I. 3. 2. Catégorie par niveaux de risque

Depuis 1975, l'Organisation mondiale de la santé propose une classification des pesticides par niveaux de risques.

I. 3. 3. Catégorie par usage

Ce regroupement s'intéresse à la cible que le pesticide est destiné à combattre. On recense ainsi :

- **les algicides**, utilisés contre les algues dans les lacs, canaux, piscines, réservoirs d'eau, etc. ;
- **les acaricides**, utilisés contre les acariens ;
- **les antimicrobiens et les bactéricides**, utilisés contre les bactéries ;
- **les corvicides ou corvifuges**, utilisés contre les corbeaux ;
- **les fongicides** pour tuer les champignons ou inhiber leur croissance (exemple, les QoI)
- **les herbicides, désherbants, phytocides ou débroussaillants** utilisés pour détruire les adventices (« mauvaises herbes »)
- **les insecticides**, utilisés contre insectes et autres arthropodes ;
- **les molluscicides**, qui tuent les limaces et les escargots (ou les éloignent dans le cas de répulsifs) ; dont les hélicides qui sont spécifiques des escargots ;
- **les nématocides**, utilisés contre les nématodes ;
- **les ovicides**, qui tuent les œufs d'insectes et d'acariens ;
- **les parasitocides**, utilisés contre les parasites ;
- **les piscicides**, utilisés contre les poissons ;
- **les rodenticides**, utilisés contre les rongeurs ;
- **les taupicides**, utilisés contre les taupes ;
- **les virucides**, terme commercial désignant des produits, solutions ou traitements censés « tuer » les virus ; ce terme est incorrect, puisqu'un virus, ne possédant pas de métabolisme interne, n'est pas considéré comme vivant au sens strict. Il peut cependant en effet être détruit ou neutralisé ;
- **les biopesticides**, divers types de pesticides dérivés de produits naturels.

Les catégories de produits suivants, sont plus spécifiquement et commercialement désigner comme « produits phytosanitaires », sont utilisées pour soigner ou prévenir les maladies des

végétaux. Ce ne sont donc pas tous des pesticides au sens strict (régulateurs hormonaux de croissance par exemple) (<https://fr.wikipedia.org/wiki/Pesticide>):

- **les anti-russetings** luttent contre la rugosité des pommes ;
- **les dessiccants et les défoliants** qui détruisent les feuillages des plantes ;
- **les répulsifs** luttent contre les insectes (moustiques), le gibier et les oiseaux ;
- **les régulateurs de croissance** sont utilisés pour la prévention de la croissance excessive d'une plante (lutte contre la verse chez le blé), les anti-germinants, les produits favorisant la résistance des plantes, le bouturage, la mise en fruit ;
- **les phéromones**, substances biochimiques qui attirent les insectes et perturbent leur comportement.

Autres produits :

- **les fumigants**, produisant des gaz ou vapeurs pour traiter bâtiments et sols contre divers bioagresseurs ;
- **les désinfectants**, pour traiter objets et matériel contre les microorganismes pathogènes ;
- les agents **antifouling**, utilisés contre les organismes qui s'attachent aux surfaces immergées, comme la coque des bateaux. (<https://fr.wikipedia.org/wiki/Pesticide>).

I. 4. Les Pesticides et l'Environnement :

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 utilisés 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. Au sein de cette classe des herbicides, plusieurs groupes de molécules sont distingués sur la base de leur séquence d'usage, correspondant par ailleurs bien souvent à une proximité chimique. Les plus critiques sont les herbicides de pré-levée, c'est-à-dire ceux utilisés sur sol nu pour empêcher ou limiter l'apparition d'adventices avant la germination des semences. Ces produits sont utilisés seuls (période hivernale) ou en mélange (printemps) suivant le démarrage des cultures mais aussi la croissance des plants. Ils sont donc particulièrement exposés à l'exportation de la parcelle nue vers le milieu naturel lors des pluies hivernales et surtout printanières, génératrices de crues.

Le devenir des herbicides dans le sol dépend fondamentalement de la capacité de transfert de chaque matière active. Elle conditionne leur mobilité, rétention et biodisponibilité. Mobile, la molécule pourra circuler entre les compartiments, eau, sol ou sédiment, air et biota. Sa capacité d'adsorption (sur un sol ou un sédiment) accroîtra sa rémanence et donc la pollution à long terme tant des sols que des eaux souterraines (Dur et al., 1998), limitant la portée des tentatives de dépollution en les renchérisant et en les rendant plus longues (Söderdren, 1984).

Suivant la méthode et les conditions extérieures, entre 0 et 23% de la dose mise au champ par aspersion est appliquée efficacement sur la plante traitée. La valeur communément admise est inférieure à 0,3% (Pimentel, 1995). Sur les 2,5 millions de tonnes épandues chaque année sur Terre, 2,49 millions sont donc « perdues », soit 0,399kg/hectare. (Gouy et al, 2001) fournissent une estimation plus récente évaluant à 5-10% les quantités absorbées par les plantes, soit une perte de 0,36kg/hectare.

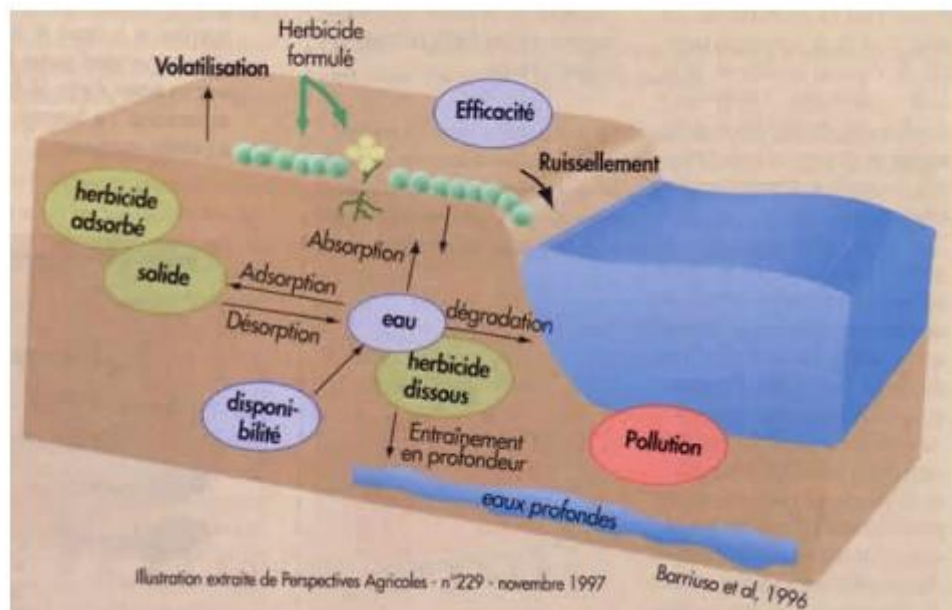


Figure 2: Processus et transfert des pesticides dans les sols (Barriuso et al., 1996)

I. 4.1. 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 faibles : 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 (J. N. Aubertot, J. M. Barbier, 2005)

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, J. M. Barbier, 2005).

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 du CORPEN15 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 (Les Produits phytosanitaires dans l'air :www.ecologie.gouv.fr/, 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 (Les Produits phytosanitaires dans l'air :www.ecologie.gouv.fr/, CORPEN – 2007)

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 liposolubles, 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 il est évident qu'un pesticide qui suit ce processus est, à exposition et toxicité égales, plus dangereux pour l'environnement qu'un autre produit qui ne s'accumule pas.

La toxicité, et notamment la toxicité chronique, se manifeste par des 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 *et al.*, 1992) 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).

I. 6. Le marché des pesticides :

Dans le monde

Il existe dans le Monde près de 100000 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 (N. A. Mokhtari, 2012).

➤ 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 pyréthrinoides, 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 million USD contre 49,4 millions USD en 2007 (N. A. Mokhtari, 2012).

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, telles 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 (A. Capry, 1999).

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 (http://www.techniques-ingenieur.fr/extraction_liquide-liquide.pdf), (A. Leo; et al; 2006).

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 (A. Leo; et al; 2006)

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 (A. Leo; et al; 2006)

En 1909, **Hertz** publia une formule reliant K au nombre d'extractions successives effectuées (A. Capry , 1999)

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

Avec :

X₀ : Quantité de la substance au départ

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 (A. Leo; et al; 2006), 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 (A. Capry , 1999), (A. Leo; et al; 2006) 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. 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 dissociant 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 dissociant 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 méthylénique (CH₂) incorporé dans la molécule. Un composé ramifié présente un K inférieur 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 (A. H. Seyd, 2013).

L'introduction d'un halogène au contraire favorise le passage en phase organique donc l'augmentation de K (A. H. Seyd, 2013).

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 quelque sorte l'énergie de liaisons 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 (A. H. Seyd, 2013).

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 :

Les modèles quantitatifs de relation structure-activité (modèles QSAR) sont des modèles de régression ou de classification utilisés dans les sciences et l'ingénierie chimique et biologique. Comme les autres modèles de régression, les modèles de régression QSAR associent un ensemble de variables «prédictives» (X) à la puissance de la variable réponse (Y), tandis que les modèles QSAR de classification associent les variables prédictives à une valeur catégorique de la variable réponse.

Dans la modélisation QSAR, les prédicateurs sont constitués de propriétés physico-chimiques ou de descripteurs moléculaires théoriques de produits chimiques; la variable de réponse QSAR pourrait être une activité biologique des produits chimiques. Les modèles QSAR résument d'abord une relation supposée entre les structures chimiques et l'activité biologique dans un ensemble de données de produits chimiques. Deuxièmement, les modèles QSAR prédisent les activités de nouveaux produits chimiques.

III.1. 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 d'un tel modèle est la suivante:

$$\text{Propriété} = f(\text{Descripteurs})$$

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 multi-linéaires (MLR) (D. Regadi, 2013) régressions aux moindres carrés partiels (PLS), arbres de décision, réseaux de neurones, algorithmes génétiques (D. Regadi, 2013)

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 les quelles la valeur expérimentale n'est pas disponible, voire 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.

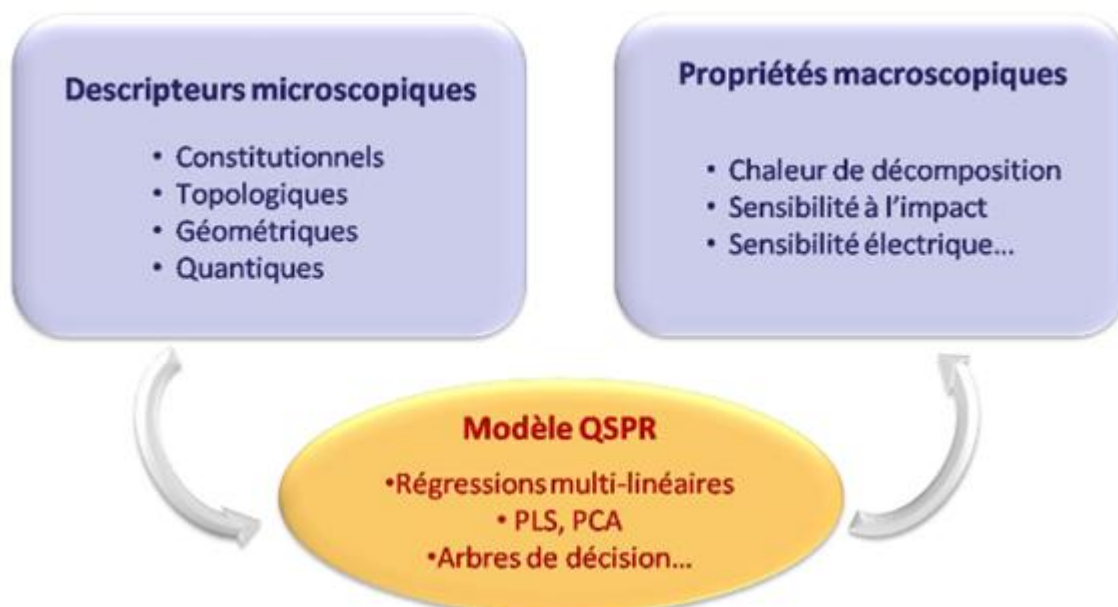


Figure 3 : Principe de la méthode QSPR

En pratique, le développement proprement dit d'un modèle débute par la collecte de données expérimentales fiables et en nombre le plus important possible. Il s'agit ensuite de développer

une série de descripteurs qui caractérisent les structures moléculaires électroniques et géométriques des composés de la base de données en vue de les relier à la propriété expérimentale étudiée. Des outils d'analyse de données sont alors employés pour aider à choisir les descripteurs adéquats et mettre en place le modèle proprement dit. Une fois développé, le modèle doit alors être validé en termes de corrélation (sur le jeu de données d'entraînement). Sa robustesse, c'est-à-dire l'influence des composés du jeu d'entraînement sur le modèle, est estimée par des méthodes de validation interne. Pour estimer son pouvoir prédictif, il est en suite nécessaire de disposer de données expérimentales supplémentaires afin de déterminer la capacité du modèle à prédire ces valeurs. Enfin, pour tout modèle, il est important de savoir pour quel type de molécules il est utilisable ou non, c'est-à-dire connaître son domaine d'applicabilité.

III.2. Importance des bases de données

Un modèle QSPR, de par sa construction, est très dépendant des données expérimentales de référence. Le choix de la base de données est donc un point critique de son développement. Dans la plupart des cas, les données expérimentales sont issues de la littérature.

Pour être de qualité, une base de données doit être composée de données expérimentales aussi fiables que possible, puisque les barres d'erreurs sur celles-ci se propageront dans le modèle final, étant donné que les paramètres de ce dernier sont ajustés par rapport à ces données. Il est donc important de choisir des données présentant des incertitudes faibles afin de limiter les barres d'erreur expérimentales. En effet, un modèle ne pourra être plus robuste statistiquement que les données théoriques à partir desquelles il a été développé: Approches de base, développement, validation et application des méthodes QSAR/QSPR.

III.3. Base de données

Le choix de la base de données expérimentale initiale est une étape critique pour le développement des modèles QSAR/QSPR.

Généralement, les composés testés ont deux origines possibles (dans la plupart des cas sont issues de la littérature), soit des produits de synthèse ou bien des produits d'extraction à partir de plantes. Quelle que soit son origine, il arrive qu'un échantillon ne soit pas pur mais corresponde à un mélange racémique. Le résultat du test d'un tel échantillon pose problème : il est impossible de savoir quelle est la contribution de chaque énantiomère dans l'activité

observée. Les structures dont la propriété étudiée est mesurée sur un mélange racémique ne peuvent pas être utilisées dans les études QSAR/QSPR (Damien Devault, 2007)

Pour être de qualité, une base de données doit être composée de données expérimentales fiables, puisque les barres d'erreurs sur celles-ci se propageront dans le modèle final. Il est donc important de choisir des données présentant de faibles incertitudes afin de limiter les barres d'erreur expérimentales.

D'autre part, l'homogénéité des données est fondamentale. Si l'on veut comparer l'activité/propriété d'une série de molécules, il faut s'assurer, si cela est possible, qu'elle est le résultat de leur interaction avec une seule et même cible et plus précisément avec le même site actif, et l'activité doit être mesurée par un seul et même test, avec des conditions expérimentales identiques pour chaque molécule.

En fin, la diversité des structures est un facteur important dans la qualité des modèles construits, elle définit l'espace chimique que l'analyse va couvrir.

III.4. 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 $i=1, \dots, n$:

$$y_i = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} + \xi \quad (2)$$

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 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 :

Les résidus (e) ont une espérance mathématique nulle :

$$E(e) = 0 \quad (3)$$

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

Les résidus sont indépendants entre eux :

$$E(e_i, e_j) = 0 \quad (4)$$

Leurs covariances sont nulles.

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 :

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 :

$$n-p-1 \geq 10 \quad (5)$$

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 à la condition* $FIV < 5$.

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

- Le coefficient de détermination multiple :

$$R^2 = 1 - \frac{SCE}{SCT} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (6)$$

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

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

$$EQMP = \sqrt{\frac{1}{n} \sum_1^n (y_i - \hat{y}_i)^2} = \sqrt{\frac{PRESS}{n}} \quad (7)$$

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}) :

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

$$EQMP_{ext} = \frac{\sum_{i=1}^{next} (y_i - \hat{y}_i)^2}{next} \quad (9)$$

La validation croisée par « leave-one-out » (LOO) 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 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, est une mesure de la dispersion des estimations. On l'utilise pour définir le coefficient de prédiction :

$$Q^2_{LOO} = \frac{SCT - PRESS}{SCT} \quad (10)$$

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 $Q^2_{LOO} > 0,5$ est considérée comme satisfaisante, et une valeur supérieure à 0,9 est excellente.

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.5. 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.

III.6. Régression Moindre Carré (PLS) :

Une alternative à la RCP consiste à utiliser la PLS (S. Wold *Chemometrics*, 1984), (F. Lindgren et al, 1994). La méthode PLS exprime une variable dépendante (y) en termes de combinaisons linéaires des variables indépendantes originales (x) selon :

$$y_1 = b_1t_1 + b_2t_2 + b_3t_3 + \dots + b_mt_m \quad (11)$$

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, \dots , 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.7. Calcul 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 6.03) 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. Les géométries ainsi optimisées ont été transférées dans le logiciel informatique DRAGON (R. Todeschini, et al, 2005) 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.

III.8. Les Types d'Apprentissages d'un Modèle QSPR :

Les techniques d'apprentissage 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.9. 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 (A.R.Leach, 2001); la comparaison se fait souvent à l'avantage des algorithmes génétiques (L. Chambers, 1995)

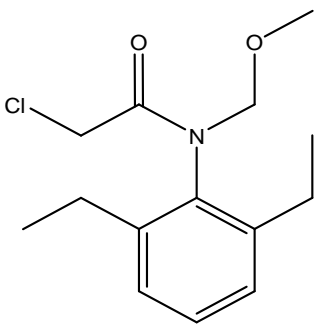
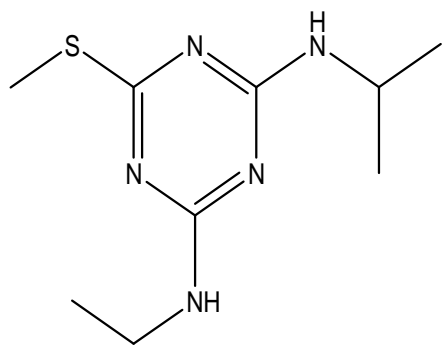
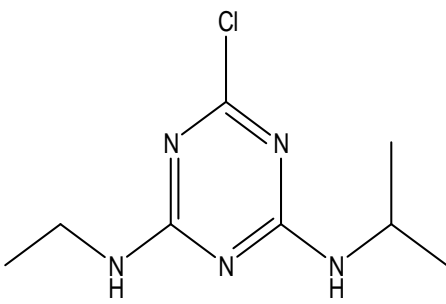
Dans un algorithme génétique adapté à l'optimisation, une solution potentielle est considérée comme un individu dans une population. La valeur de la fonction de coût associée à une solution mesure « l'adaptation » de l'individu associé à son environnement. Un algorithme génétique simule l'évolution,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

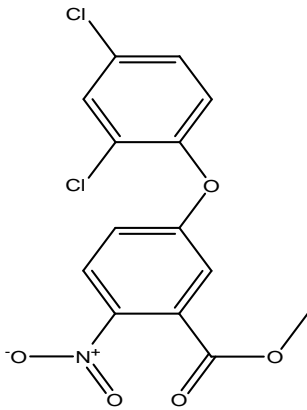
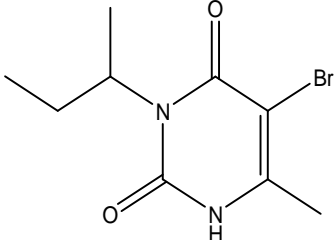
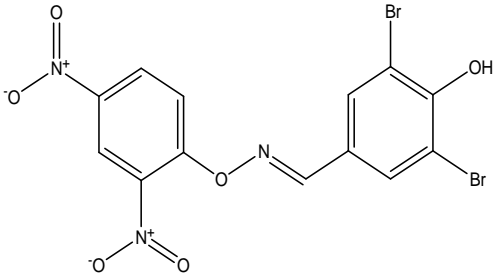
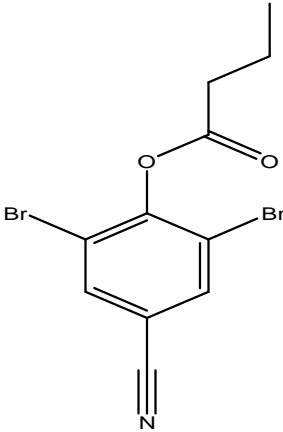
Dans le présent travail, la sélection des descripteurs a été réalisée par algorithme génétique, dans la version MOBY DIGS de Todeschini (R. Todeschini, 2004) en maximisant Q_{LOO}^2 .

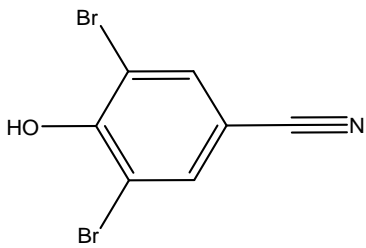
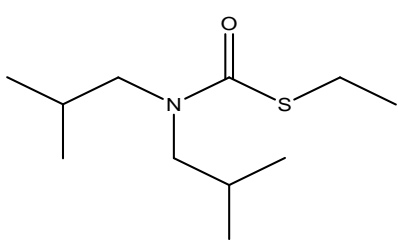
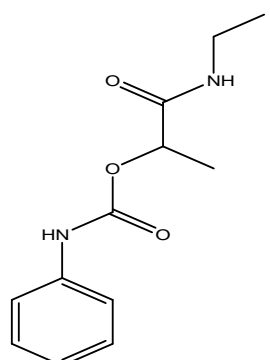
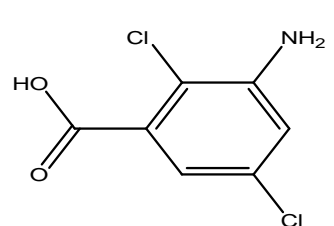
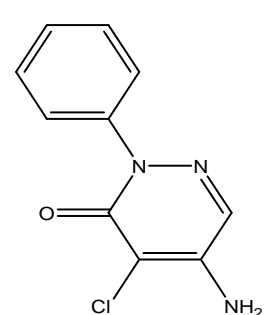
Dans cette partie expérimentale, nous allons déterminer la constante de partage octanol-eau d'une série de pesticides par modélisation d'un ensemble d'estimation. Le meilleur modèle obtenue sera validé par l'approche hybride algorithme génétique et régression multilinéaire GA/RLM.

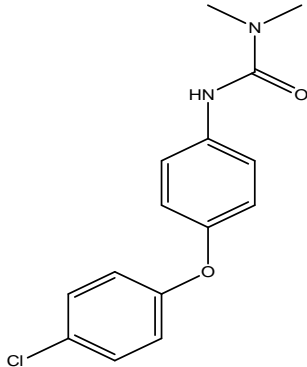
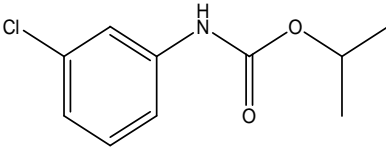
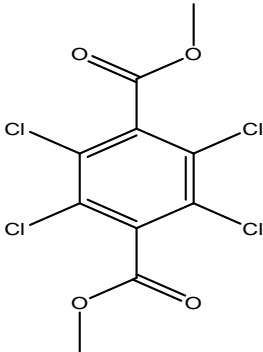
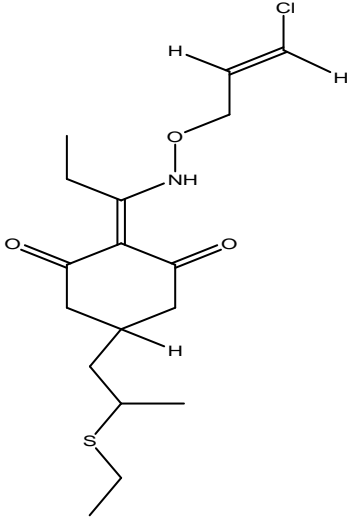
La liste des molécules étudiées dans notre travail présente dans le Tableau I, cette base de données contient 24 molécules, avec la propriété à étudier (la constante de partage n octanol-eau (K_{ow})) et les variables explicatives (descripteurs) sélectionnées par algorithme génétique.

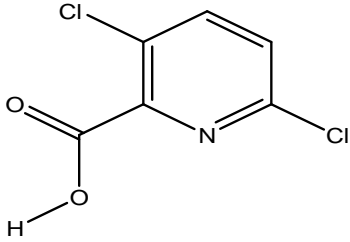
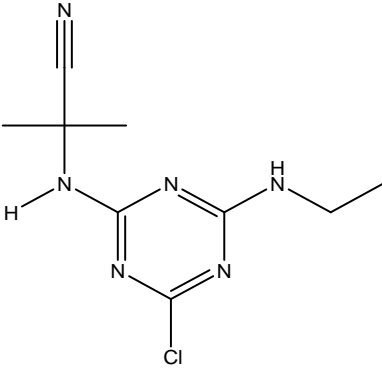
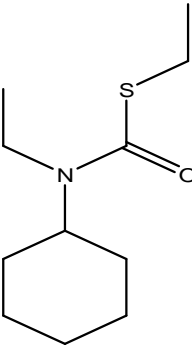
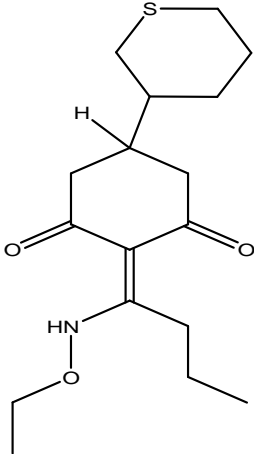
Tableau I : Structure des composés étudiés

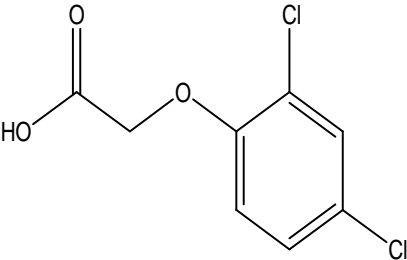
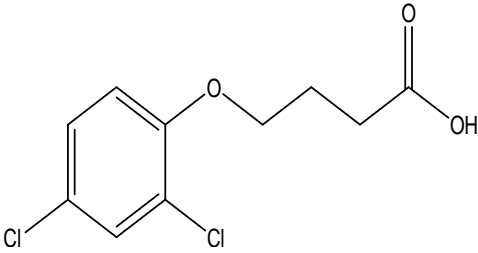
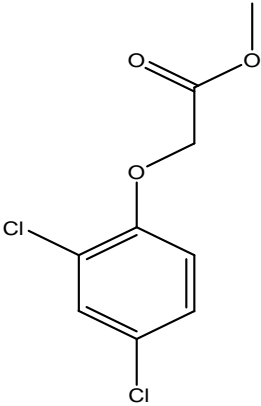
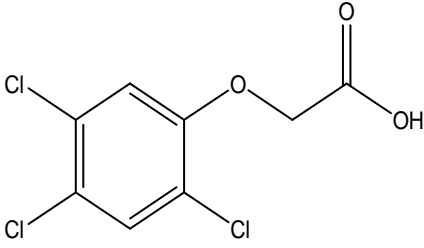
N°	Composés	Structure	log Cte Henry
01	alachlor		-2,6458
02	ametryn		-3,1586
03	atrazine		-3,3439

N°	Composés	Structure	log Cte Henry
04	bifenox		-3,7798
05	bromacil		-4,8507
06	bromofenoxim		-8,9586
07	bromoxynil butyrate		-1,752

N°	Composés	Structure	log Cte Henry
08	bromoxynil		-4,0599
09	butylate		0,4393
10	carbetamide		-4,8356
11	chloramben		-5,6716
12	chloridazon		-6,1784

N°	Composés	Structure	log Cte Henry
13	chloroxuron		-5,7986
14	chlorpropham		-2,5391
15	chlorthal dimethyl		-2,1655
16	clethodim		-5,9281

N°	Composés	Structure	log Cte Henry
17	clopyralid	 <chem>CC(=O)OC1=CC=C(Cl)N=C1Cl</chem>	-3,3018
18	cyanazine	 <chem>CCN1C=NC2=C1N=CN2C#NCl</chem>	-6,7258
19	cycloate	 <chem>CCN(CC)C(=O)SCC1CCCCC1</chem>	0,0863
20	cycloxydim	 <chem>CCN(CC)C(=O)C1CCCCC1C(C)CC</chem>	-6,0154

N°	Composés	Structure	log Cte Henry
21	2,4-D		-3,0301
22	2,4-DB		-3,6345
23	2,4-D-methyl		-0,5243
24	2,4,5-T		-3,1598

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 8.0). 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 (R.Todeschini *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 17.

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 Henry (log cte H).

N°	Descripteur	Classe	Signification
1	AAC	Information indices	mean information index on atomic composition
2	RDF010u	RDF descriptors	Radial Distribution Function - 1.0 / unweighted
3	Mor13u	3D-MoRSE descriptors	3D-MoRSE - signal 13 / unweighted

Les valeurs du logarithme de la constante de Henry (log cte H) et les valeurs des descripteurs sélectionnés par algorithme génétique sont présentées dans le tableau III au dessous.

Tableau III : Valeurs de log cte H , AAC, RDF010u, Mor 13u

Composés	log Cte Henry	AAC	RDF010u	Mor13u
Alachlor	-2.6458	1.518	5.081	1.1
Ametryn	-3.1586	1.574	5.697	0.578
Atrazine	-3.3439	1.632	4.876	0.396
bifenox	-3.7798	1.875	2.93	-0.006
bromacil	-4.8507	1.768	3.87	0.333
bromofenoxim	-8.9586	2.047	5.03	-0.975
butylate	0.4393	1.369	5.485	1.734
carbetamide	-4.8356	1.597	5.93	0.287
chloramben	-5.6716	2.013	3.582	-0.41
chloridazon	-6.1784	1.829	4.138	-0.332
chloroxuron	-5.7986	1.666	5.236	0.077
chlorpropham	-2.5391	1.691	4.198	0.266
chlorthal dimethyl	-2.1655	1.888	1.546	-0.281
clopyralid	-3.3018	2.074	1.66	-0.067
cyazifluoproc	-6.7258	1.681	6.495	0.568
cycloxydim	-6.0154	1.48	6.94	0.953
2,4-D-methyl	-0.5243	1.765	2.303	0.301
bromoxynil butyrate*	-1.752	1.821	2.419	0.18
Bromoxynil*	-4.0599	1.921	1.69	-0.571
Clethodim*	-5.9281	1.605	6.992	1.205
Cycloate*	0.0863	1.407	4.792	0.812
2,4-D*	-3.0301	1.813	2.409	0.013
2,4-DB*	-3.6345	1.716	3.357	0.276
2,4,5-T*	-3.1598	1.873	2.058	0.058

(*) Composé 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 IV

Tableau IV : Matrice de corrélation

	log cte H	AAC	RDF010u
AAC	-0,710 0,000		
RDF010u	0,284 0,136	0,117 0,546	
Mor13u	-0,342 0,069	0,144 0,455	0,389 0,037

Les trois descripteurs AAC, RDF010u, Mor13u (tableau IV) permettent de corrélérer log cte H des 17 composés des pesticides, la corrélation entre le log cte H et le AAC est trop forte par contre elle est faible avec le RDF010u et le Mor13u et malgré sa porte une certaine information complémentaire à la constante de Henry. Les trois descripteurs sont faiblement corrélés entre eux.

II- 2- Equation et analyse de régression :

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

$$\log \text{cte H} = 17.8(\pm 6.13) - 8.90 (\pm 3.07)\text{AAC} - 1.57(\pm 0.19)\text{RDF010u} + 1.42(\pm 0.81)\text{Mor13u} \quad (12)$$

Tableau V: Analyse de régression

Régresseur	Coef	SE Coef	T	P	VIF
Constante	17.846	6.130	2.91	0.012	
AAC	-8.903	3.072	-2.90	0.012	8.648
RDF010u	-1.5666	0.1986	-7.89	0.000	2.269
Mor13u	1.4151	0.8132	1.74	0.105	5.933

Les valeurs de T sont presque proche a part le RDF010u qui un peu éloigné se qui confirme que la contribution de se dernier et très importante. Les valeurs des VIF sont assez grande supérieur à 5, ce qui traduit la non indépendance des trois descripteurs intervenant dans le modèle.

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

Tableau VI : Les paramètres statistiques pour l'ensemble de calibration (17 composés) et celui de validation (7 composés).

Les paramètres statistiques	
F	37,12
R ²	89,55
R _{adj} ²	87,13
Q _{LOO} ²	82,65
Q _{ext} ²	80,01
S	0,845
EQMC	0,739
EQMP	0,952
EQMP _{ext}	0,896

Les paramètres statistiques reproduit dans le tableau ci-dessus montrent une bonne explication de la variabilité cte H, par les descripteurs choisis, de l'ordre de 87 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 petites et proches; en outre ce modèle est significatif avec une valeur du paramètre de Fisher : (F=37,12), Les statistiques

montrent aussi que le modèle a 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 : 0,2 pour les deux ensembles (calibration et validation) ; les valeurs des erreurs standardisées ($e_{i\text{std}}$) en fonction des valeurs de Leviers (h_i),

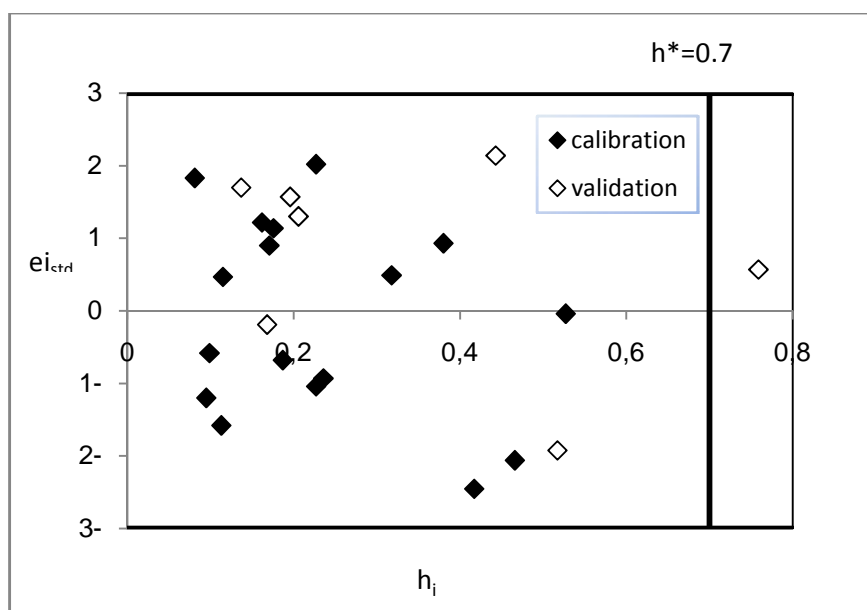


Figure 4: Diagramme de Williams

D'après le graphe on remarque que tous les erreurs standardisé sont comprise entre les limites ± 3 , donc une absence des points aberrant.

Le diagramme fait ressortir un point influent de l'ensemble de validation, qui renforce le modèle.

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 constante de Henry.

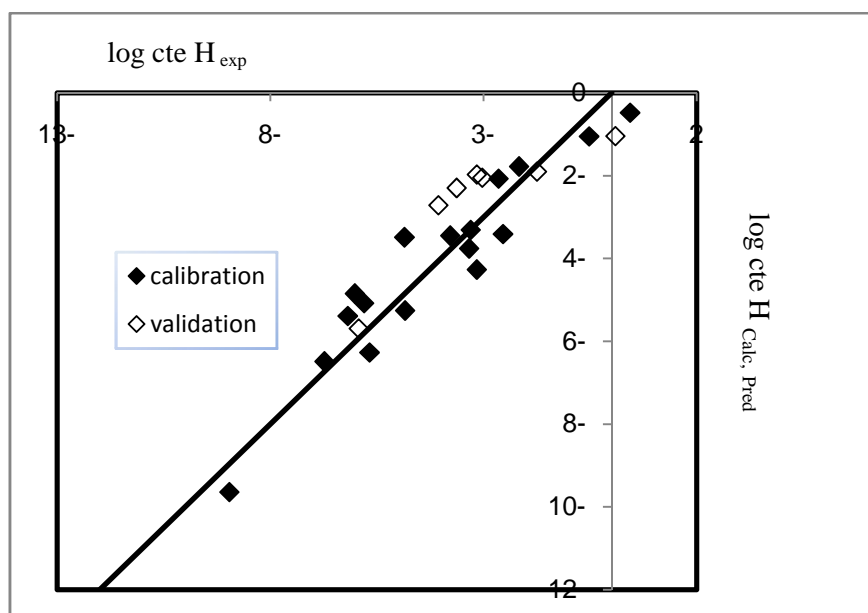


Figure 5 : Graphe des valeurs calculées, prédites en fonction des valeurs observées du $\log cte H$

D'après le graphe on remarque une certaine dispersion autour de la première bissectrice. Ceci est confirmé par la valeur 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 appliqués le test de randomisation de y (Figure 6), dans cette figure on à présenté le graphe des coefficients statistiques Q^2 et R^2 .

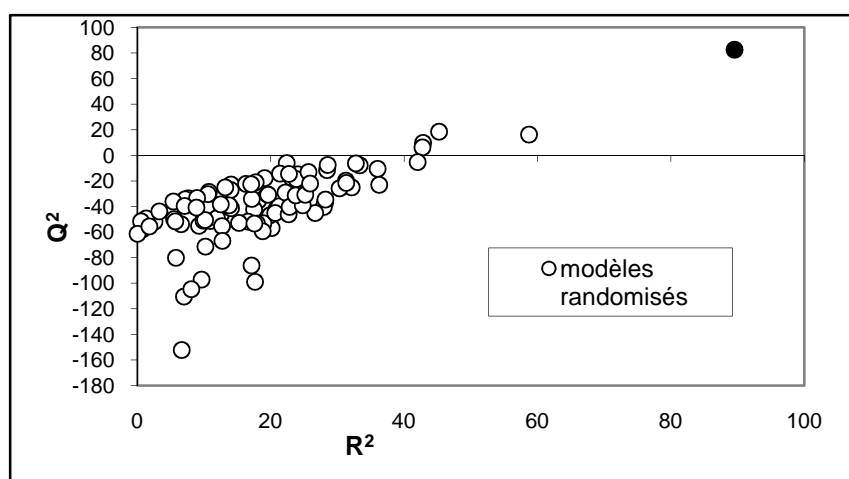


Figure 6 : Test de randomisation

Il est clair que les statistiques obtenues pour les cte H 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 :

Dans notre travail nous avons utilisé la méthodologie QSPR pour prédire la constante de Henry H d'une série de 24 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 constante d'Henry H et les variables explicatives (AAC, RDF010u, et Mor13u) elle est fondamentalement linéaire.

Références bibliographiques

- A. Capry, Importance de la lipophilie en modélisation moléculaire, LPTC, UFR de chimie, Université de Bordeaux I, France. (p 3-5. p 27), 1999.
- A. K. Abdesslem, Universités Paris-Est Et Tunis El Manar- Pour Obtenir Le Grade De Docteur par sujet : Dégradation des pesticides chlortoluron, carbofurane et bentazone en milieux aqueux par les procédés d'oxydation avancée, (p 2-5). Le 12 Décembre 2008.
<https://tel.archives-ouvertes.fr/tel-00470543/document.09/05/2018>
- 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 .
- A. R. Leach. Modeling. Principales and Application, 2nded Pearson PrinticeHall, London 2001.
- A.H. Seyd, Mémoire Présenté pour l'obtention du diplôme de MAGISTER, Simulation du rendement et du coefficient de distribution dans une extraction liquide-liquide.(p 12), 2007.
- A.R. Leach, V.L. Gillet An Introduction to Chemoinformatics : Revised Edition ; Springer, Dordrecht, 2007.
- D. Regadi, Développement de Modèle QSPR Pour la Prédiction des Propriétés Physiques des Quelque Composés Odorants, 2013.
- Damien Devault. Approche spatio-temporelle de la contamination par les herbicides de pré-levée du biotique de la Garonne Moyenne these Doctorat Toulouse, 2007.
- F. Lindgren, P. Gelada, S. Wold. J. Chemometrics 8, 349, 1994.
- Hyperchem™Release 6.03 for windows, Molecular Modeling System, 2000.
- J. N. Aubertot, J. M. Barbier Expertise scientifique collective INRA – Cemagref, 2005, Alain Carpentier, 2005.
- L. Chambers. Practical of Handbook of Gentic Algorithms. Lewis Publishing, 1995.
- L. Eriksson, J. Jaworska, A. P. Worth, D. Cronin, R. M. Mc Dowell, P. Gramatica Méthods for Relability and Uncertainty Assesement and for Applicability Evaluation of Classification and Regression Based QSARS. Environmental Health Perspectives (111), 1361-1375, 2003.
- les pesticide <https://fr.wikipedia.org/wiki/Pesticide>, Catégories des pesticides,15/04/2018

- les pesticide <https://fr.wikipedia.org/wiki/Pesticide>, Historique, 15/04/2018
- les pesticide <https://fr.wikipedia.org/wiki/Pesticide>, Définition, 15/04/2018
- Les Produits phytosanitaires dans l'air : origine, surveillance et recommandations pratiques en agriculture », CORPEN – 2007.
- N. A. Mokhtari, université d'oran Laboratoire De Synthèse Organique Appliquée (LSOA) par sujet : Identification et dosage des Pesticides dans l'Agriculture et les problèmes d'Environnement liés présentée.(p21.22), le 15 octobre 2012.
- R. Todeschini, B. Ballabio, V. Consonni, A. M. Pavan. Mobydigs Software for Multilinear Regression Analysis and Variable Subest Selection by Genetic Algorithm. Release 1.0 for Windows, Milano, 2004.
- R. Todeschini, V. Consonni, M. Pavan. DRAGON, Software for the Calculation of Molecular Descriptors. Release 5.3 for windows, Milano, 2005.
- R. Wehrens, H. Putter, L. M. C. Buvidens. Chemom. Int. Lab. Syst, (p,35-52,54), 2000.
- S. Wold Chemometrics : Mathematics and Statistics in Chemistry. Reidel, Dordrecht, The Netherlands, 1984.

ETUDE
BIBLIOGRAPHIQUE

APPLICATION

*CONCLUSION
GENERALE*

REFERENCES
BIBLIOGRAPHIQUES