



République Algérienne Démocratique Et Populaire  
Université Abbes Laghrour khenchela



Faculté des sciences et technologie  
Département de sciences de la matière  
Spécialité : chimie analytique

Intitulé:

**Prédiction de la solubilité aqueuse d'une série des  
composés organiques utilisant une régression  
linéaire multiple**

Présenté par :

- DJAGHROURI Loubna
- TOUMI Anouar

Diriger par :

Mme Bouakkadia. A

Devant le jury composé de :

<i>SAMAI Salima</i>	<i>MCB</i>	<i>U.A.L. Khenchela</i>	<i>Présidente</i>
<i>BOUAKKADIA Amel</i>	<i>MCA</i>	<i>U.A.L. Khenchela</i>	<i>Encadreur</i>
<i>KERTIOU Nour Eddine</i>	<i>MCB</i>	<i>U.A.L. Khenchela</i>	<i>Examineur</i>

Année universitaire :2022 / 2023

# Remerciements

Avant tout l'éloge, A Dieu qui nous a donné et accordé la force, le courage et la santé pour compléter ce travail.

Nous adressons aussi nos profonds remerciements à notre encadreur Mme Bouakkadia Amel pour avoir dirigé notre travail et pour le temps qui nous a consacré pour nous transmettre une partie de son savoir.

Nous voulons aussi remercier tous les membres de jury

Nous remercions aussi tous nos enseignants qui nous ont accompagnés dans notre parcours éducatif à l'université d'Abbes Laghrour Khenchela.

Et sans oublier de remercier tous ceux qui ont contribués de près ou de loin à l'élaboration de ce travail.

# Dédicaces

Je dédie ce modeste travail :

A La source d'affection éternelle, à celle qui dieu a fait le paradis sous ses pieds, à celle qui a sacrifié pour moi et pour mon éducation....ma très chère mère, j'espère qu'Allah la garde et la protège.

A ceux qui ont contribué à mon éducation, à ceux qui m'ont considéré leur fille....mon oncle , mes tantes.

A mon marie et mes enfants.

A ma belle-famille.

A tous mes amies de la promotion de master « Chimie Analytique ».

**Djaghrouri Loubna**

# Dédicaces

C'est Avec un énorme plaisir, un cœur ouvert et une joie immense, que Je dédie ce modeste travail à ma femme qui m'as tant soutenu et qui m'as incité à poursuivre mes études, ainsi que mes enfants qui m'ont encouragé et à mes défunts parents qui m'ont soutenu tout au long de leurs vies, et qui ont toujours voulu que j'aïlle très loin dans mes études, et à toute personnes qui m'ont encouragé et participés à réussir ce travail.

A vous tous je vous dis merci de votre soutien.

**Toumi Anouar**

**Résumé :**

Le logiciel Minitab utilise la méthode QSPR pour relier linéairement la solubilité d'un ensemble homogène de certains composés contaminants à des descripteurs moléculaires théoriques sélectionnés

L'ensemble de données a été divisé au hasard en deux sous-ensembles, un ensemble de 24 composés pour la construction du modèle et un ensemble de 10 composés pour la validation.

Différentes statistiques établies pour les ensembles d'étalonnage et de validation (coefficients de détermination et de prédiction :  $R^2$  ;  $Q^2$  et écart-type  $s$ ) seront confirmées.

Les domaines d'application sont discutés à l'aide de diagrammes de Williams et de la qualité de l'ajustement.

**Abstract:**

Minitab software uses the QSPR method to linearly relate the solubility of a homogeneous set of certain contaminant compounds to selected theoretical molecular descriptors.

The data set was randomly divided into two subsets, a set of 24 compounds for model building and a set of 10 compounds for validation.

Different statistics established for the calibration and validation sets (coefficients of determination and prediction:  $R^2$ ;  $Q^2$  and standard deviation  $s$ ) will be confirmed.

Application areas are discussed using Williams's plots and goodness of fit.

## ملخص:

يستخدم برنامج Minitab طريقة QSPR لربط قابلية ذوبان مجموعة متجانسة من بعض المركبات الملوثة للواصفات الجزيئية النظرية المختارة خطياً  
تم تقسيم مجموعة البيانات بشكل عشوائي إلى مجموعتين فرعيتين، مجموعة من 24 مركباً  
لبناء النموذج ومجموعة من 10 مركبات للتحقق من صحتها.  
سيتم تأكيد الإحصائيات المختلفة التي تم إنشاؤها لمجموعات المعايرة والتحقق من الصحة  
(معاملات التحديد والتنبؤ:  $R^2$  ؛  $Q^2$  والانحراف المعياري).  
تتم مناقشة مجالات التطبيق باستخدام مخططات ويليامز وجودة الملاءمة.

## Sommaire :

I- Introduction générale :	2
I- Généralités sur les PCBs :	4
I.1 Définition :	4
I.2 Propriétés physico-chimiques des PCBs :	4
I.2.1 Solubilité des PCBs :	5
I.2.2 Pression de vapeur :	5
I.2.3 Coefficient de partage :	6
II- Usage des PCBs :	7
III- Sources de contamination :	8
III.1 Historiques :	8
III.2 Sources de contamination actuelles :	9
IV- Normes des PCB :	10
V- Effets des PCBs sur l'environnement :	10
V.1 Effets sur l'eau :	10
V.2 Effets sur l'air :	11
V.3 Effets sur les animaux :	12
V.4 Effets sur les êtres humains :	13
VI- Traitement des PCB :	15
I- Relation Quantitative Structures Propriétés :	17
I.1 Principe et Théories :	17
I.2 Méthodologie QSPR :	17
I.3 Importance de la base de données :	18
II- Les descripteurs moléculaires :	18
II.1 Descripteurs constitutionnels :	19
II.2 Descripteurs topologiques :	19
II.3 Descripteurs géométriques :	19
II.5 Descripteurs thermodynamiques :	20
III- Validation des modèles QSPR :	20
III.1 Réduction de nombre de variables :	20
III.2 Validation interne :	20

<b>III.3 Validation externe :</b> .....	21
<b>IV- Statistique et méthode d'analyse de données :</b> .....	21
<b>IV.1 Introduction :</b> .....	21
<b>IV.2 La régression linéaire simple :</b> .....	22
<b>IV.3 La régression linéaire multiple :</b> .....	22
<b>IV.3.1 Validation du modèle :</b> .....	22
<b>I- Calcul du modèle :</b> .....	29
<b>I.1 Choix du modèle :</b> .....	29
<b>I.2 Choix d'une famille de descripteurs :</b> .....	29
<b>II- Développement du modèle :</b> .....	31
<b>II.1 Matrice de corrélation :</b> .....	31
<b>II.2 Equation et analyse de régression :</b> .....	32
<b>II.4 Qualité de l'ajustement :</b> .....	33
<b>II.5 Test de randomisation :</b> .....	34
<b>Conclusion générale :</b> .....	35

**Symboles et abréviations :**

<b>AESA :</b>	Agence Européenne de la Sécurité Aérienne
<b>ANOVA :</b>	Analysis Of Variance
<b>CITEPA :</b>	Centre Interprofessionnel Technique d'Etude de la Pollution Atmosphérique
<b>DRIRE :</b>	Direction Régionale de l'Industrie , de la Recherche et de l'Environnement
<b>EQMC :</b>	Erreur Quadratique Moyen Calculée sur l'ensemble de calibrage
<b>EQMP :</b>	Erreur Quadratique Moyen de Prédiction
<b>EQMP<sub>ext</sub> :</b>	Erreur Quadratique Moyen de Prédiction sur l'ensemble de validation
<b>F :</b>	Test de Fisher
<b>HOMO :</b>	Highest Occupied Molecular Orbital
<b>ICPE :</b>	Installation Classées Pour l'Environnement
<b>IUPAC :</b>	International Union of Pure and Applied Chemistry
<b>LMO :</b>	Leave Many Out
<b>LOO :</b>	Leave One Out
<b>LUMO :</b>	Lowest Unoccupied Molecular Orbital
<b>PACA :</b>	Provence Alpes Côte d'Azur
<b>PCB :</b>	Polychlorobiphényle
<b>POPs :</b>	Polychlorobiphényle Organiques Poluants
<b>Q<sup>2</sup><sub>ext</sub> :</b>	Coefficients de prédiction sur l'ensemble de validation
<b>Q<sup>2</sup><sub>Lo0</sub> :</b>	Coefficients de prédiction sur l'ensemble de calibrage
<b>QSPR :</b>	Relation Quantitative Structures Propriétés
<b>R<sup>2</sup> :</b>	Coefficients de détermination
<b>R<sup>2</sup><sub>adj</sub> :</b>	Coefficients de détermination d'ajustement
<b>S :</b>	L'écart type
<b>THT :</b>	Très Haute Tension

## Liste des tableaux :

<b>Tableau I :</b> Solubilité des PCB pour chaque groupe d'isomères (données uniques sont des moyennes).....	5
<b>Tableau II :</b> Pressions de vapeurs des PCB pour chaque groupe d'isomères (données uniques sont des moyennes).....	6
<b>Tableau III :</b> Coefficient de partage des PCB pour chaque groupe d'isomères (données uniques sont des moyennes).....	7
<b>Tableau IV :</b> Quelques emplois de PCB depuis leur mise sur le marché. ....	9
<b>Tableau V :</b> Apport en PCB totaux imputable à l'eau de consommation pour différentes classes d'âge.....	11
<b>Tableau VI :</b> Comparaison des apports en PCB par l'air ambiant en fonction des classes d'âge et habitations.....	12
<b>Tableau VII :</b> Résultats des principales études concernant l'absorption des PCB chez l'homme et l'animal.....	14
<b>Tableau VIII :</b> Liste des molécules étudiées et leurs solubilités. ....	24
<b>Tableau IX :</b> Descripteurs moléculaires intervenant dans la modélisation de la solubilité ....	30
<b>Tableau X :</b> Valeurs de log S, X <sub>2v</sub> , Mor <sub>12v</sub> .....	30
<b>Tableau XI :</b> Matrice de corrélation.....	31
<b>Tableau XII :</b> Analyse de régression .....	32
<b>Tableau XIII :</b> Les paramètres statistiques pour l'ensemble de calibration (24 composés) et celui de validation (10 composés) .....	32

**Liste des figures :**

<b>Figure 1 :</b> Représentation de la structure générale d'une molécule de PCB ( $x + y = 10$ ) .....	4
<b>Figure 2 :</b> Diagramme de Williams .....	33
<b>Figure 3 :</b> Qualité de l'ajustement.....	34
<b>Figure 4:</b> Test de randomisation.....	34

# **Introduction générale**

### I- Introduction générale :

Les bis phényles poly chlorés (PCB) ont été découverts au XIXe siècle. Compte tenu de leurs propriétés physico-chimiques, d'isolation électrique et de propriétés thermiques ininflammables, ils sont largement utilisés industriellement comme fluides caloporteurs pour l'isolation électrique et le refroidissement des transformateurs, condensateurs, etc.

En conséquence, les PCB ont connu un développement industriel considérable entre 1930 et 1980. Principaux pays producteurs : Autriche, Chine, Tchécoslovaquie, France, Allemagne, Italie, Japon, Russie, Espagne, Royaume-Uni, États-Unis... Les bis phényles poly chlorés sont des polluants organiques persistants (POP), c'est-à-dire des substances chimiques qui sont persistantes et s'accumulent dans les tissus biologiques des animaux et des humains, des écosystèmes terrestres et aquatiques.

En fait, ils ont tous en commun une stabilité chimique et physique élevée et une faible biodégradabilité, ce qui explique pourquoi ils s'accumulent dans les organismes qui se trouvent successivement au bout de la chaîne alimentaire où se trouvent les humains. A ce titre, ils sont omniprésents dans la chaîne alimentaire ; ce phénomène de bioaccumulation dans les tissus biologiques a été mis en évidence en 1966 par des chimistes suédois. Cette découverte soulève les premières interrogations et inquiétudes quant à l'éventuel impact de ces molécules sur la santé humaine.

Les pouvoirs publics sont de plus en plus conscients de ce problème et des risques liés à l'exposition aux PCB. Ainsi, la France, comme les États-Unis, a pris un certain nombre de mesures pour limiter l'exposition (rapport 2008, PCB Programme Steering Committee) : • Interdiction des PCB présents dans les encres, les adhésifs, les additifs et certaines huiles depuis 1979. En 1987, la vente, l'acquisition ou la commercialisation d'équipements contenant des PCB a été interdite • En 2003, le plan national de décontamination et d'élimination des équipements contenant des PCB (élimination des équipements contenant des PCB en concentration supérieure à 500mg/Kg avant le 31 décembre 2010) Cependant, ces efforts se sont avérés insuffisants. En fait, l'UE a récemment abaissé la concentration maximale autorisée de PCB dans les poissons destinés à la consommation humaine.

Ce durcissement des normes sanitaires pourrait conduire à des interdictions locales de pêche et/ou de commercialisation de poisson de consommation, notamment dans le Rhône. C'est pourquoi le gouvernement a lancé fin 2008 un plan d'action interministériel pour améliorer la connaissance de ces molécules, notamment le comportement des PCB dans

'environnement. Parmi les mesures phares du programme figure un programme d'échantillonnage à grande échelle sur l'ensemble de la France visant à collecter des données qui permettront de mieux identifier les voies d'exposition de l'environnement et des populations aux PCB et de quantifier les niveaux de ces expositions. Industrie (carte d'exposition aux PCB). La collecte de données devrait permettre de modéliser les transferts entre les sédiments et les poissons pour mieux déterminer l'exposition par l'alimentation.

Ainsi, l'objectif principal de ce travail est étudié.

**Ce mémoire est constitué de deux parties :**

- A la suite d'une introduction générale, la première partie est une étude bibliographique qui consiste des généralités sur les PCB et aussi une étude consacrée à différentes méthodes de traitement QSPR.
- La deuxième partie présente la partie application
- Le mémoire est enfin complet par une conclusion générale.

# **Partie bibliographique**

# Chapitre 01

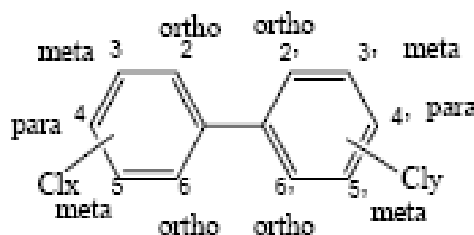
## I- Généralités sur les PCB :

### I.1 Définition :

Le terme PCB est un terme générique qui regroupe un groupe de produits chimiques définis comme des hydrocarbures halogénés de haut poids moléculaire [3].

Les biphényles polychlorés (PCB) sont des composés aromatiques organochlorés synthétisés à partir du bi phényle comme matière première, ils sont synthétisés sous forme de mélange et la formule chimique est  $C_{12}H_{(10-n)} Cl_n (1 \leq n \leq 10)$ . Ils forment une famille de composés ayant la même structure chimique (chaque noyau benzénique peut contenir jusqu'à 5 atomes de chlore à la place des atomes d'hydrogène). Il y a donc 10 degrés de chloration, du monochlorobiphényle au décachlorobiphényle [4]. Selon le nombre et la position des atomes de chlore, il existe en théorie 209 congénères. Par conséquent, chaque PCB est généralement défini par un nombre compris entre 1 et 209 selon la nomenclature IUPAC.

En effet, les contraintes thermodynamiques, la chimie de la réaction de chloration radicalaire du biphényle, et les contraintes de configuration stérique ont permis la synthèse d'environ 150 congénères différents [5].



**Figure 1 :** Représentation de la structure générale d'une molécule de PCB ( $x + y = 10$ )

### I.2 Propriétés physico-chimiques des PCB :

Les propriétés physicochimiques des PCB dépendent du degré de substitution du chlore ainsi que de la position des biphényles. Ces derniers peuvent être divisés en deux grandes catégories : les biens à l'origine d'un usage industriel et des propriétés qui expliquent et modélisent leur comportement dans l'environnement, à savoir :

- D'une part, la solubilité et les coefficients de partage octanol/eau, qui permettent de révéler l'affinité des différents congénères pour les phases polaires ou non polaires, exprimant leur capacité à s'accumuler dans les graisses

- La pression de vapeur saturante et la constante de Henry, en revanche, peuvent indiquer la capacité du congénère à se volatiliser, à passer à l'état gazeux, et ainsi caractériser le degré d'exposition « à l'air »

Les PCB ont des propriétés qui affectent leurs activités aux niveaux environnemental et biologique et expliquent leur toxicité pour l'homme, à savoir une stabilité chimique élevée, une faible solubilité dans l'eau, une forte lipophile et une tendance à la bioaccumulation.

### I.2.1 Solubilité des PCB :

Selon leur teneur en chlore, les PCB commerciaux se présentent sous forme de liquides plus ou moins visqueux ou de produits résineux. Ils sont incolores ou jaune pâle et ont une odeur aromatique caractéristique. Ils ont une très faible solubilité dans l'eau (0,007 à 5,9 mg/L), en particulier pour la plupart des composés chlorés. Les PCB sont également insolubles dans la glycérine et les glycols, mais solubles dans les huiles et la plupart des solvants organiques.

**Tableau I :** Solubilité des PCB pour chaque groupe d'isomères (données uniques sont des moyennes)[6].

Groupe de congénères	Solubilité à 25°C (g/m <sup>3</sup> )
<b>Monochlorobiphényle</b>	1,21 - 5,5
<b>Dichlorobiphényle</b>	0,06 - 2
<b>Trichlorobiphényle</b>	0,015 – 0,4
<b>Tétrachlorobiphényle</b>	0,0043 – 0,01
<b>Pentachlorobiphényle</b>	0,004 – 0,02
<b>Hexachlorobiphényle</b>	0,0004 – 0,0007
<b>Heptachlorobiphényle</b>	0,000045
<b>Octachlorobiphényle</b>	0,0002 – 0,0003
<b>Nonachlorobiphényle</b>	0,00018 – 0,0012
<b>Décachlorobiphényle</b>	0,000001

### I.2.2 Pression de vapeur :

La pression de vapeur des chlorobiphényles doit dépendre du nombre d'atomes de chlore présents dans l'isomère. La pression de vapeur des PCB en général est comprise entre  $10^{-3}$  et  $10^{-6}$  mm Hg, mais augmente jusqu'à  $10^{-1}$  mm Hg à plus de 100 °C. Ainsi, les PCB exposés à l'air sont capables de pénétrer dans l'atmosphère à des températures normales à des taux

modérés. On a estimé que la perte totale de PCB dans l'atmosphère est d'environ  $1,5$  à  $2,5 \times 10^3$  tonnes / an

Le faible solubilité dans l'eau et la faible pression de vapeur des PCB, associées aux processus de transport dans l'air, l'eau et les sédiments, signifient qu'ils sont facilement transportés des sites de contamination locaux ou régionaux vers des régions éloignées.

**Tableau II :** Pressions de vapeurs des PCB pour chaque groupe d'isomères (données uniques sont des moyennes)[6].

Groupe de congénères	Pression de vapeur (Pa) à 25°C
<b>Monochlorobiphényle</b>	0,9-2,5
<b>Dichlorobiphényle</b>	0,008-0,6
<b>Trichlorobiphényle</b>	0,003-0,22
<b>Tétrachlorobiphényle</b>	0,002
<b>Pentachlorobiphényle</b>	0,0023-0,051
<b>Hexachlorobiphényle</b>	0,0007-0,012
<b>Heptachlorobiphényle</b>	$1,3 \cdot 10^{-4}$
<b>Octachlorobiphényle</b>	$2,8 \cdot 10^{-5}$
<b>Nonachlorobiphényle</b>	$6,3 \cdot 10^{-6}$
<b>Décachlorobiphényle</b>	$1,4 \cdot 10^{-6}$

### I.2.3 Coefficient de partage :

Le coefficient de partage n-octanol/eau (log) est un paramètre utile pour l'évaluation du devenir et de l'impact dans l'environnement des contaminants traces xénobiotiques

Le coefficient de partage n-octanol/eau (log) est une propriété physico-chimique importante adaptée à la caractérisation de la lipophilie des composés tels que les PCB.

**Tableau III :** Coefficient de partage des PCB pour chaque groupe d'isomères (données uniques sont des moyennes) [6].

Groupe de congénères	Coefficient de partage octanol /eau (logKow)
<b>Monochlorobiphényle</b>	4,3-4,6
<b>Dichlorobiphényle</b>	4,9-5,3
<b>Trichlorobiphényle</b>	5,5-5,9
<b>Tétrachlorobiphényle</b>	5,6-6,5
<b>Pentachlorobiphényle</b>	6,2-6,5
<b>Hexachlorobiphényle</b>	6,7-7,3
<b>Heptachlorobiphényle</b>	6,7-7
<b>Octachlorobiphényle</b>	7,1
<b>Nonachlorobiphényle</b>	7,2-8,16
<b>Décachlorobiphényle</b>	8,26

## II- Usage des PCB :

Les PCB sont utilisées comme isolants électriques presque ininflammables et pour leurs excellentes caractéristiques diélectriques et de conduction thermique, les PCB ont été massivement utilisées des années 1930 aux années 1970 dans :

- Des transformateurs électriques .
- Des condensateurs .
- Des sectionneurs de puissance .
- Des isolateurs, dans les environnements à très haute tension (THT), ils étaient alors mélangés à des chlorobenzènes (également toxiques).
- Des fluides caloporteurs, par exemple dans les environnements à risque d'incendie, dont les navires transportant des carburants (pétrole, gaz).
- Des fluides hydrauliques en contexte à risque ou de contraintes thermiques (mines...).
- Des moteurs de pompe .
- Des fours à micro-ondes .
- Comme additifs d'huiles minérales ou de produits de soudures .
- Dans certains adhésifs .
- Dans des peintures .

- Et jusque dans des papiers autocopiants [1, 2].

La toxicité aigüe des PCB a été constatée dès les années 1930 chez les ouvriers fabriquant et manipulant les produits.

### **III-Sources de contamination :**

On a deux grandes catégories de sources de contamination : les sources historiques dues à la production de PCB de 1929 à 1987 et les sources actuelles de contamination qui permettent aujourd'hui encore la libération des PCB dans l'environnement.

#### **III.1 Historiques :**

Les PCB sont conçus pour trois types d'applications, qui peuvent être classées selon leur impact sur l'environnement :

- **Systèmes fermés :** Ces systèmes sont généralement de gros équipements (transformateurs et condensateurs) qui sont contrôlés pour éviter les fuites. Ils contiennent suffisamment de PCB pour justifier leur recyclage.
- **Systèmes fermés non contrôlés :** Ces systèmes impliquent de petits condensateurs tels que l'huile hydraulique et les fluides caloporteurs utilisés dans l'industrie minière. Des fuites sont possibles et le recyclage des PCB n'est pas pratique.
- **Systèmes ouverts non maîtrisés :** Ces systèmes correspondent à des applications dispersées comme les additifs stabilisants ou retardateurs de flamme dans les encres, peintures, plastiques. Ces systèmes ouverts utilisent des PCB qui sont directement introduits et dispersés dans l'environnement, les rendant irrécupérables [7].

**Tableau IV** : Quelques emplois de PCB depuis leur mise sur le marché [8].

Système clos	Système Clos non contrôlables	Système ouvert non contrôlables
fluides diélectriques : transformateurs, condensateurs de puissance ou pour l'électroménager et l'éclairage fours à micro-ondes air conditionné moteurs électriques Ballasts pour lampes électriques Electro aimants	fluides hydrauliques fluides de transfert de chaleurs Commutateurs Régulateurs de voltage Disjoncteurs Câbles électriques fluides industriels : pompes à vide, huiles hydrauliques, huiles de coupe des métaux	Lubrifiants Cires Retardateurs de flamme Matériaux isolants Pesticides Teintures Asphaltes Condensats de gazoducs additifs ignifugeants : matière plastique, agents plastifiants et/ou adhésifs

### III.2 Sources de contamination actuelles :

La France interdit la production et l'utilisation des PCB depuis 1987. Toutefois, l'utilisation d'équipements mis en service avant 1987 est toujours autorisée (Décret n087-59 du 2 février 1987). Si la production de PCB est interdite maintenant, l'environnement ne sera pas protégé. Le but de cette section est d'examiner les diverses sources de pollution environnementale actuelle. La contamination actuelle est en grande partie le résultat de l'utilisation des derniers équipements en PCB, des traitements de décontamination et des actes de négligence et de vandalisme. En effet, l'utilisation des équipements PCB les plus récents concerne principalement les anciens transformateurs, qui sont aujourd'hui à l'origine d'accidents tels que fuites, explosions, incendies et décharges accidentelles. Aujourd'hui, le problème n'est plus la production de PCB, mais la décontamination des PCB. Par conséquent, les activités de décontamination et de traitement des déchets génèrent des émissions dans l'environnement.

Selon le CITEPA, l'incinération des déchets industriels et ménagers sera à l'origine de l'essentiel des émissions dans l'atmosphère, soit 83 % des émissions de PCB en France. Les « broyeurs » de déchets dangereux sont à l'origine de rejets de PCB dans l'environnement en décontaminant les appareils électriques contaminés (condensateurs, transformateurs) puis en les incinérant. Les PCB sont également présents dans les effluents des stations d'épuration et

les boues d'épuration, contribuant à la diffusion de ces composés dans l'environnement par rejet direct ou diffusion. Les personnes qui vandalisent, abandonnent les transformateurs ou recherchent du métal sont également des sources de pollution. Par exemple, les récentes explosions de mines de cuivre ont incité certains criminels à récolter le métal rouge en ouvrant des transformateurs et en se débarrassant frénétiquement des BPC contenus dans l'environnement.

D'après les informations recueillies auprès de la DRIRE Rhône-Alpes et de la PACA, les lâchers sauvages ou accidentels actuels sont largement supérieurs aux lâchers autorisés. En effet, les friches industrielles, les anciennes usines, les anciens supermarchés, et même les anciennes décharges, tant qu'il n'y aura pas de surveillance, tant qu'il y aura des transformateurs ou d'autres équipements, constitueront un réservoir de pollution, qui peut être facilement ouvert et détruit par des personnes cherchant pour les métaux [9].

Par ailleurs, certaines activités qui échappent parfois à la réglementation des ICPE peuvent également être source de contamination, notamment les ICPE pour la ferraille et le recyclage des métaux. Les usines telles que l'usine Arkema de Saint-Auban qui produit des composés chlorés qui se dégradent facilement en PCB peuvent être une source de rejets de PCB dans l'environnement extérieur.

Les rejets industriels sont aujourd'hui réglementés et sous contrôle mais ceci n'empêche pas la pollution de l'environnement.

Enfin, d'après le Citepa, le secteur résidentiel participerait à 17% des émissions de PCB dans l'atmosphère en France par la combustion de bois, la consommation de tabac, l'usage de produits domestiques etc.

#### **IV- Normes des PCB :**

En Europe, la Dose Tolérable hebdomadaire (Tolérable Weekly Intake ou TWI) a été fixée à 14 pg TEQ/kg pc. Selon l'AESA, en 2010, entre 1 % et 52,9 % de la population étaient en surpoids selon le groupe étudié [12]. En Europe, une directive interdit l'utilisation de produits alimentaires pour animaux contenant des substances indésirables au-delà d'un certain niveau.

#### **V- Effets des PCB sur l'environnement :**

##### **V.1 Effets sur l'eau :**

Dans les lacs et les rivières, les PCB se fixent aux sédiments, où ils peuvent rester enfouis pendant de longues périodes avant d'être finalement rejetés dans l'eau ou dans

l'atmosphère. Les PCB en suspension dans l'air peuvent atteindre le sol par la pluie et la neige, ou seulement lorsque les particules se déposent sous l'effet de la gravité terrestre.

Les concentrations de PCB dans l'eau de mer à proximité des zones industrielles sont souvent élevées. La production de PCB est limitée depuis les années 1970 et les concentrations dans les sédiments fluviaux et les sédiments de poissons frais ont progressivement diminué.

**Tableau V :** Apport en PCB totaux imputable à l'eau de consommation pour différentes classes d'âge [10].

Classes d'âge	Quantité d'eau consommée (L/jour)	Apports en PCB totaux par l'eau potable (ng/jour)
0 à 6 mois	0,75	0,0023 – 0,23
6 mois à 4 ans	0,8	0,0024 – 0,24
5 à 11 ans	0,9	0,0027 – 0,27
12 à 19 ans	1,3	0,0039 – 0,39
20 ans et plus	1,5	0,0045 – 0,45

## V.2 Effets sur l'air :

Dans l'air, les PCB se décomposent sous l'action indirecte de la lumière solaire. Selon le type de PCB, le temps nécessaire pour décomposer la moitié de la quantité qui existait à l'origine varie de quelques jours à quelques mois. Dans l'eau, les PCB se dissolvent plus lentement, peut-être en raison de la lumière du soleil ou de l'action microbienne. Ces organismes jouent un rôle important dans la décomposition des PCB dans le sol et les sédiments. À l'extérieur, les concentrations de PCB dans l'air étaient plus faibles dans les régions rurales et éloignées que dans les régions urbaines et industrielles. En règle générale, la concentration de PCB dans l'air intérieur est au moins dix fois plus élevée que dans l'air extérieur.

**Tableau VI :** Comparaison des apports en PCB par l'air ambiant en fonction des classes d'âge et habitations [11].

Classes d'âge	Taux d'inhalation (m <sup>3</sup> / jour)	Apports en PCB totaux par l'air ambiant (ng/jour)	
		Zone industrialisée	Zones urbaines ou rurales
0 à 6 mois	2,1	1,5 – 22,7	0,2 – 0,9
6 mois à 4 ans	9,3	6,5 – 100,4	0,7 – 4,0
5 à 11 ans	14,5	10,2 – 156,6	1,2 – 6,2
12 à 19 ans	15,8	11,1 – 170,6	1,3 – 6,8
20 ans et plus	15,8	11,1 – 170,6	1,3 – 6,8

- Les concentrations de PCB dans l'air ambiant mesurées dans une zone à proximité des usines de General Motors, de Reynolds Metals et de Domtar paper mill variaient de 0,7 à 10,8 ng/ m<sup>3</sup> de décembre 1993 à novembre 1994
- Les concentrations de PCB dans l'air ambiant variaient de 0,08 à 0,43 ng/m<sup>3</sup>

Ce tableau illustre l'impact des zones industrialisées sur l'apport total de PCB dans l'air ambiant pour les populations vivant à proximité d'activités industrielles. Ce groupe peut avoir jusqu'à dix fois la consommation de ceux qui vivent dans des zones moins exposées. Outre les effets des activités industrielles, certaines habitations construites ou utilisant des matériaux contenant des PCB peuvent être une source d'apport quotidien de ces polluants.

### V.3 Effets sur les animaux :

Les PCB peuvent s'accumuler chez les animaux au fil du temps tout au long de la chaîne alimentaire. Les PCB se trouvent dans les tissus adipeux des animaux terrestres et aquatiques, en particulier ceux qui se trouvent au sommet de la chaîne alimentaire. Par conséquent, les humains peuvent également accumuler des PCB à partir des aliments qu'ils consomment. Certaines espèces animales, notamment les oiseaux, les poissons, les mammifères, les insectes et autres invertébrés, peuvent éliminer ou transformer certains PCB dans leur organisme. Chez les animaux, une seule exposition à de fortes doses de PCB peut causer de la diarrhée, des difficultés respiratoires, de la déshydratation, une diminution

de la sensibilité à la douleur et même le coma. On a observé que les PCB affectent principalement les poumons, l'estomac et le pancréas.

Prendre de faibles niveaux de PCB sur une période de temps peut interférer avec le fonctionnement normal du foie et de la glande thyroïde et, à long terme, entraîner un cancer du foie. Des effets sur la fertilité des femelles, les organes reproducteurs et l'activité hormonale ont également été observés chez des animaux de laboratoire exposés aux PCB par l'alimentation. La progéniture d'animaux exposés aux PCB dans leur alimentation pendant la gestation et la lactation peut également être affectée, nombre d'entre eux présentant des difficultés d'apprentissage et de comportement. Chez les jeunes, l'exposition aux BPC peut aussi avoir des effets sur le système immunitaire et le développement de certains organes comme le foie, la thyroïde et les reins.

Les effets observés des PCB sur le système immunitaire comprennent une diminution de la production d'anticorps, une sensibilité accrue aux maladies et une diminution du poids du thymus. Les animaux adultes semblent moins sensibles aux effets des PCB que les fœtus. Les PCB se comportent différemment selon leur structure chimique. Certains PCB se comportent comme des dioxines et peuvent augmenter le risque de cancer. D'autres peuvent affecter le développement du système nerveux lorsque les expositions sont élevées.

#### **V.4 Effets sur les êtres humains :**

Les humains peuvent absorber les PCB en mangeant ou en buvant des aliments contaminés et, dans une moindre mesure, en respirant de l'air contaminé ou à travers la peau. Une fois absorbés, les PCB traversent les membranes cellulaires et pénètrent dans les vaisseaux sanguins et le système lymphatique. Les niveaux les plus élevés de PCB se trouvaient généralement dans le foie, le tissu adipeux, le cerveau et la peau. Ils sont également présents dans le sang. Chez les femmes enceintes ou jeunes mères, on a observé que les PCB passent dans le sang du cordon ombilical, le placenta et le lait maternel. Chez les humains et les animaux, les PCB se transforment facilement en substances aux structures chimiques modifiées, difficiles à éliminer, mais qui seront retenues et accumulées dans les tissus et les fluides corporels.

Les PCB peuvent également être transformés en substances qui se combinent avec d'autres pour ensuite être éliminées de l'organisme dans les urines et les selles.

**Tableau VII** : Résultats des principales études concernant l'absorption des PCB chez l'homme et l'animal

Facteurs jouant un rôle sur l'adsorption		Homme	Animal
<b>Voie digestive</b>	<p>nombre de chlores sur les molécules (+ il y a de chlores et plus le taux d'absorption est important)</p> <p>contenu du tube digestif concentrations PCB dans le sang et les aliments car le transfert se fait de manière passive donc dépend du gradient de concentration</p>	<p><b>Abraham et al, 1994</b> : étude sur des bébés de 1 à 5 mois, taux d'absorption par téter de lait d'environ 98% de 3 congénères (PCB 138,153 et 180)</p> <p><b>Dahl et al, 1995</b> : fractions d'absorption moyenne par allaitement de 56 congénères varient de 0,9 à 1 pour des bébés de 1,3 et 6 mois</p> <p><b>Moser et MC Lachlan, 2001</b> : bilan massique entre PCB ingérés et dosés dans les fèces chez 5 adultes ayant accepté une alimentation enrichie en PCB indique une fraction d'absorption &gt;80%</p>	
<b>Voie respiratoire</b>	<p>Nombre de chlores : moins il y en a et plus ils sont volatils</p> <p><b>Mandalakis et Stephanou, 2002</b> : congénères avec 2-3 chlores représentent plus de 50% des PCB inhalés</p> <p>Congénères les plus chlorés adsorbés aux particules aériennes qui diminuent la fraction inhalée PCB sous forme gazeuse peu hydrosolubles donc doivent atteindre la région alvéolaire pour être absorbés</p>	<p>Pas de données permettant de déterminer la fraction d'absorption de PCB inhalés</p> <p>Par analogie avec les études existantes sur d'autres particules semi-volatiles possibilité d'extrapoler les données aux PCB :</p> <ul style="list-style-type: none"> <li>-70% des PCB sous forme gazeuse sont inhalés</li> <li>-&lt; 40% pour PCB adsorbés aux particules</li> </ul>	

<b>Voie cutanée</b>	<p><b>Erickson, 2001</b> : près de 99% des molécules de PCB auxquelles les individus sont exposés par la peau sont adsorbées à des poussières de sol qui diminuent grandement leur potentiel d'absorption cutanée</p> <p><b>Kissel et Fenske, 2000</b> : Les quantités de contaminants absorbés par la peau humaine sont fortement dépendantes du temps de contact</p>	<p>Aucune donnée expérimentale chez l'humain ne permet de déterminer la fraction d'absorption des PCB par la peau.</p> <p><b>Wester et al, 1990, 1993</b> : 2 études in-vitro avec Arcelor 1242 ou 1254 marqués au 14C indiquent une absorption voisine de 45% lorsque les PCB sont appliqués dans l'eau alors qu'elle est inférieure ou égale à 10% en cas de solubilisation dans l'huile minérale</p>	<p><b>Attention</b> : les animaux absorbent plus les substances chimiques que les humains entre autre à cause de leur grande pilosité</p> <p><b>Wester et al, 1983</b> : étude in vivo : cobayes taux d'absorption varient de 33 à 56% / singetaux d'absorption varient de 15 à 34%</p>
---------------------	--	---	---

## VI- Traitement des PCB :

Deux actions distinctes doivent être entreprises, sous réserve d'approbation :

- Récupération de matériaux :

L'équipement, les outils et les matériaux susceptibles d'avoir été en contact avec des PCB et des PCT sont décontaminés (ou décontaminés) par nettoyage au solvant dans un autoclave sous vide après la mise hors service de l'équipement.

Nous découvrons :

Une opération d'appoint consistant à remplacer l'huile contaminée par de l'huile neuve (décontamination par fluide de remplacement) est un exemple de méthode de décontamination.

- Dés halogénéation (traitement chimique) :
- Des réactions chimiques sont utilisées pour éliminer les PCB.
- Procédé APROCHIM :
- Grâce à la désorption sous vide poussé, la toxicité et le volume de déchets peuvent être considérablement réduits Remarque : Toutes ces méthodes doivent garantir que la teneur en PCB de l'huile est inférieure à 50 % après six mois de fonctionnement.

- Traitement de l'énergie :
- Les PCB sont éliminés par incinération, dans des autoclaves de décontamination, dans des installations classées et agréées par la législation. Entre autres entreprises, Tridi incinère les déchets contenant des PCB.

**Autres traitements :**

- Les PCB peuvent être traités physico chimiquement dans des unités de déchloration.
- Le Groupe Chimirec a développé un procédé appelé « APROCHIM » pour la déchloration des déchets contenant des PCB

# Chapitre 02

## Introduction :

Ce n'est qu'en 1964 que le modèle de "contribution de groupe" a été développé, qui, avec les travaux de Hansch [13], a constitué le début de la modélisation QSPR. Depuis, l'essor de nouvelles techniques de modélisation par apprentissage, d'abord linéaire puis non linéaire, a permis la mise en œuvre de nombreuses méthodes ; elles reposent principalement sur la recherche de la relation entre un ensemble de nombres réels de descripteurs moléculaires, et de relation entre eux. L'utilisation de la méthode QSPR n'a cessé de progresser ces dernières années. Il est même devenu incontournable en chimie en général.

Ces informations peuvent alors être utilisées pour prédire les propriétés physicochimiques de nouveaux composés ainsi que pour concevoir de nouvelles structures.

## I- Relation Quantitative Structures Propriétés :

### I.1 Principe et Théories :

Le principe de la méthode QSPR est de réaliser une relation mathématique liant des descripteurs moléculaires (compositionnels, topologiques, géométriques, spectroscopiques, quantique, thermodynamique ...) avec des propriétés macroscopiquement observables, par exemple des propriétés physico-chimiques, ou la solubilité, ou : Pour une série de composés similaires utilisant une méthode d'analyse des données statistiques.

La forme générale d'un modèle QSPR est :

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

Ainsi, le but de cette méthode est d'analyser des données structurales afin de détecter des déterminants des propriétés mesurées. Pour cela, différents types d'outils sont disponibles : régression linéaire simple [14, 15].

### I.2 Méthodologie QSPR :

En pratique, le développement d'un modèle commence par la collecte d'autant de données expérimentales fiables que possible. L'étape suivante consiste à développer une série de descripteurs décrivant les structures moléculaires électroniques et géométriques des composés de la base de données afin de les relier aux propriétés expérimentales étudiées. Des outils d'analyse de données sont ensuite utilisés pour aider à sélectionner les descripteurs appropriés et à mettre en œuvre le modèle. Pour s'assurer que le modèle QSPR développé est fiable, il doit ensuite être validé par rapport à la corrélation (sur l'ensemble de données d'apprentissage). Pour estimer son pouvoir prédictif, des données expérimentales supplémentaires sont nécessaires pour déterminer la capacité du modèle à prédire ces

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

### **I.3 Importance de la base de données :**

Les modèles QSPR reposent sur des données expérimentales de référence, et le choix de la base de données est un point clé dans le développement de ces modèles. Dans la plupart des cas, les données expérimentales proviennent de la littérature et pour obtenir une bonne qualité, la base de données doit être constituée de données expérimentales aussi fiables que possible, car les barres d'erreurs de ces données seront propagées dans le modèle final au fur et à mesure que les paramètres de ce dernier sont ajustés en fonction de ces données. Par conséquent, il est très important de choisir des données à faible incertitude pour limiter les barres d'erreur expérimentales. De plus, les données doivent être obtenues selon un protocole expérimental unique. En effet, les conditions expérimentales ont une forte influence sur les valeurs obtenues. La définition des propriétés en fonction des conditions expérimentales est également un point important de cette méthode

## **II- Les descripteurs moléculaires :**

Au cours des dernières décennies, de nombreuses recherches ont été menées pour trouver la meilleure façon de représenter l'information contenue dans les structures moléculaires, et les structures elles-mêmes, dans un ensemble de nombres réels appelés descripteurs moléculaires, qui visent à décrire la structure avec nombres pour une molécule, une fois ces nombres disponibles, les outils de modélisation classiques peuvent être utilisés pour établir des relations entre ces nombres et les propriétés moléculaires. Ainsi, ces descripteurs numériques encodent des informations chimiques sous forme de vecteurs de nombres réels. Il existe aujourd'hui plus de 6 000 types [16], qui quantifient les caractéristiques physico-chimiques ou structurales d'une molécule. Ils peuvent être obtenus de manière empirique ou non, et ils permettent des prédictions sans avoir à synthétiser des molécules.

Nous décrirons brièvement les descripteurs moléculaires les plus courants, en commençant par les descripteurs les plus simples, qui nécessitent peu de connaissances sur la structure moléculaire. Nous verrons ensuite comment les progrès de la modélisation moléculaire ont

permis d'accéder à la structure 3D d'une molécule et de calculer des descripteurs à partir de cette structure.

### II.1 Descripteurs constitutionnels :

Les descripteurs constitutionnels sont directement liés à la formule brute de la molécule, à l'aide de la composition moléculaire, c'est-à-dire les atomes qui le constituent, Il s'agit de :

- La masse molaire
- Les nombres d'atomes (C, H, O, S, N, F, Cl, Br, I, P...).
- Les nombres de groupes fonctionnels (NH<sub>2</sub>, COOH, OH...).
- Les nombres de liaisons (simples, doubles, aromatiques...).
- Les nombres de cycles (aromatiques ou non).

Ces descripteurs sont très utilisés du fait de leur extrême simplicité non seulement d'un point de vue conceptuel mais surtout calculatoire.

### II.2 Descripteurs topologiques :

Les descripteurs topologiques, décrivent la connectivité atomique dans la molécule, ils sont obtenus à partir de la structure 2D de la molécule, et donnent des informations sur sa taille, sa forme globale et ses ramifications.

### II.3 Descripteurs géométriques :

Les descripteurs géométriques d'une molécule sont évalués à partir des positions relatives de ses atomes dans l'espace, et décrivent des caractéristiques plus complexes; leurs calculs nécessitent de connaître, la géométrie 3D de la molécule, par modélisation moléculaire empirique, Ces descripteurs s'avèrent donc relativement coûteux en temps de calcul, mais apportent davantage d'informations, et sont nécessaires à la modélisation de propriétés qui dépendent de la structure 3D. On distingue plusieurs descripteurs importants, le volume moléculaire, la surface accessible au solvant, le moment d'inertie. Les distances, les angles de liaisons ou angles dièdres dans la molécule.

### II.4 Descripteurs quantiques :

Les descripteurs de chimie quantique donnent des informations importantes pour la molécule. Ces descripteurs permettent de quantifier différents types d'interactions inter- et intramoléculaires, de grande influence sur des propriétés physicochimiques de molécules. Le calcul de la plupart de ces descripteurs nécessite la recherche de la géométrie pour laquelle l'énergie est minimale, et fait appel aux approches de chimie quantique. On cite toutes les données énergétiques, vibrationnelles et orbitales du système par exemple, les énergies de

la plus haute orbitale moléculaire occupée HOMO et de la plus basse vacante LUMO, le moment dipolaire, la polarisabilité, le potentiel d'ionisation.

### II.5 Descripteurs thermodynamiques :

Les descripteurs thermodynamiques sont calculés sur la base de la fonction de partition totale  $Q$  de la molécule [17, 18]. La fonction de partition commande la façon avec laquelle l'énergie d'un système de molécules est répartie par miles individus moléculaires. Sa valeur dépend du poids moléculaire, de la température, du volume moléculaire, des distances internucléaires, des mouvements moléculaires et des forces inter moléculaires. La fonction de partition est le point le plus commode entre les propriétés microscopiques des molécules individuelles (niveaux d'énergie, moments d'inertie) avec les propriétés macroscopiques (chaleur spécifique, entropie).

## III- Validation des modèles QSPR :

### III.1 Réduction de nombre de variables :

Dans un calcul quanto-chimique un grand nombre de descripteurs sont collectés pour la modélisation d'une grandeur donnée, car les facteurs déterminants du processus étudié ne sont a priori pas connus. Cependant, les descripteurs envisagés n'ont pas tous une influence significative sur la grandeur modélisée, et les variables ne sont pas toujours mutuellement indépendantes. De plus, le nombre de descripteurs, c'est-à-dire la dimension du vecteur d'entrée, détermine la dimension du vecteur des paramètres à ajuster. Si cette dimension est trop importante par rapport au nombre d'exemples de la base d'apprentissage, le modèle risque d'être sur ajusté à ces exemples, et incapable de prédire la grandeur modélisée sur de nouvelles observations. Il est donc nécessaire de réduire la dimensionnalité des variables d'entrée, plusieurs approches sont possibles pour résoudre ce problème.

- Réduire la dimension de l'espace des entrées.
- Remplacer les variables corrélées par de nouvelles variables synthétiques, obtenues à partir de leurs combinaisons.
- Sélectionner les variables les plus pertinentes.

### III.2 Validation interne :

La technique la plus employée pour déterminer la stabilité du modèle prédictif est de tester l'influence de chaque échantillon sur le modèle final. Pour ce faire, on emploie une technique de validation croisée (cross validation ou CV). Ce processus consiste à extraire un certain nombre  $k$  de molécules du jeu initial à  $N$  molécules et à construire un nouveau modèle avec les  $N-k$  molécules restantes à l'aide des descripteurs choisis (seules les

constantes de la régression changent), ce nouveau modèle est alors utilisé pour la phase de prédiction sur les molécules retirées. Ce processus est ensuite réitéré pour retirer et prédire les valeurs de toutes les molécules du jeu d'entraînement, en fonction du nombre de molécules retirées à chaque itération, on parlera de Leave-One-Out (LOO) ou de Leave-Many-Out (LMO) [19]. Selon qu'une ou plusieurs molécules est (sont) retirée(s). D'une manière générale, ces techniques de validation interne permettent l'évaluation de la robustesse du modèle, autrement dit la stabilité des paramètres du modèle QSPR vis-à-vis des molécules du jeu d'entraînement

Cela dit, elles ne permettent en aucun cas de démontrer le pouvoir prédictif des modèles

### III.3 Validation externe :

Afin de tester de manière fiable le pouvoir prédictif, l'utilisation d'un jeu de validation externe, non employé pour le développement du modèle, est nécessaire, pour que le jeu de données initial soit suffisamment large, cette validation est caractérisée par les paramètres  $R^2(\text{test})$  ;  $R^2$  Par conséquent, d'autres paramètres doivent être vérifiés pour cet objectif. Ces paramètres sont connus sous le nom «critères de validation externe» ou souvent appelés «critères de Tropsha » (Tropshacriteria) [20, 21].

$R^2$  Coefficient de corrélation pour les molécules de la série de test.

$R^2$  coefficient de corrélation entre les valeurs prédites et expérimentales pour la série de test.

## IV- Statistique et méthode d'analyse de données :

La notion de régression est fondamentale dans toutes les sciences appliquées puisqu'elle consiste à analyser une relation entre deux ou plusieurs variables quantitatives et à l'exploiter pour estimer la valeur inconnue de l'une à l'aide de la valeur connue de l'autre. Elle est couramment utilisée dans les domaines de la chimie et de la biologie, Nous formalisons ici la démarche utilisée dans ce chapitre pour calculer l'équation de régression linéaire simple et multiple [22, 23].

### IV.1 Introduction :

La méthode de régression linéaire est classée parmi les méthodes d'analyses multivariées qui traitent des données quantitatives. Cette méthode de calcul date de la fin du 18<sup>me</sup> siècle. Elle est essentiellement, due à A. M. Legendre 1805 et C. F. Gauss 1809 [24, 25]. Le mot régression vient, de la fin du 19<sup>me</sup> siècle, et est dû à F. Galton 1886 [26]. Galton a introduit le mot régression pour expliquer la diminution progressive des écarts par rapport à la moyenne, d'une génération à la suivante, dans des problèmes d'hérédité, Le

terme régression a ensuite été utilisé d'une manière tout à fait générale, L'analyse de la régression est une méthode statistique qui permet d'étudier le type de relation pouvant exister entre une ou plusieurs variables explicatives (notée  $X_i$ ) et une variable à expliquer (notée  $Y$ ). En d'autres termes, l'analyse de la régression permet d'étudier les variations de la variable dépendante en fonction des variations connues des variables indépendantes.

#### **IV.2 La régression linéaire simple :**

La régression linéaire dite simple si elle permet de prédire les valeurs d'une variable (expliquée ( $Y$ )) à partir des valeurs prises par une autre variable (explicative ( $X$ )). Lorsque cette dépendance est exacte, la liaison entre les deux variables est «fonctionnelle » : à chaque valeur de  $X$  correspond une et une seule valeur possible de  $Y$  [27, 28].

#### **Modélisation par régression linéaire simple :**

Le modèle de régression est simplement une équation censée représenter cette relation entre les deux variables. Il s'écrit :

$$Y = f(x) + \varepsilon \dots\dots\dots (1)$$

La variable  $Y$  est donc supposée approximativement égale à une fonction  $f(x)$  ,

Le terme  $\varepsilon$  caractérisant la marge d'erreur ou d'imprécision du modèle.

#### **IV.3 La régression linéaire multiple :**

##### **Modélisation par régression linéaire multiple :**

Le modèle de régression multiple est une généralisation du modèle de régression simple lorsque les variables explicatives sont en nombre fini [29, 30].

##### **IV.3.1 Validation du modèle :**

###### **IV.3.1.1 L'analyse de la variance :**

L'analyse de la variance (terme souvent abrégé par le terme anglais (ANOVA) : (Analysis OF Variance)) est un test statistique permettant de vérifier que plusieurs échantillons sont issus d'une même population.

Ce test s'applique lorsque l'on mesure une ou plusieurs variables explicatives catégorielles qui ont de l'influence sur la distribution d'une variable continue à expliquer. Nous présentons dans le tableau suivant l'analyse de variance.

IV.3.1.2 Coefficient de Régression  $R^2$  :

$$R^2 = \frac{\|\hat{Y} - \bar{y}\|^2}{\|Y - \bar{y}\|^2} \dots \dots \dots (2)$$

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(Y_i - \bar{y})^2} = \frac{b_1 \sum(x_i - \bar{x})^2}{\sum(Y_i - \bar{y})^2} = \frac{\text{cov}(x_i, y_i)^2}{\text{var}(x_i)\text{var}(y_i)} \dots \dots \dots (3)$$

Le  $R^2$  peut aussi s'écrire en fonction des résidus :

$$R^2 = \frac{\text{SCE}}{\text{SCT}} \dots \dots \dots (4)$$

On a :

$$\text{SCT} = \text{SCE} + \text{SCR} \dots \dots (5)$$

$$R^2 = \frac{\text{SCE}}{\text{SCT}} = 1 - \frac{\text{SCR}}{\text{SCT}} \dots \dots (6)$$

$$R^2 = 1 - \frac{\|\hat{\varepsilon}\|^2}{\|Y - \bar{y}\|^2} \dots \dots \dots (7)$$

Un modèle, que l'on qualifiera de bon, possédera des estimations  $y_i$  proches des Vraies valeurs de  $y_i$ . Les deux quantités SCT et SCE modèle sont des sommes de carrés donc toujours positives ou nulles et telles que :  $\text{SCE} < \text{SCT}$ . Le rapport de ces deux quantités est compris entre 0 et 1 que l'on appelle le coefficient de détermination  $R^2$ .

## IV.3.1.3 Critère de validation croisée : PRESS

La somme des erreurs de prédiction « Prediction Sum of Squares » (PRESS) est définie par

$$\text{PRESS} = \sum_i \varepsilon_i^2 \dots \dots \dots (8)$$

Ce critère permet de sélectionner les modèles ayant un bon pouvoir prédictif. (On cherche toujours le PRESS le plus petit).

## IV.3.1.4 La valeur p "P-Value" :

Pour éviter de raisonner seulement sur F, les programmes destinés à l'analyse statistique fournissent la p-value associée au F observé. Le p-value est le niveau de significativité du test de Fisher-Snedecor, c'est-à-dire la probabilité de dépasser le F observé si l'hypothèse nulle est vraie.

On compare la p-value au risque  $\alpha$  choisi (par exemple  $\alpha=0.05$ ),

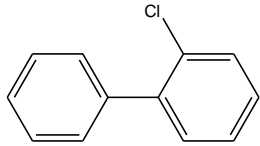
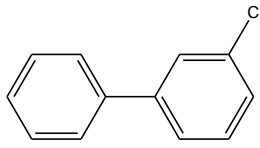
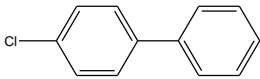
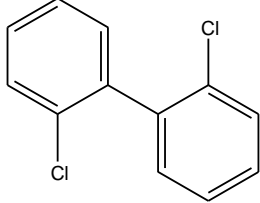
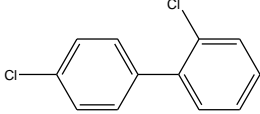
Si  $P\text{-value} \leq \alpha$  alors on rejette l'hypothèse nulle  $\beta_1 = \beta_2 = \beta_3 \dots = \beta_p = 0$ .

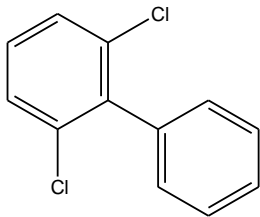
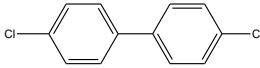
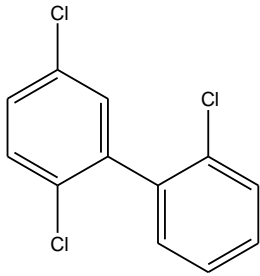
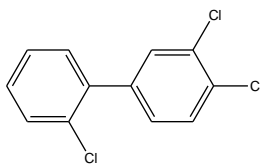
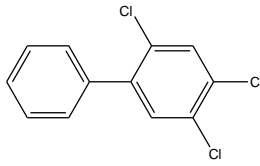
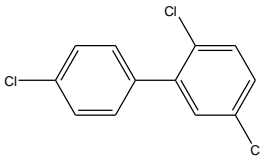
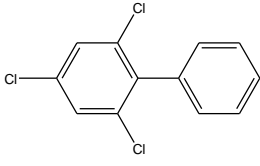
# **Partie application**

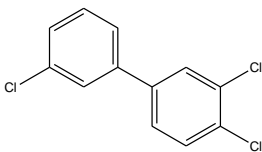
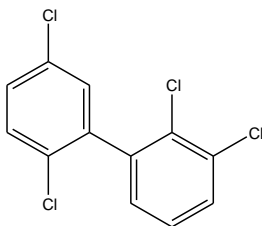
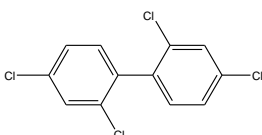
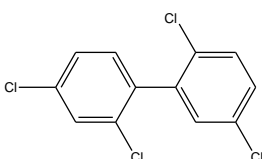
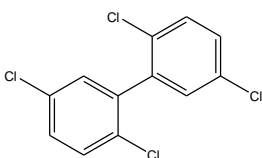
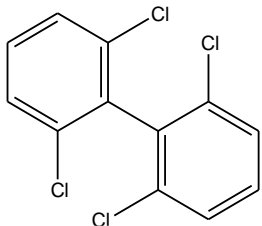
Dans cette partie expérimentale, nous allons prédire la solubilité d'une série des polychlorobiphényles (PCB).

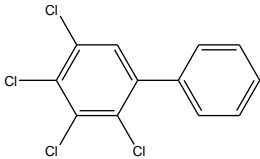
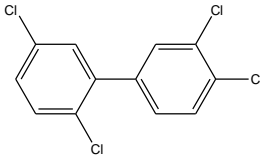
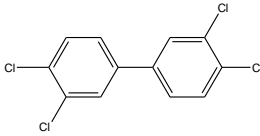
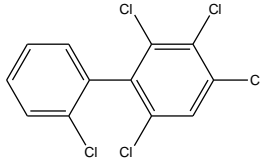
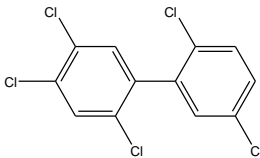
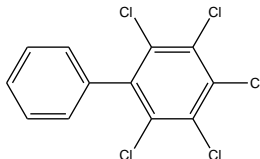
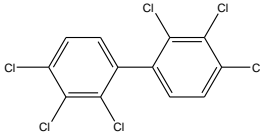
La liste des molécules étudiées dans notre travail présentée dans le Tableau VIII, cette base des données contient 34 composés prélevés de la littérature [31], avec les valeurs de leurs solubilités. Dans cette étude nous avons utilisé le logarithme de la solubilité

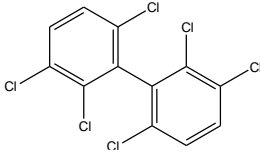
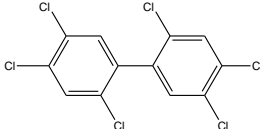
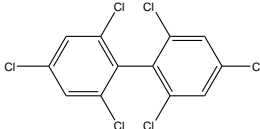
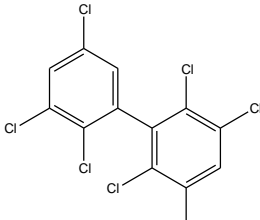
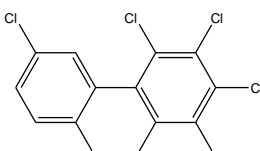
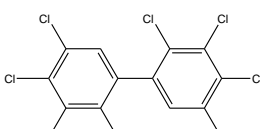
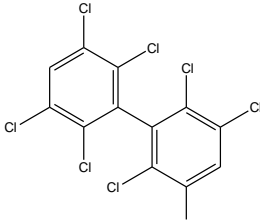
**Tableau VIII** : Liste des molécules étudiées et leurs solubilités.

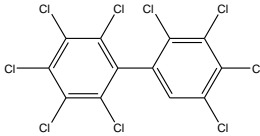
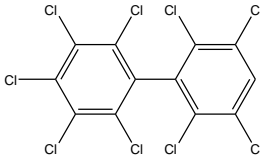
numéro	composé	Formule générale	structure	Log S
1	2-chlorobiphenyl	$C_{12}H_9Cl$		-4,63
2*	3-chlorobiphenyl	$C_{12}H_9Cl$		-5,16
3	4-chlorobiphenyl	$C_{12}H_9Cl$		-5,33
4	2,2'-dichlorobiphenyl	$C_{12}H_8Cl_2$		-5,45
5*	2,4'-dichlorobiphenyl	$C_{12}H_8Cl_2$		-6,56

numéro	composé	Formule générale	structure	Log S
6*	2,6-dichlorobiphenyl	$C_{12}H_8Cl_2$		-5,63
7	4,4'-dichlorobiphenyl	$C_{12}H_8Cl_2$		-6,62
8	2,2',5-trichlorobiphenyl	$C_{12}H_7Cl_3$		-6,62
9*	2,3,4'-trichlorobiphenyl	$C_{12}H_7Cl_3$		-6,52
10	2, 4,5-trichlorobiphenyl	$C_{12}H_7Cl_3$		-6,45
11	2,4',5-trichlorobiphenyl	$C_{12}H_7Cl_3$		-6,44
12	2, 4,6-trichlorobiphenyl	$C_{12}H_7Cl_3$		-6,06

numéro	composé	Formule générale	structure	Log S
13	3,3',4-trichlorobiphenyl	$C_{12}H_7Cl_3$		-7,22
14	2,2', 3,5'-tetrachlorobiphenyl	$C_{12}H_6Cl_4$		-6,23
15	2,2', 4,4'-tetrachlorobiphenyl	$C_{12}H_6Cl_4$		-6,78
16*	2,2',4',5-tetrachlorobiphenyl	$C_{12}H_6Cl_4$		-7,3
17	2,2', 5,5'-tetrachlorobiphenyl	$C_{12}H_6Cl_4$		-7,28
18	2,2', 6,6'-tetrachlorobiphenyl	$C_{12}H_6Cl_4$		-8,03

numéro	composé	Formule générale	structure	Log S
19	2, 3, 4,5-tetrachlorobiphenyl	$C_{12}H_6Cl_4$		-7,18
20	2,3',4',5-tetrachlorobiphenyl	$C_{12}H_6Cl_4$		-7,25
21	3,3', 4,4'-tetrachlorobiphenyl	$C_{12}H_6Cl_4$		-7,41
22*	2,2', 3, 4,6-pentachlorobiphenyl	$C_{12}H_5Cl_5$		-7,44
23	2,2', 4, 5,5'-pentachlorobiphenyl	$C_{12}H_5Cl_5$		-7,89
24	2, 3, 4, 5,6-pentachlorobiphenyl	$C_{12}H_5Cl_5$		-7,38
25	2,2',3, 3', 4,4'-hexachlorobiphenyl	$C_{12}H_4Cl_6$		-8,91

numéro	composé	Formule générale	structure	Log S
26*	2,2', 3,3', 6,6'- hexachlorobiphenyl	$C_{12}H_4Cl_6$		-7,9
27	2,2', 4,4', 5,5'- hexachlorobiphenyl	$C_{12}H_4Cl_6$		-8,48
28	2,2', 4,4', 6,6'- hexachlorobiphenyl	$C_{12}H_4Cl_6$		-8,52
29	2,2', 3,3', 5,5',6- heptachlorobiphenyl	$C_{12}H_3Cl_7$		-8,6
30*	2,2', 3, 4, 5,5',6- heptachlorobiphenyl	$C_{12}H_3Cl_7$		-8,93
31	2,2', 3,3', 4,4', 5,5'- octachlorobiphenyl	$C_{12}H_2Cl_8$		-9,54
32*	2,2', 3,3', 5,5', 6,6'- octachlorobiphenyl	$C_{12}H_2Cl_8$		-9,5

numéro	composé	Formule générale	structure	Log S
33	2,2', 3,3', 4,4', 5,5',6-nonachlorobiphenyl	$C_{12}HCl_9$		-9,77
34*	2,2', 3,3', 4, 5,5', 6,6'-nonachlorobiphenyl	$C_{12}HCl_9$		-10,4

\*Composés de validation

## I- Calcul du modèle :

### I.1 Choix du modèle :

Les molécules sont dessinées par le logiciel Chem Draw [32] puis elles sont optimisées en utilisant le logiciel Hyper Chem [33] en utilisant une méthode semi empirique PM3. Les descripteurs moléculaires ont été calculés à l'aide du logiciel informatique Dragon [34], plus de 1600 descripteurs sont calculés. Nous avons éliminé les descripteurs qui portent les mêmes informations (deux tiers de la totalité présente la même valeur)

Le nombre de descripteurs est réduit en 1090 descripteurs

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.

Le choix des descripteurs intervenant dans le modèle a été fait par les méthodes de Pas à Pas dans le logiciel Minitab [35].

Le modèle obtenu à pour deux descripteurs, sa validation est effectuée par 10 composés, alors que son calibrage par 24 composés.

### I.2 Choix d'une famille de descripteurs :

Les descripteurs entrant dans le modèle obtenus par la méthode Pas à Pas, leurs classes et de brèves définitions sont donnés dans le tableau IX

**Tableau IX :** Descripteurs moléculaires intervenant dans la modélisation de la solubilité (Log S).

N°	Descripteur	Classe	Signification
1	X2v	Connectivity indices (block 4)	Valence connectivity index chi-2
2	Mor12v	3D-MoRSE descriptors (block14)	3D-MoRSE – signal 12 / weighted by atomic van der waals volumes

Les valeurs du logarithme de la solubilité et les valeurs des descripteurs sont présentées dans le tableau X au-dessous :

**Tableau X :** Valeurs de log S, X2v, Mor12v

molécule	Log S	X2v	Mor12v
2-chlorobiphenyl	-4,63	3,253	0,027
4-chlorobiphenyl	-5,33	3,309	0,054
2,2'-dichlorobiphenyl	-5,45	3,774	-0,072
4,4'-dichlorobiphenyl	-6,62	3,887	0,006
2,2',5-trichlorobiphenyl	-6,62	4,355	-0,153
2, 4,5-trichlorobiphenyl	-6,45	4,334	-0,058
2,4',5-trichlorobiphenyl	-6,44	4,411	-0,084
2, 4,6-trichlorobiphenyl	-6,06	4,361	-0,153
3,3',4-trichlorobiphenyl	-7,22	4,390	-0,005
2,2', 3,5'-tetrachlorobiphenyl	-6,23	4,854	-0,188
2,2', 4,4'-tetrachlorobiphenyl	-6,78	4,936	-0,194
2,2', 5,5'-tetrachlorobiphenyl	-7,28	4,936	-0,245
2,2', 6,6'-tetrachlorobiphenyl	-8,03	4,822	0,055
2, 3, 4,5-tetrachlorobiphenyl	-7,18	4,755	-0,051
2,3',4',5-tetrachlorobiphenyl	-7,25	4,911	-0,105
3,3', 4,4'-tetrachlorobiphenyl	-7,41	4,887	-0,024
2,2', 4, 5,5'-pentachlorobiphenyl	-7,89	5,436	-0,141
2, 3, 4, 5,6-pentachlorobiphenyl	-7,38	5,201	-0,178
2,2', 3,3', 4,4'-hexachlorobiphenyl	-8,91	5,772	-0,185

2,2', 4,4', 5,5'-hexachlorobiphenyl	-8,48	5,936	-0,317
2,2', 4,4', 6,6'-hexachlorobiphenyl	-8,52	5,991	-0,192
2,2', 3,3', 5,5',6-heptachlorobiphenyl	-8,60	6,384	-0,352
2,2', 3,3', 4,4', 5,5'-octachlorobiphenyl	-9,54	6,778	-0,301
2,2', 3,3', 4,4', 5,5',6-nonachlorobiphenyl	-9,77	7,224	-0,406
3-chlorobiphenyl*	-5,16	3,313	0,069
2,4'-dichlorobiphenyl*	-6,56	3,830	-0,033
2,6-dichlorobiphenyl*	-5,63	3,777	-0,103
2,3',4'-trichlorobiphenyl*	-6,52	4,330	-0,041
2,2',4',5-tetrachlorobiphenyl*	-7,30	4,936	-0,210
2,2', 3, 4,6-pentachlorobiphenyl*	-7,44	5,301	-0,288
2,2', 3,3', 6,6'-hexachlorobiphenyl*	-7,90	5,821	-0,097
2,2', 3, 4, 5,5',6-heptachlorobiphenyl*	-8,93	6,303	-0,356
2,2', 3,3', 5,5', 6,6'-octachlorobiphenyl*	-9,50	6,827	-0,304
2,2', 3,3', 4, 5,5', 6,6'-nonachlorobiphenyl*	-10,40	7,248	-0,347

\*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 XI

**Tableau XI** : Matrice de corrélation

	Log S	X2v
X2v	-0.944 0.000	
Mor12v	0.672 0.000	-0.843 0.000

L'application de la méthode Pas à Pas par la sélection des descripteurs dans notre modèle conduit à un modèle linéaire avec deux descripteurs (X2v , Mor12v)

La matrice de corrélation présente dans le tableau XI montre une grande corrélation entre les deux descripteurs, pour cela nous avons éliminé le descripteur Mor12v et on a garde le X2v parce que sa corrélation avec la solubilité est très importante par rapport le Mor12v

## II.2 Equation et analyse de régression :

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

$$\text{Log } S = -1,23 (\pm 0,4586) - 1,22 (\pm 0,09078) X_{2v}$$

**Tableau XII :** Analyse de régression

Predictor	Coef	SE Coef	T	P	VIF
Constant	-1,2274	0,4586	-2,68	0,014	
X2v	-1,21628	0,09078	-13,40	0.000	1,000

**Tableau XIII :** Les paramètres statistiques pour l'ensemble de calibration (24 composés) et celui de validation (10 composés)

<b>Les paramètres statistiques</b>	
F	179,51
$R^2$	89,1 %
$R^2_{adj}$	88,6 %
$Q^2_{Loo}$	87,3 %
$Q^2_{ext}$	94.72 %
S	0,439
EQMC	0,421
EQMP	0,454
$EQMP_{ext}$	0,293

Les valeurs de  $R^2$  et  $R^2_{adj}$  montrent la qualité de l'ajustement, alors que la petite différence entre  $R^2$  et  $Q^2_{Loo}$  renseigne sur la robustesse du modèle qui est, en outre, hautement significatif (grande valeur de la statistique F de Fisher). Les valeurs assez proches de EQMC et EQMP signifient que la capacité de prédiction interne du modèle n'est pas trop dissemblable de son pouvoir d'ajustement.

La validation statistique externe ( $Q^2_{ext}$  ;  $EQMP_{ext}$ ) atteste de la bonne capacité prédictive des composés n'ayant pas participé au calcul du modèle.

### II.3 Diagramme de Williams :

Le domaine d'application a été discuté à l'aide du diagramme de Williams, ou nous avons présenté les erreurs standardisées de prédiction en fonction des valeurs de levier.

L'analyse de la figure 2, montre une absence totale des points aberrants, tous les points sont compris entre les limites  $\pm 2,5$ . Le diagramme fait ressortir deux points influents, un de l'ensemble de calibrage et le deuxième de l'ensemble de validation. Ces deux points présentent un bras de levier important  $h > h^*$ .

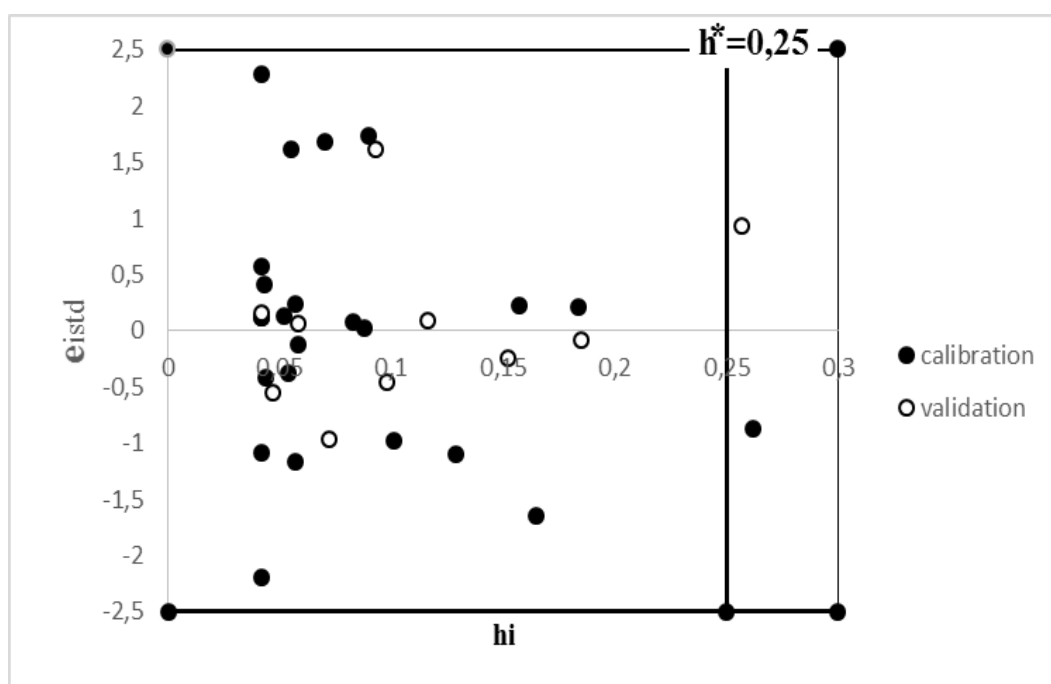


Figure 2 : Diagramme de Williams

### II.4 Qualité de l'ajustement :

La représentation des valeurs prédites du logarithme de la solubilité en fonction des valeurs expérimentales (Figure 3) montre un bon ajustement, une faible dispersion autour de la première bissectrice pour les deux ensembles de calibrage et de validation, ceci est confirmé par la valeur de  $Q^2$  qu'elle est autour de 90%.

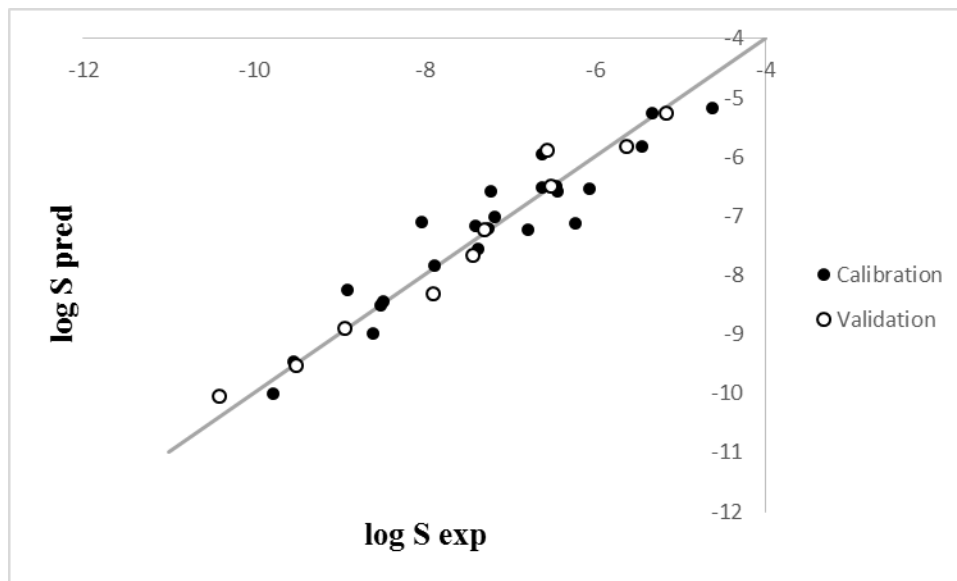


Figure 3 : Qualité de l'ajustement

### 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 04), dans cette figure on à présenter le graphe des coefficients statistiques  $Q^2$  et  $R^2$ .

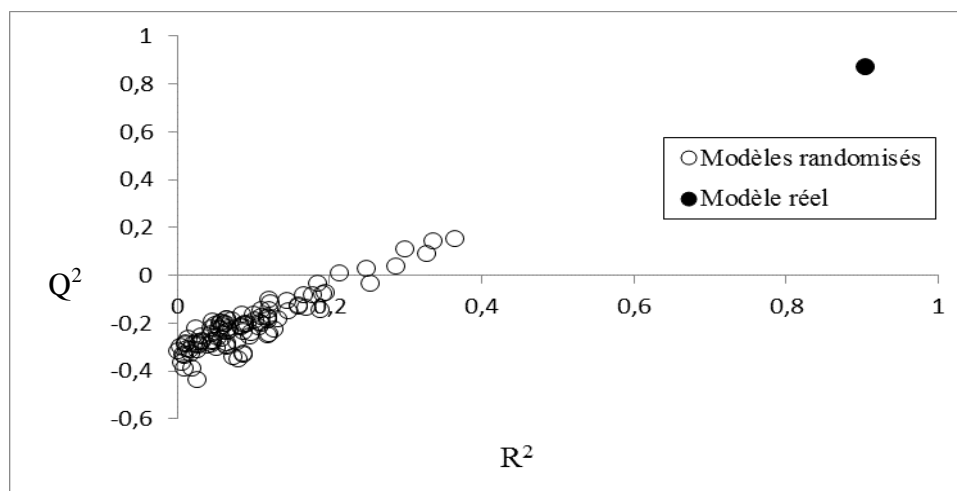


Figure 4: Test de randomisation

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

# **Conclusion générale**

### **Conclusion générale :**

Dans notre travail, nous utilisons la méthode QSPR pour prédire la solubilité d'une série de 34 polluants organiques (biphényles polychlorés), pour lesquels nous relierons cette propriété à des descripteurs moléculaires théoriques.

Cette étude a adopté la méthode de régression linéaire multiple et les descripteurs impliqués dans le modèle ont été sélectionnés par le logiciel Minitab.

Le modèle choisi a donné de bons résultats : ajusté pour la qualité, le pouvoir prédictif, etc., démontrant que la relation entre la solubilité et la variable explicative ( $X_{2v}$ ) est essentiellement linéaire.

# **Références bibliographiques**

### Références bibliographiques :

- [1] (15 mars 2009). "The toxicity of certain benzene derivatives and related compounds". US EPA National Center for Environmental Assessment. *Journal of Industrial Hygiene and Toxicology*. V 13. pp 87-96.
- [2] C. K. Drinker, M. F. Warren ET G. A. Bennett. (1937). "The Problem of Possible Systemic Effects from Certain Chlorinated Hydrocarbons". *Journal of industrial hygiene and toxicology*. V19 No 7. pp 283-311
- [3] K. Abbes, L. Bizzozero. (2010). "Les polychlorobiphényles, Un problème nouveau pour une pollution ancienne". Ecole des Hautes Etudes en Santé Publique.
- [4] C Dargnat. & C Fisson. (2010). "Les PCB dans le bassin de la Seine et son estuaire ". Etude réalisée par le GIP Seine-Aval. pp 134
- [5] P, Meunier. (2008). "Rapport d'information sur le Rhône et les PCB : une pollution au long cours". Assemblée Nationale. n° 998.
- [6] Larry. W Robertson, L.G Hansen. (2001). "PCBs: recent advances in environment toxicology and health aspects". United Environment Programme : rapport sur les polluants organiques persistants. The University Press of Kentucky
- [7] Martin Hirsch. (2002). " la question d'un éventuel lien entre exposition à l'aspartame et tumeurs du cerveau ". Rapport de l'Agence Française de Sécurité Sanitaire des Aliments
- [8] Nations Unies -PNUE. (2001). " Manuel de formation visant la préparation d'un Plan National pour la Gestion ... projet pilote pour la gestion écologiquement rationnelle des PCBs et des équipements contaminés aux PCs dans le cadre de la mise en oeuvre de la convention de Bâle ".
- [9] Observatoire régional de la santé du Rhône-Alpes. (Décembre 2008). "Les PCB (Polychlorobiphényles) ". Les dossiers de l'environnement de l'ORS. N°2.
- [10] MP Caron, P Levallois, S Gingras, D Phaneuf. (2004). " Étude sur la qualité de l'eau potable dans sept bassins versant en surplus de fumier et impacts potentiels sur la santé - Étude de la consommation d'eau dans la population adulte ". Institut national de santé publique du Québec.
- [11] G.M Richardson. (1997). *Compendium of Canadian Human Exposure Factors for Risk Assessment*. O'Connor Associates Environmental Inc.
- [12] "Scientific report of Efsa; Update of the monitoring of levels of dioxins and PCBs in food and feed". (2012). *EFSA Journal*. Vol 10 issue 7. pp 2832

- [13] L. P. Hammett. "The Effect of Structure upon the Reactions of Organic Compounds, Benzene Derivatives". *Journal of the American Chemical Society*. V 59. pp 96-103.
- [14] M Goodarzi., T. Chen, Matheus P. Freitas. (2010). *Laboratory Systems*. V 104. pp 260-264.
- [15] A Danial, M. A. Sobati. (2014). *International Journal of Cancer*. V 40. pp 282.
- [16] T Roberto., V. Consonni. (2000). "Handbook of Molecular Descriptors". WILEY-VCH.
- [17] A Akhiezer. I., Peltminskii, S.V. (1981). "Methods of Statistical Physics". Pergamon Press, Oxford.
- [18] P. W Atkins. (1982). "Physical Chemistry". Ch. 20. 2nd edition. W.H., Freeman and Co, San Francisco
- [19] D. M Allen. (1974). "The relationship between variable selection and data augmentation and a method for prediction". *Technometrics*. V 16. pp 125-127.
- [20] A Golbraikh., A. Tropsha, Beware. (2002). *Journal of Molecular Graphics and Modelling*. V 20. pp 269-276.
- [21] A Tropsha. P. Gramatica, and V. K. Gombar. (2003). "The importance of being earnest, validation is the absolute essential for successful application and interpretation of QSPR models, QSAR"
- [22] A Antoniadis, J Berruyer, R Carmona. (1992). "Régression non linéaire et applications".
- [23] D Birkes, Y Dodge. (1993). "Alternative methods of regression". *Wiley Series in Probability and Statistics*. 1st Edition by Birkes
- [24] A. M Legendre. (1805). "Nouvelles Méthodes pour la Détermination des Orbites des Comètes Courcier Paris "
- [25] C. F Gauss. (1809). "Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium Frid". Perthes et I.H.Besser, Hamburgi.
- [26] Francis Galton. (1886). "Regression towards Mediocrity in Hereditary Stature". *Journal of the Anthropological Institute*. V 15. pp 246.
- [27] C Alvin. Rencher and G. Bruce Schaalje. (2008). "Linear Models in Statistics". Second Edition John Wiley & Sons
- [28] S Weisberg. (1985). "Applied Linear Regression". New York: John Wiley & sons
- [29] D. F Morrison. (1983). "Applied Linear Statistical Methods". *Journal of the Royal Statistical Society: Series A (General)*. V 147. pp 530
- [30] (a) Christensen R. (1996). "Plane Answers to Complex Questions: The Theory of Linear Models". (2nd ed.). New York: Springer-Verlag. (b) (1997). "Log-Linear Models and

Logistic Regression". (2nd ed.). New York: Springer-Verlag.

[31] N Nagmany , Nirmalakhandan and Richard E, Speece. (1989). "Prediction of aqueous solubility of organic chemicals based on molecular structure". Environ Sci Technol

[32] ChemDraw ultra, Release 7.0

[33] Hyperchem<sup>TM</sup>, Release 6.03, for Windows, (2000)

[34] R Todeschini, V Cosonni, Dragon, P.M. (2006). Software for the calculation of molecular descriptors. Release 5.3 for windows, Milano.

[35] MINITAB, Release 16, Statistical software, (2000)