وزارة التعليم العالى و البحث العلمى

Université BADJI Mokhtar – Annaba
BADJI Mokhtar – Annaba University

جامعة باجي مختار – عنابة

Faculté des Sciences
Département de Chimie

MEMOIRE

Présenté pour l'obtention du diplôme MAGISTER

Par M^r^. BOUFENAYA Hamza

DES en Chimie

Option : Chimie et environnement

T H E M E

Modèle prédictif du facteur de capacité de phénols séparés par CLHP-PI avec une phase mobile méthanol-eau

Président :	**Pr M^r^. D. MESSADI**	**pr.**	**UBMA**
Examinateurs :	**Mme.s.Ali-MOUKHNACHE**	**Pr.**	**UBMA**
	Mr. A.DJALEL	**MC**	**UBMA**
Rapporteur	**Mr.KHATMI Djemeleddine**	**pr.**	**UBMA**

Année 2010

Remerciement

Cette étude a été réalisée au laboratoire de ***sécurité environnementale et alimentaire*** de l'université d'Annaba sous la direction de monsieur le ***Pr.Djeloul MESSADI*** que je remercie vivement, pour le bienveillant intérêt qu'il a accordé à ce travail.

Je souhaiterais tout d'abord adresser mes sincères remerciements à monsieur ***KHATMI Djameleddine***.

Je remercie aussi tous les membres de jury :

Le ***Pr MESSADI Djelloul*** qui me fait l'honneur de le présider ; madame ***ALIMOKHNACH Salima*** *et* ***Mr DJALLAL Ahmed*** pour avoir accepté d'examiner et de juger mon travail.

Je tiens enfin à remercier tous les membres du laboratoire ***LASEA***.

Dédicaces

Je dédie ce travail :

A mon père

A ma mère qui m'a éclairé mon chemin et qui ma encouragé et soutenue tout au long de mes études.

A mes frères

A mes sœurs

A tous mes amis.

A toute l'équipe du ***LASEA***

ملخص :

الاستبقاء (log k) من خليط غير متجانس من الفينولات مفصولة بنظام isochratique بواسطة CLHP على عمود (Partisil ODS) مع الطور المتحرك ميثانول/ماء كان متصلا ، وشارك فيφ المياه للشروط التحليلية (T درجة الحرارة ، وكسر حدة التخزين، φ ، مذيب عضوي) ، ومعامل تفريق الأوكتانول / Ghose-Crippen-Viswanadhan (ALOG P) محسوبة باستخدام برمجيات الكمبيوتر (DRAGON).

معايرة مجموعة (40 عنصرا)، التي تم الحصول عليها من خلال تطبيق الخوارزمية دوبلكس ، تحسب نموذجا تلبية افتراضات النموذج الإحصائي الخطي مع الآثار ، وقوية ثابتة ، والتي قدرتها التنبؤية الداخلية ليست غاية تختلف عن قوة التكيف. إحصائية التحقق من صحة الخارجية على اختبار مجموعة من 26 عناصر ، تبين مدى قدرة تنبؤية جيدة مسجلة لا تستخدم في حساب النموذج.

الكلمات الدالة :

الفينولات - CLHP/PI - الإستبقاء - نموذج.QSRR

Abstract:

The retention (log k) of a heterogeneous mixture of phenols separated by HPLC system isochratique - PI on a Partisil ODS column with a mobile phase methanol - water has been linked , Co-organic solvent)φto analysis conditions (temperature T, volume fraction, φ and the partition coefficient n-octanol / water Ghose-Crippen-Viswanadhan (A log P) calculated using the software DRAGON computer.

The calibration set (40 elements), calculated by the algorithm DUPLEX, calculates a model satisfying the assumptions of a linear statistical model with fixed effects, robust, and whose internal predictive ability is not too dissimilar power adjustment. Statistical external validation on a test set of 26 elements, demonstrates good predictive ability of log k n'ayant not used in calculating the model.

Key - words: phenol - HPLC / IP - Retention - Model QSRR.

Résumé :

La rétention (log k) d'un mélange hétérogène de phénols séparés en régime isochratique parCLHP–PI, sur une colonne Partisil ODS, avec une phase mobile méthanol – eau a été reliée aux conditions d'analyse (température T ; fraction volumique, φ, du co-solvant organique) et au coefficient de partage n-octanol / eau de Ghose-Crippen-Viswanadhan (A log P) calculé à l'aide du logiciel informatique DRAGON.

L'ensemble de calibration (40 éléments), obtenu en appliquant l'algorithme DUPLEX, permet de calculer un modèle vérifiant les hypothèses d'un modèle statistique linéaire à effets fixes, robuste, et dont la capacité de prédiction interne n'est pas trop dissemblable de son pouvoir d'ajustement. La validation statistique externe, sur un ensemble test de 26 éléments, atteste de la bonne capacité prédictive des log k n'ayant pas servi au calcul du modèle.

Mots – clés : Phénols – CLHP / PI – Rétention – Modèle QSRR.

Sommaire

Introduction générale

On appelle phénols les dérivés hydroxylés du benzène et des hydrocarbures aromatiques, dans lesquels le groupe OH est lié à un atome de carbone du cycle benzénique. Les dérivés polyhydroxylés sont appelés polyphénols. Rappelons que chez les alcools le groupe OH est lié à un atome de carbone saturé.

Le phénol est un produit de synthèse. Pur, il se présente à la température ordinaire comme un solide blanc cristallisé. C'est un composé toxique qui provoque des brûlures graves sur la peau. Il doit être manipulé en utilisant des gants et des lunettes de protection. Ses solutions (acide phénique) ont été parmi les premiers antiseptiques utilisés en médecine. On l'utilise dans l'industrie comme réactif de base dans la synthèse du cyclohexanol dont la coupure oxydante conduit au Nylon 6,6.

D’autres composés phénoliques sont utilisés pour le tannage, en cosmétique, dans l’industrie organique (fabrication de matières plastiques,produits,pharmaceutiques, explosifs...) ,ainsi que pour le développement photo, ce qui en fait d’importants polluants potentiels de l’environnement

Les relations structure/Retention quantitatives, désignées par l'abréviation QSRR (Quantitave Structure/Retention Relationships), très utilisées depuis une vingtaine d'années, constituent des modèles mathématiques pour l'approximation des relations, souvent complexes, entre la structure caractérisée par des descripteurs moléculaires, et la rétention en chromatographie .

Les techniques les plus courantes pour établir des modèles QSRR utilisent l'analyse de régression (régression multilinéaire : MLR) et les réseaux neuronaux (RNA) pour ne citer que cela.

Nous avons imposé un ensemble de régresseurs (3 descripteurs: température de la colonne, fraction volumique du méthanol dans la phase mobile et le coefficient de partage

n-octanol-eau) qui donnent une reconstitution satisfaisante de la variable à expliquer et nous avons appliqué la régression linéaire multiple (RLM) pour modélisé le facteur de capacité(variable à expliquer) de 10 phénols séparés par CLHP-PI.

Notre mémoire comporte en plus de la bibliographie, de l'introduction et de la conclusion générale ; deux grandes parties :

Dans la première on rappelé quelques les généralités sur les phénols (Définitions et historique, les phénols et la santé…), brèves principes sur la chromatographie liquide haute performance ainsi qu'une définition des modèles QSAR et QSPR et les méthodes statistiques pour leur développement et l'évaluation de leur qualité.

Dans la deuxième partie (expérimentale) nous avons modélisé les facteurs de rétention (ou de capacité) de 10 phénols séparés par CLHP – PI, à différentes températures, et pour différentes compositions de la phase mobile méthanol – eau.

I. Les phénols :

I-1 –Définition et historique

Au 17ème siècle, Glauber obtient le phénol à l'état impur, à partir des produits issus de la distillation sèche de la houille. Il le décrit «comme une huile vive et rouge de sang qui assèche et guérit tous les ulcères humides». Deux siècles plus tard, un chirurgien anglais, Joseph Lister redécouvrira les vertus antiseptiques du phénol. Encore appelé acide phénique, le phénol a été très utilisé comme antiseptique pour soigner les blessures envenimées ; il est abandonné aujourd'hui en raison de sa toxicité.

Les phénols sont des composés aromatiques possédant un ou plusieurs groupements hydroxyle, -OH, substitués sur le(s) cycle(s)

Des phénols sont obtenus par distillation sèche de la houille (phénol, crésols, α et β naphtols) ou du bois et plus précisément de la lignine (phénol, crésols, créosols, gaïacol). Des phénols polyfonctionnels, de structures souvent complexes, sont très répandus dans le règne végétal, le plus souvent sous forme d'esters ou de glucosides. Parmi les plus simples d'entre eux, on peut citer la vanilline des gousses de vanille, l'eugénol des clous de girofle, le thymol du thym et l'acide gallique de la noix de galle.

I-2- STRUCTURE ET NOMENCLATURE

Nous présentons dans la figure 1 la structure et la nomenclature de quelques phénols :

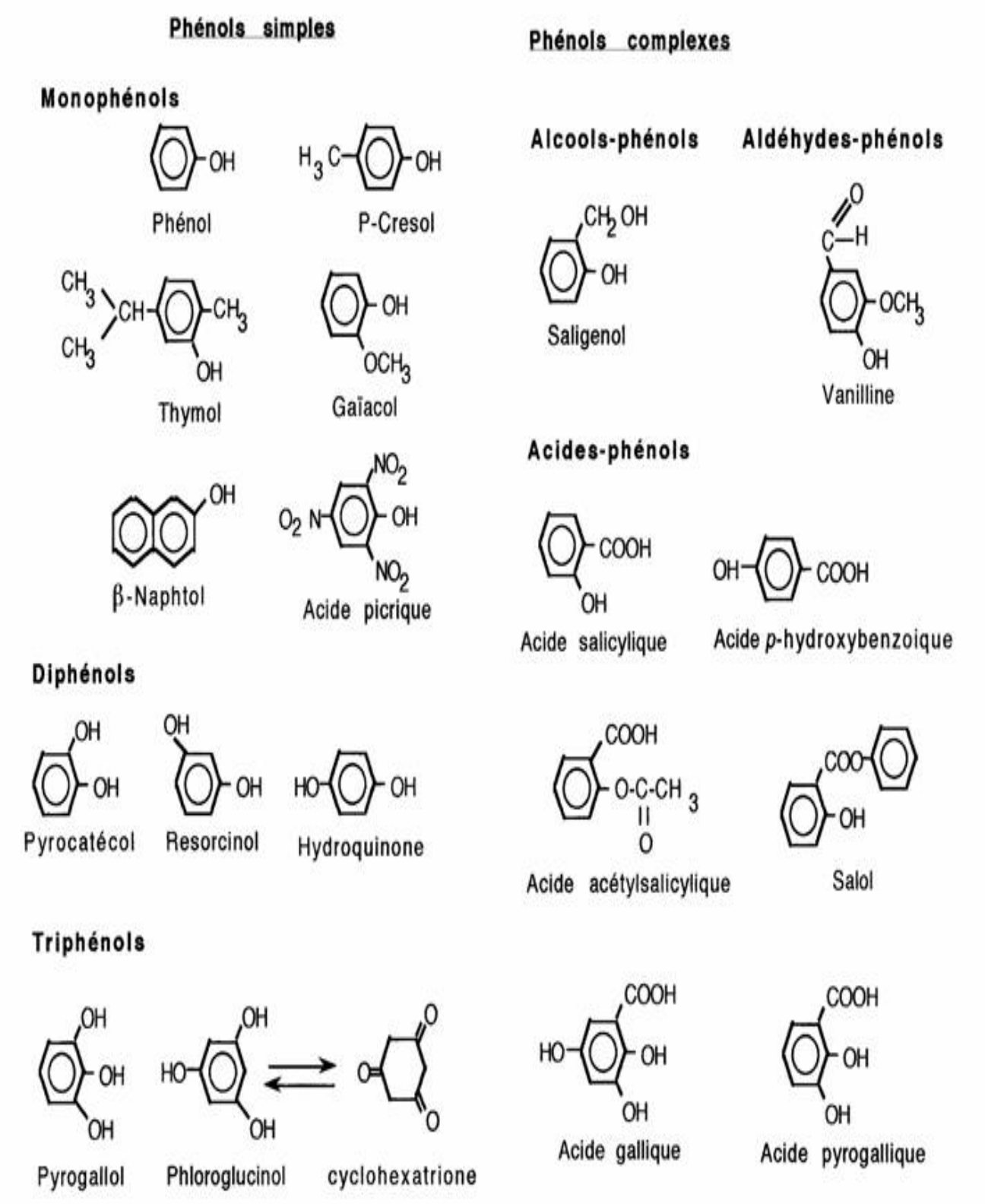


Figure 1 : *Structures et nomenclatures de quelques phénols*

I-3- PROPRIETES PHYSIQUES

L'introduction d'un hydroxyle sur un cycle aromatique augmente la possibilité de formation de liaisons intermoléculaires sur des composés qui étaient déjà des liquides ou des solides. Les phénols sont donc des solides cristallins. Ils possèdent en général une forte odeur caractéristique (de "gouache").

Leur insolubilité dans l'eau les différencie nettement des alcools.

Le phénol ("acide phénique") est faiblement hydrosoluble mais très hygroscopique: on prépare le "phénol aqueux" en chauffant 9g de phénol avec lg d'eau, il cristallise vers 15°C et est utilisé pour préparer l'eau phéniquée officinale.

I- 4- PROPRIETES CHIMIQUES

La réactivité des phénols tient de celle des alcools et de celle des dérivés benzéniques mais elle offre également de grandes particularités liées à la conjugaison des doublets électroniques de l'oxygène avec le cycle. Il s'ensuit que:

1) le clivage C-O est impossible,

2) les réactions de substitution électrophile sont facilitées et orientées en ortho et para,

3) les phénols sont plus acides que les alcools: pKa = 9-10. Seules les réactions entrainant le clivage O-H seront envisagées; elles peuvent être hétérolytiques ou homolytiques.

I-5- PROPRIETES BIOLOGIQUES

I-5.1- Usages

Les phénols simples (c'est-à-dire sans autre fonction chimique) sont utilisés comme antiseptiques et éventuellement insecticides externes.

- monophénols
- eau phénolée (1 à 2%) : antiseptique, calmant et antipungineux, mais risque de nécrose cutanée.
- crésylol (mélange des créosols, o, m et p).
- thymol
- béta-naphtol et benzonaphtol (benzoate de (3-naphtol): exception = antiseptique intestinal.
- polyphénols
- pyrocathécol, pyrogallol, gaïacol, hexylrésorcinol.

- goudrons: produits phénoliques bruts obtenus par pyrogénation de la houille ou du bois (goudron de cade, créosotre de hêtre ....).

I-5.2- Toxicité

La toxicité des phénols peut se manifester localement, principalement à l'occasion de leur utilisation comme antiseptique, ou de façon générale après ingestion, inhalation, ou contact cutané prolongé.

- toxicité locale: dermatose puis gangrène.
- toxicité générale: les phénols induisent des troubles nerveux (convulsions puis coma) et des œdèmes pulmonaires.

II-LA CHROMATOGRAPHIE LIQUIDE-LIQUIDE

II-1 INTRODUCTION

Découverte par **MARTIN** et **SYNGE** en 1941, la chromatographie liquide- liquide appelée aussi chromatographie de partage est l'une des puissantes méthodes de séparation à résolution élevée.

Il y a quelques années, cette méthode n'était pas très utilisée à cause de difficulté d'immobiliser les phases stationnaires comme dans le cas de chromatographie sur couche mince. Grâce à l'évolution de la théorie et aux grands progrès de la technologie dans la fabrication des colonnes, de pompes performantes et de détecteurs sensibles, la chromatographie de partage, est devenue bien plus facile, et plus commode [1].

Il est à remarquer que la chromatographie liquide- liquide et la chromatographie gaz-liquide , présentent des analogies, notamment en ce qui concerne l'efficacité des colonne et les temps d'analyse.

La chromatographie liquide- liquide est utilisée essentiellement pour la séparation des molécules très polaires dont le poids moléculaire est inférieur a 1000 et pour les homologues d'une même série mal séparés par chromatographie liquide- solide [1]

II-2 - SUPPORT ET PHASE STATIONNAIRE

En chromatographie liquide- liquide le support se compose d'un lit de fines particules solides, qui présente une très grande surface pour retenir une grande quantité de phase immobile dans un petit volume.

Il est nécessaire que ces supports ne réagissent pas avec les solutés, et que leurs propriétés adsorptives soient totalement masquées.

Les supports sont imprégnés de phases stationnaires qui sont en général des substances dans lesquelles les composés à séparer sont solubles.

Les phases mobile qui traverse la colonne a une très grande interface de contact avec la phase stationnaire, ce qui permet une distribution de soluté ente les deux phases.

En chromatographie liquide haute performance, il existe deux types de phases stationnaires :

-la phase normale :

Constituée de gel de silice, matériau très polaire. Il faut donc utiliser un éluant apolaire. Ainsi alors de l'injection d'une solution, les produits polaires sont retenus dans la colonne, contrairement aux produits apolaires qui sortent en tête.

L'inconvénient d'une telle phase, réside dans détérioration rapide au cours du temps du gel de silice, ce qui se traduit par un manque de reproductibilité des séparations.

-la phase inverse :

Elle est majoritairement composée de silice greffée par des chaînes linéaires de 8 ou 18 atomes de carbone (C_8 et C_{18}). Cette phase et apolaire et nécessite donc un éluant polaire : acétonitrile (ACN), méthanol, eau. Dans ce cas, ce sont les composés polaires qui seront élués en premier.

Contrairement à une phase normale, la phase stationnaire inverse n'évolue pas au cours du temps, et la qualité de la séparation est maintenue pratiquement constante.

II-3-PHASE MOBILE

Le choix et les conditions d'emploi des solvants comme liquide vecteur ou phase mobile dans la chromatographie liquide- liquide sont fondés sur les considérations Suivantes [2] :

- le solvant doit être chimiquement inerte vis-à-vis de l'échantillon à séparer.
- Il doit être compatible avec le système de détection.
- Il doit être insoluble dans la phase stationnaire.

La dernière considération est très difficile à réaliser car deux liquides non miscibles présentent tout même une légère solubilité entre eux.

L'interaction plus ou moins forte entre la phase mobile et la phase stationnaire normale ou à polarité inversée se répercute sur les temps de rétention des solutés. La polarité de la phase stationnaire permet de distinguer deux situations de principe :
-si la phase stationnaire est polaire, on utilisera une phase mobile peu polaire, et la chromatographie est dite en phase normale ;
-si la phase stationnaire est très polaire, on choisira une phase mobile polaire (le plus souvent des mélanges de méthanol ou d'acétonitrile avec de l'eau), c'est la chromatographie en phase inverse. En modifiant la polarité de la phase mobile, on agit sur les facteurs de rétention K des composés.

Pour minimiser ou diminuer la miscibilité des deux phases, et pour assurer une longue durée de vie aux colonne (éviter la sortie de la phase stationnaire de la colonne durant l'utilisation de cette dernière), il faut prendre quelques précautions particulières, telle que la saturation des phases les unes avec les autres par des contacts préalables. Et cela peut se réaliser en plaçant les deux phases ensemble dans une enceinte fermée (ampoule à décantation par exemple) pendant au moins un jour, en les agitant fréquemment pour les séparer ensuite.

En chromatographie liquide haut performance, pour assurer l'équilibre complet des deux phases, il est recommandé de faire passer tout d'abord la phase mobile à travers une pré colonne de saturation. C'est une petite colonne beaucoup plus chargée en phase stationnaire et qui se place juste avant la colonne, la phase mobile, après la traversée de cette dernière,arrive à la colonne chromatographique déjà saturée, ce qui diminue les risques de détérioration de la phase stationnaire.

Les silices greffées conduisent en général à une perte importante de polarité. Avec une phase greffée, l'ordre d'élution est opposé à celui auquel on est habitué avec les phases normales. Ainsi avec un élution polaire, un composé polaire migre plus vite qu'un composé apolaire. Dans ces conditions les hydrocarbures sont fortement retenus. On réalise des gradients d'élution en diminuant au cours de la séparation la polarité de l'éluant (ex : mélange eau / acétonitrile dont la concentration en acétonitrile va en croissant au cours de l'élution).

II-4 CARACTERISTIQUES PRINCIPALES DE LA CHROMATOGRAPHIE LIQUIDE-LIQUIDE

La chromatographie liquide- liquide, est l'une des méthodes chromatographiques les plus souples, par suite de la possibilité de choisir les phases de partage pour une séparation donnée. Elle permet l'analyse d'une grande variété d'échantillons polaires, et non polaire. Si la phase stationnaire est polaire alors que la phase mobile est peu ou non polaire(ce qui permet la séparation des composés polaires),il s'agit de la chromatographie en phase normale. Par contre si la phase stationnaire est non polaire alors la phase mobile est polaire (ce qui permet la séparation des composés non polaires), il s'agit de phase inverse.

Contrairement à la chromatographie liquide- solide, il n'existe pas un système de partage universel pour tous les solutés, comme le gel de silice. Mais il y a presque une infinité de couples de liquides de partage possibles pour la séparation.

Puisque les analyses se font généralement à température ambiante, avec des supports chimiquement inertes, on peut séparer un grand nombre de solutés réactifs ou dégradables. De plus en chromatographie liquide- liquide, on ne peut pas rencontrer d'effets catalytiques qui proviendraient de l'hydrolyse ou d'autres réaction comme en chromatographie liquide- solide.

Un des avantages pratiques essentiels de la chromatographie liquide- liquide est la très proche reproductibilité de remplissage, c'est-à-dire, la possibilité de fabriquer des colonnes qui donnent des séparations reproductibles.

L'évolution de la technique au cours de ces dernières années, a permis d'obtenir des séparation aussi rapides qu'avec les autres méthodes chromatographies et cela par variation des interactions sélectives qui donnent des valeurs élevées pour le facteur de sélectivité de la colonne .

La chromatographie de partage est une méthode utilisée pour résoudre les problèmes d'analyses quantitatives qui nécessitent des essais de grande précision, et pour l'analyse de composés à des concentrations de quelques ppm.

II-5- GRADIENT D'ELUTION

Pour avoir une bonne résolution par unité de temps pour une large gamme d'échantillons, il est commode d'augmenter la force d'élution durant l'analyse, et cela s'effectue par variation de la composition du liquide vecteur pendant la séparation (en mélangeant deux solvants de polarités différentes), cette méthode s'appelle gradient d'élution.

En conclusion, on peut dire que la chromatographie liquide- liquide, est une des méthodes les plus précises de la chromatographie en général et de la chromatographie en phase liquide en particulier.

II-6-INSTRUMENTATION [3]

La figure 2 représente le schéma de principe d'un chromatographe pour CLHP

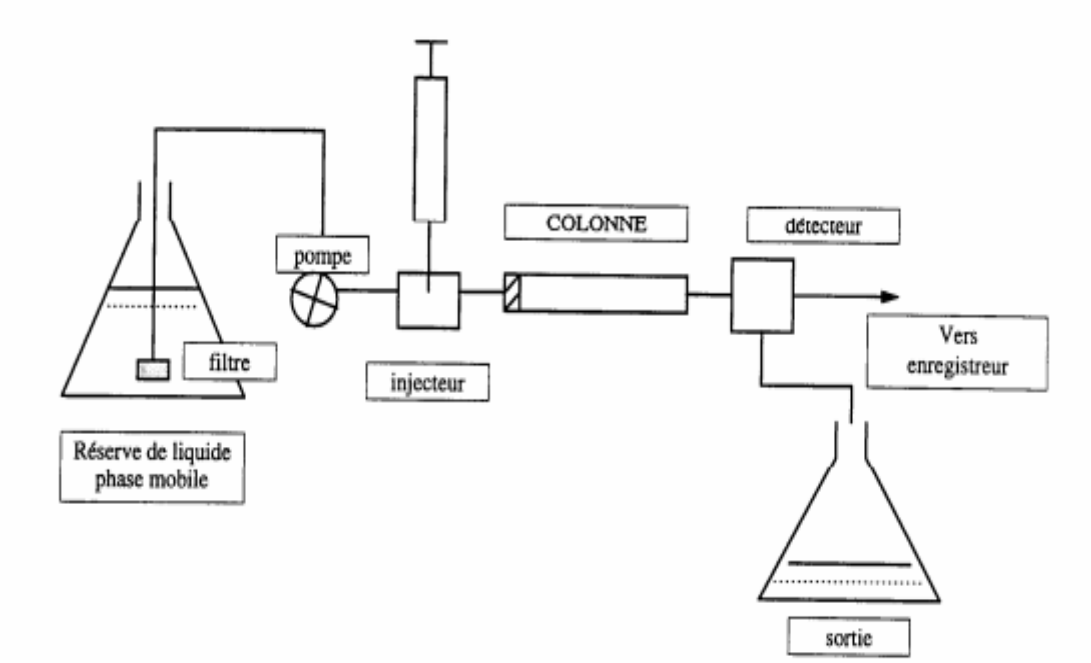


Figure 2 : *principe de fonctionnement de la CLHP*

II-7 DEFFERENTES PARTIE D'UN CHROMATOGRAPHE POUR CLHP

II-7-1- Un réservoir de solvant (éluant)

qui contient la phase mobile en quantité suffisante. Plusieurs flacons d'éluants (solvants de polarités différentes) sont disponibles pour pouvoir réaliser des gradients d'élution (mélange de plusieurs solvants à des concentrations variables)

II-7-2- La pompe :

elle est muni d'un système de gradient permettant d'effectuer une programmation de la nature du solvant. Elle permet de travailler:

- en mode isocratique, c'est-à-dire avec 100% d'un même éluant tout au long de l'analyse
- en mode gradient, c'est-à-dire avec une variation de la concentration des constituants du mélange d'éluants.
- Les pompes actuelles ont un débit variable de quelques ml à plusieurs ml/min.

II-7-3 –vanne d'injection :

C'est un injecteur à boucles d'échantillonnage. Il existe des boucles de différents volumes. Le choix du volume de la boucle se fait en fonction de la taille de la colonne et de la concentration supposée des produits à analyser. Le système de la boucle d'injection permet d'avoir un volume injecté constant, ce qui est important pour l'analyse quantitative.

II-7-4 -La colonne :

Une colonne est un tube construit dans un matériau le plus possible inerte aux produits chimiques, souvent en inox ou en verre. Sa section est constante, de diamètre compris entre 4 et 20 mm pour des longueurs généralement de 15 à 30 cm. Au delà,des pertes de charges exigeraient des pressions de liquide beaucoup trop élevées.

II-7-5 –le détecteur :

Différent type de détecteurs sont possibles. Le détecteur à absorption UV, travaillant à une longueur d'onde fixe mais réglable dans la gamme 190-800 nm , est le plus utilisé en CLHP. Il est constitué d'une cuve à circulation en quartz, d'une capacité d'environ 10µl, traversée en continu par le faisceau U.V.

II-8- LA RETENTION

II-8-1- le temps de rétention

Le principe de séparation chromatographie est fondé sur le partage dynamique des composée à analyser contenus dans une colonne de faible diamètre, entre deux phase non miscibles, l'une des phases qui est stationnaire peut être un liquide immobilisé sur un support inactif (verre), ou un solide doué de propriétés adsorbantes, l'autre phase est mobile, renouvelée constamment et réglée afin de rester constante pendant tout la durée de l'analyse.

Lorsque la solution contenant les composés à séparer est introduite dans le système chromatographique (colonne) elle se retrouve en contact avec une phase stationnaire, choisie préalablement en raison de son affinité pour ces composés, des différences de distribution d'équilibres naissant des interactions moléculaires ' composés – phase stationnaire ' s'établissent, et les composés se retrouvent ainsi retenus, chacun selon son affinité pour cette phase. Mais le renouvellement continu de la phase mobile remet en cause les équilibres précédents et provoque la migration des composés le long de phase stationnaire, et en raison des distributions préférentielles de ces composées, pour l'une ou l'autre phase, il se produit des différences de migration. Ainsi les substances qui on une affinité préférentielle pour la phase stationnaire migrent plus lentement ce qui se traduit par une rétention plus longue, tandis que celles qui préfèrent la phase mobile vont se déplacer plus vite.

Pour qualifier ce phénomène de rétention on utilise le temps de rétention :

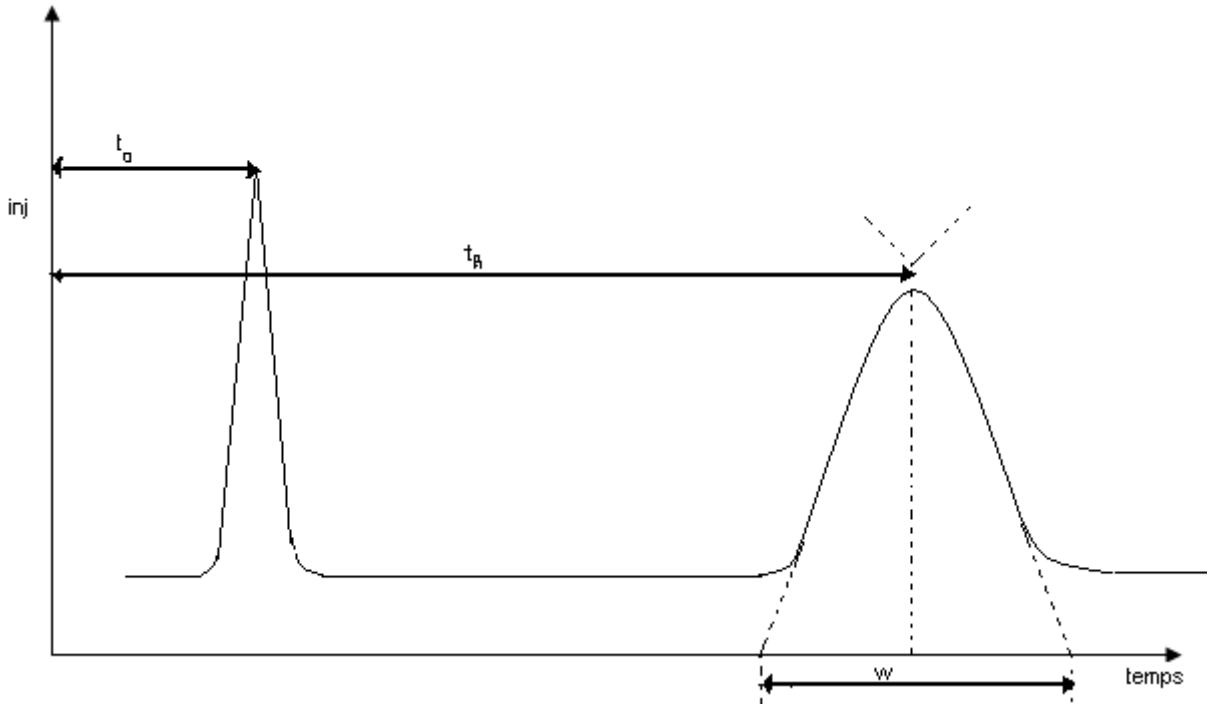


Figure 3 : *chromatogramme d'une substance à un seul composé.*

t_R : temps de rétention brut

w : base de pic appelée élargissement du pic

t_0 : temps de rétention de la phase mobile, ou plutôt d'un composé non retenu.

Le temps de rétention brut, c'est le temps écoulé entre l'introduction du composé d'un échantillon dans la colonne et le moment de l'apparition du sommet du pic.

Le temps de rétention t_0 correspond au temps mis par la phase mobile pour parcourir la colonne (temps mort), ce temps n'est dû qu' à la confection de la colonne. Le temps de rétention net (réduit ou corrigé), c'est la différence entre le temps de rétention brut et le temps mort.

$$t'_R = t_R - t_0 \qquad (1)$$

Le temps de rétention étant fonction de la vitesse de la phase mobile, le volume de rétention V_R s'obtient par le produit du débit de la phase mobile et du temps de rétention

t_R, V_R c'est le volume de la phase mobile nécessaire à élution du composé considéré. Le volume mort s'obtient par le produit du débit de la phase mobile par le temps mort.

II-8-2-Expression du volume de rétention en fonction du coefficient de partage

Pour un procédé chromatographique la formule fondamentale de la rétention est la suivante :

$$V_R = V_M + K V_S \qquad (2)$$

V_S : volume de la phase stationnaire : qui ne contient pas le support inerte, par exemple en chromatographie liquide- liquide le volume de phase stationnaire représente le volume du liquide déposé.

K : coefficient de distribution est linéaire, de la forme :

$$C_S = K\, C_m \qquad (3)$$

C_S : concentration du composé dans la phase stationnaire.

C_M : concentration du composé dans la phase mobile.

Selon cette formule il y a une augmentation proportionnelle des quantités d'échantillon dans la phase mobile et stationnaire, et la représentation graphique de cette équation à une température donnée se traduit par une droite linéaire appelée «isotherme de distribution » **[2]** caractérisant la chromatographie d'élution linéaire qu'on veut obtenir dans tout procédé chromatographique. Donc le coefficient de distribution présente le rapport des concentrations d'échantillon dans la phase stationnaire et la phase mobile.

II-8-3 Expression du volume de rétention en fonction du facteur de capacité K'

$$k' = k \frac{V_S}{V_M} \qquad (4)$$

Ce facteur de capacité est un paramètre important parce qu'il représente la manière dont un composé est retenu le long de la colonne.

D'après les équations (2) et (4) on déduit la relation suivante

$$V_R = V_M(1 + K') \qquad (5)$$

Le volume mort ne dépend que de la confection de la colonne.

Lorsque le facteur de capacité est nul, le composé n'est pas retenu par la phase stationnaire, et l'augmentation du facteur de capacité K' entraîne une rétention du composé. Comme le temps de rétention est proportionnel au volume de rétention, l'influence du facteur de capacité est importante dans la réalisation d'une séparation.

Pour un débit constant la relation (4) peut être formulée de manière suivante :

$$k' = \frac{t_R - t_0}{t_0} \qquad (6)$$

On déduit :

$$t_R = (1+k')t_0 \qquad (7)$$

Cette formule donne une autre définition du facteur de capacité qui est le rapport entre le temps de rétention corrigé d'un composé et le temps mort.

Le temps mort est donné par l'équation :

$$t_0 = \frac{L}{V} \qquad (8)$$

t_0 : Longueur de colonne

V : Vites de la phase mobile

Des expressions (7) et (8) on peut tirer l'équation qui relie le temps de rétention à la longueur de la colonne et la vitesse de phase mobile :

$$t_R = (1+k')\frac{L}{V} \qquad (9)$$

III- LES MODELES QSAR/QSPR

Au cours des décennies passées, les Relations Quantitatives Structure- Activité/ Propriété (QSAR/QSPR) sont devenues un puissant outil théorique, alternatif à la mécanique quantique, pour la description et la prédiction des propriétés des systèmes moléculaires complexes dans différents environnements. L'approche QSAR/QSPR procède de l'hypothèse d'une correspondance univoque entre n'importe quelle propriété physique, affinité chimique, ou activité biologique d'un composé chimique et sa structure moléculaire [4]. Cette dernière peut être représentée par la composition chimique, la connectivité des atomes, la surface d'énergie potentielle, et la fonction d'onde électronique d'un composé. Différents descripteurs moléculaires physico- chimiques reflétant la structure peuvent être déterminés empiriquement ou en utilisant des méthodes théoriques et computationnelles de différentes complexités. Il est à souligner que la connaissance de la constitution chimique exacte et/ou de la structure moléculaire tridimensionnelle des composés chimiques étudiés est un pré- requis à l'application de l'approche QSAR/QSPR.

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

Les descripteurs empiriques peuvent être divisés en deux classes générales (tableau 1), la première reflète les interactions électroniques intramoléculaires **(descripteurs structurels)** alors que la seconde tient compte des interactions intermoléculaires dans les milieux condensés tels que les liquides et les solutions **(descripteurs de solvatation)**

Tableau 1: Classification d'ensemble des descripteurs moléculaires empiriques

classe	Sous- classe
Descripteurs structurels	- Constantes d'induction - Constantes de résonance - Constantes stérique
Descripteurs de solvatation	- Echelles de polarité - Echelles de polarisabilité - Echelles d'acidité - Echelles de basicité - Echelles mixtes

Les descripteurs structurels les plus répandus ont été définis pour quantifier les propriétés d'induction, l'effet mésomère ou de résonance, et les effets stériques des composés chimiques. Les descripteurs de solvatation reflètent les interactions du soluté avec la masse du solvant environnant **(effets de solvant macroscopiques** ou **non spécifiques)**, et les liaisons spécifiques, souvent des liaisons hydrogène entre le soluté et les molécules individuelles de solvant **(effets de solvant spécifiques** ou **microscopiques)**. Les effets de solvant macroscopiques sont quantifiés en utilisant diverses échelles de polarité et de polarisabilité. Les descripteurs des effets de solvant microscopiques impliquent les échelles générales d'acidité et de basicité. Certaines échelles empiriques d'effets de solvant (échelles mixtes) peuvent impliquer en même temps ces deux effets macroscopique et microscopique. Le coefficient de partage octanol/ eau, log P, est le représentant typique de tels descripteurs.

Les descripteurs moléculaires théoriques peuvent, conventionnellement, être répartis en un certain nombre de classes, selon leur complexité ou leur méthode de calcul. Les descripteurs théoriques les plus simples sont des **descripteurs constitutionnels** qui peuvent être construits à partir de l'information sur la composition chimique du composé considéré. Les nombres, absolus et relatifs, des différents types d'atomes et de liaisons chimiques, la masse molaire, et le nombre de différents cycles dans le composé représentent quelques descripteurs constitutionnels typiques. **Les descripteurs**, **ou indices**, **topologiques** décrivent la connectivité des atomes dans la molécule. On a avancé [4] que les indices topologiques pouvaient encoder des interactions moléculaires subtiles et non pas seulement renseigner sur

le degré de ramification des liaisons chimiques ou la distribution de la masse spécifique dans la molécule. **Les descripteurs géométriques** sont obtenus à partir de la structure tri-dimensionnelle des molécules définie par les coordonnées des noyaux atomiques et la grosseur de la molécule représentée, par exemple, par le rayon atomique de Van der Waals. Les molécules de la plupart des composés chimiques possèdent une certaine flexibilité conformationnelle et les surfaces de potentiels moléculaires respectives possèdent de multiples minima locaux. Selon la structure de la molécule, le nombre de ces minima peut être très grand et, par conséquent, il est plutôt difficile de trouver le minimum d'énergie global pour des conditions expérimentales établies.
Evidemment, les descripteurs géométriques peuvent varier de façon significative selon les conformations utilisées dans le calcul de ces descripteurs. Dans une certaine mesure, **les descripteurs théoriques liés à la distribution de charge** peuvent également dépendre de la conformation. Ces descripteurs sont basés sur la structure tri- dimensionnelle et la distribution des charges dans la molécule. Ces dernières peuvent se présenter comme charges atomiques partielles obtenues à partir d'un schéma empirique ou en utilisant des fonctions plus sophistiquées basées sur la fonction d'onde de la molécule calculée par la chimie quantique.

Un certain nombre de **descripteurs quanto- chimiques basés sur les OM** ont été employés dans le développement d'équations QSAR/QSPR. Les plus utilisés sont les énergies des OM frontières, c'est-à-dire, l'énergie calculée de la plus basse orbitale moléculaire inoccupée ($\varepsilon_{LUMO)}$, et l'énergie de la plus haute orbitale moléculaire occupée (ε_{HOMO}), et la différence entre ces énergies. De même, différents indices de réactivité déduits de la théorie de la superdélocalisabilité de Fukui ou d'autres constructions théoriques ont gagné en popularité parmi les chercheurs.

Tous les descripteurs théoriques ne peuvent être strictement classés selon le schéma présenté dans le tableau 2. Par exemple, les indices topographiques sont déduits de l'information contenant à la fois la topologie et la géométrie des molécules. **Les indices électrotopologiques** sont fondés sur la topologie et la distribution de charge alors que les aires de surfaces partielles chargées sont des descripteurs qui encodent à la fois la distribution de charge et la géométrie des molécules. De tels descripteurs peuvent être classés comme **descripteurs moléculaires mixtes ou combinés**.

Les descripteurs moléculaires peuvent être définis pour tout le système moléculaire étudié ou pour n'importe laquelle de ses parties (fragments). Par exemple, la majorité des descripteurs empiriques structurels sont reliés à des fragments moléculaires appelés substituants. En conséquence, les molécules d'une série congénère de composés chimiques sont divisées formellement en deux ou plusieurs fragments qui correspondent à une unité structurale constante Y (c'est-à-dire le centre de réaction) et à des unités structurales variables Xi (les substituants). Les relations QSAR/QSPR sont ainsi présentées comme suit :

$$P = P_0^{(Y)} + \sum_i \sum_k \alpha_{ik}^{(Y)} D_{ik}^{(X)} \qquad (10)$$

Où $P_0^{(Y)}$ est l'ordonnée à l'origine correspondant au fragment moléculaire constant Y, les $D_{ik}^{(X)}$ sont les descripteurs moléculaires de type k pour les fragments variables X_i, et les $\alpha_{ik}^{(Y)}$ sont les coefficients de développement caractéristiques d'une série donnée de composés $X_i Y$.

La plupart des descripteurs théoriques qui apparaissent dans le tableau 2 peuvent être calculés soit pour la molécule entière soit pour un fragment moléculaire pré- défini.

L'hydrophobicité est une propriété physicochimique moléculaire rendant compte de l'affinité relative d'un soluté pour les phases organique et aqueuse respectivement. Par conséquent, l'hydrophobicité encode la plupart des forces intermoléculaires qu'il peut y avoir entre un soluté et un solvant. Par plusieurs évidentes approches [5], il a été démontré que Log P peut être factorisé en un terme de volume ou stérique et un terme électronique ou polaire. Par cette affirmation et du fait que la rétention en chromatographie est le résultat, entre autres, d'un de ces facteurs ou leurs associations avec d'autres ; il apparaît clairement que l'introduction du log P comme prédicteur dans un modèle QSRR est une initiative légitime.

Les valeurs du coefficient de partage octanol –eau utilisée pour ce travail ont été calculées par les deux modèles que nous décrivons brièvement.

Modèle de Ghose-Crippen-Viswanadhan :

Le coefficient de partage octanol –eau (ALOGP) de Ghose, Crippen et Viswanadhan est calculé à partir de l'équation de régression basée sur les contributions à l'hydrophobicité de la molécules des 120 types d'atomes pouvant la composer [6,7]. Chaque atome de n'importe quelle structure est classé comme un des 120 types. Et le log P de cette

structure est estimé par simple sommation des contributions des types d'atomes qui la constituent par l'équation :

$$ALOGP=\sum n_i a_i \qquad (11)$$

où n_i est le nombre d'atomes de type " i " et a_i est la constante d'hydrophobicité correspondante à ce type.

Les coefficients du modèle qu'utilise le logiciel DRAGON, dont on a usé pour le calcule des différents log P usité dans ce travail, ont étés estimés sur un ensemble de calibration de 2648 composés (de log P expérimentaux connus) pris dans la base de données NCI.

Modèle de Moriguchi :

Le coefficient de partage de Moriguchi (MLOGP) est calculé par un modèle qui consiste ne une équation de régression multiple à 13 paramètres structuraux [8,9]. Les coefficients de régression ont étés évalués sur un ensemble d'estimation de 1230 molécules organiques : Aliphatiques, aromatiques, hétérocycliques ; contenant les atomes C, H, N, O, S, P, F, Cl, Br et I. L'équation du modèle de Moriguchi est la suivante :

$$MLOGP = -1{,}041 + 1{,}244(CX)^{0,6} - 1{,}017(NO)^{0,9} + 0{,}406(PRX) - 0{,}145(UB)^{0,8} + 0{,}511(HB) + 0{,}268(POL) - 2{,}215(AMP) + 0{,}912(ALK) - 0{,}392(RNG) - 3{,}684(QN) + 0{,}474(NO2) + 1{,}582\ (NCS) + 0{,}733(BLM) \qquad (12)$$

où par exemple :

CX : Est la somme pondérée des nombres de carbones et d'halogène.

NO : Nombre totale d'azotes et d'oxygène.

HB : Existence, HB=1(0 dans le cas contraire), de liaison hydrogène intramoléculaire.

Le modèle utilisé dans le logiciel DRAGON a été évalué sur un ensemble de 2648 composés, pris à partir de la base de données libre NCI.

Ces descripteurs ne fournissent par assez d'information sur la structure des molécules pour l'élaboration de modèles prédictifs plus complexes ; il est nécessaire d'ajouter d'autres types de descripteurs.

Tableau 2 : Classification générale des descripteurs moléculaires théoriques

classe	Sous- classe
Descripteurs constitutionnels	- Dénombrement des atomes ou des liaisons. - Descripteurs basés sur les masses atomiques.
Descripteurs topologiques	- Indices topologiques (connectivité). - Descripteurs théoriques d'information. - Descripteurs topochimiques.
Descripteurs géométriques	- Descripteurs liés à la distance. - Descripteurs liés à l'aire de la surface. - Descripteurs liés au volume. - Descripteurs du champ stérique moléculaire.
Descripteurs liés à la distribution de charge	- Charges atomiques partielles. - Moments électriques moléculaires - Polarisabilités moléculaires. - Descripteurs du champ électrique moléculaire.
Descripteurs liés aux orbitales moléculaires	- Energie des OM frontières - Ordres de liaison - Indices de réactivité de Fukui.
Descripteurs température dépendants	- Fonctions thermodynamiques. - Descripteurs facteurs de Boltzmann pondérés.
Descripteurs de solvatation	- Energie électrostatique de solvatation. - Energie de dispersion de solvatation. - Enthalpie libre de formation de cavité. - Descripteurs de liaison hydrogène. - Entropie de solvatation. - Descripteurs d'énergie de solvatation linéaire théorique.
Descripteurs mixtes	- Descripteurs topographiques. - Descripteurs électrotopologiques. - Descripteurs de la charge partielle de l'aire de la surface.

IV- METHODES UTILISEES POUR LE DEVELOPPEMENT DE MODELES QSAR/QSPR

IV-1- Introduction

L'application pratique des gammes des descripteurs moléculaires dans le développement de modèles QSAR/QSPR n'est pas une tâche aisée [4]. Tout d'abord, un très grand nombre (>3000) de descripteurs moléculaires, de différentes complexités et de conceptions diverses ont été imaginés et proposés au cours des (60 dernières) années. Ensuite, pendant ce temps, aucune règle stricte n'a été établie, ni même proposée, pour la sélection de descripteurs adaptés parmi la myriade de descripteurs disponibles. Ce choix a souvent été basé sur l'intuition chimique des chercheurs, ou en se pliant à la tradition.

Une autre difficulté dans la sélection des descripteurs QSAR/QSPR découle de la non standardisation des gammes de descripteurs. Les gammes empiriques des constantes d'induction, de résonance et d'effet stérique des constituants, ou les échelles empiriques d'effets de solvant comportent des erreurs intrinsèques liées aux erreurs respectives des mesures expérimentales. Par ailleurs, les méthodes quanto- mécaniques appliquées aux calculs des descripteurs moléculaires et aux distributions de charges liés aux OM sont souvent basées sur différents paramètres semi- empiriques, ou l'utilisation de différents ensembles de base dans les calculs ab- initio. Naturellement, un descripteur construit à l'aide de différentes méthodes expérimentales ou théoriques, pour divers composés, ne peut être utilisé pour le calcul d'un modèle QSAR/QSPR unique. Une approche systématique pour la sélection de gammes de descripteurs pour le calcul de modèles QSAR/QSPR est basée sur la discrimination statistique entre de larges ensembles de descripteurs.

Dans ce qui suit nous passerons en revue diverses approches utilisées pour le développement des " meilleures" équations QSPR dans de grands espaces de descripteurs.

En dernier ressort, les modèles QSAR/QSPR peuvent être développés selon des modèles mathématiques différents, généralement en relation avec l'analyse statistique multivariée. Le premier modèle, et le plus largement utilisé, consiste en une équation (multi) linéaire obtenue par régression des données expérimentales en fonction d'un ensemble de

descripteurs pré- sélectionnés (ou d'un seul), en utilisant la méthode des moindres carrés ordinaires (MCO). Dans quelques cas, les modèles physiques ou chimiques connus du phénomène étudié laissent prévoir certaines formes mathématiques non linéaires (exponentielles ou logarithmiques) de la dépendance entre les données expérimentales et les descripteurs moléculaires. Les modèles QSAR/QSPR peuvent alors être établis à l'aide de la technique de régression par les moindres carrés non linéaires. D'autres modèles ont été développés en utilisant l'analyse factorielle ou l'analyse en composantes principales. L'intérêt de ces méthodes est qu'elles évacuent le problème de multicolinéarité inhérent aux méthodes de régression linéaires. Cependant, l'interprétation des équations QSAR/QSPR est alors

entravée par la nature formelle des facteurs ou des composantes principales. Une alternative aux méthodes très classiques de régression linéaire multiple (RLM) et d'analyse en composantes principales (ACP) est la technique de régression par les moindres carrés partiels (MCP ou PLS) [10-15].

V-METHODOLOGIE

V-1- Calcul et sélection des descripteurs moléculaires*:*

Nous avons utilisé le logiciel de modélisation moléculaire Hyperchem^{TM} 6.03 [16] pour représenter les molécules puis, à l'aide de la méthode semi-empirique AM1 (Austin Model), obtenir les géométries finales. Tous les calculs ont été menés dans le cadre du formalisme RHF (pour restricted Hartree-Fock ou formalisme de Hartee-Fock avec contrainte de spin) 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 version 5.3 [17] pour le calcul de 1664 descripteurs appartenant à différentes classes. Les descripteurs d'un même groupe, à valeur constante (écarts types inférieurs à 0,001), et ceux hautement corrélés ($R \geq 0,95$) ont été exclus.

V-2- Développement et validation du modèle.

L'analyse de régression multi linéaire a été réalisée avec le logiciel MOBYDIGS [18] en utilisant la méthode des moindres carrés ordinaires.

La qualité de l'ajustement a été évaluée par le coefficient de détermination, R^2, et l'écart quadratique moyen calculé sur l'ensemble de calibration :

$$EQMC = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \qquad (13)$$

y_i et $\hat{y}_i$ étant les valeurs mesurées et calculées de la variable dépendante.

Les techniques de validation croisée ont été exploitées pour l'évaluation de la prédiction interne $\left(Q^2_{LMO}; \text{bootstrap}\right)$, et de la robustesse $\left(Q^2_{LOO}\right)$ du modèle,et le Y-scarambling.

La validation croisée par "leave-one-out" (LOO) [19] consiste a recalculer le modèle sur (n-1) objets, et utiliser le modèle ainsi obtenu pour prédire la valeur de la variable dépendant du composé écarté. Le procédé est répété pour chacun des objets de l'ensemble de calibration. La somme des carrés des erreurs de prédiction (désignée par l'acronyme PRESS pour Predictive Residual Sum of Squares) est une mesure de la dispersion des estimations. On

l'utilise pour définir le coefficient de prédiction $\left(Q^2_{LOO}\right)$, et l'écart quadratique moyen de prédiction (ou EQMP) :

$$Q^2_{LOO} = 1 - \frac{PRESS}{SCT} = 1 - \frac{\sum_1^n \left(y_i - \hat{y}_{i/i}\right)^2}{\sum_1^n \left(y_i - \bar{y}\right)^2} \quad (14)$$

$$EQMP = \sqrt{\frac{1}{n} PRESS} \quad (15)$$

$y_{i/i}$ désignant la réponse du i-ème objet estimée en utilisant un modèle obtenu sans faire intervenir cet i-ème objet, et $\bar{y}$ la valeur moyenne des n observations; la sommation porte sur l'ensemble des composés de calibration; SCT est la somme des carrés totale.

Une valeur de $Q^2_{LOO} > 0{,}5$ est, en général, considérée comme satisfaisante et une valeur $Q^2_{LOO} > 0{,}9$ est excellente [20].

En fait, si une forte valeur de Q^2 est une condition nécessaire d'une possible capacité prédictive élevée d'un modèle, cette condition seule n'est pas suffisante.

Pour éviter une surestimation de la capacité prédictive du model nous avons également appliqué la procédure "leave-more-out" (LMO), en excluant jusqu'à 50 % des objets à chaque étape $\left(Q^2_{LMO/50}\right)$. La procédure est répétée 8000 fois dans le présent travail.

Dans la technique de validation par bootstrap on simule- de nouveaux échantillons de taille (n), par tirages aléatoires avec remise. De cette façon l'ensemble de calibration, qui conserve sa taille initiale (n), se compose, en général, d'objets répètes, l'ensemble d'évaluation rassemblant les objets exclus [21]. Le modèle est calculé sur l'ensemble de calibration et les réponses prédites pour l'ensemble d'évaluation. Tous les carrés des différences entre valeurs prédites et réelles des objets de l'ensemble d'évaluation sont collectés dans le PRESS. Cette procédure de construction des ensembles de calibration et d'évaluation est répétée plusieurs milliers de fois (5000 dans cette étude), les PRESS sont additionnés, et une capacité de prédiction moyenne calculée [22].

L'application du modèle, calculé sur l'ensemble de calibration, aux observations de l'ensemble de validation, permet de vérifier de manière fiable la capacité prédictive du modèle obtenu.

L'équation (16) permet le calcul de Q^2_{ext} :

$$Q^2_{ext} = 1 - \frac{\sum_{i=1}^{n_{ext}} (\hat{y}_{i/i} - y_i)^2 / n_{ext}}{\sum_{i=1}^{n_{tr}} (y_i - \bar{y}_{tr})^2 / n_{tr}} = 1 - \frac{PRESS / n_{ext}}{SCT / n_{tr}} \qquad (16)$$

L'indice (ext) se rapportant aux objets de l'ensemble de validation externe (ou à ceux de l'ensemble d'évaluation obtenu par bootstrap), et l'indice tr à ceux de l'ensemble de calibration (training set).

$$EQMP_{ext} = \sqrt{\frac{1}{n_{ext}} \sum_{i=1}^{n_{ext}} (y_i - \hat{y}_i)^2} \qquad (17)$$

La somme portant sur les objets de l'ensemble test (n_{ext})

Notre écriture du programme pour l'algorithme DUPLEX s'appuie sur sa description faite par R.D. Snee [23] qui mit au crédit de R.W. Kennard son développement. L'édition et l'exécution de ce programme utilisent l'environnement du logiciel MATLAB [24].

L'algorithme DUPLEX commence par une transformation des données (variables explicatives seulement) qui est faite comme suit :

1. Standardisation des données par la formule :

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j \sqrt{(n-1)}}$$

s_j : Ecart-type du *j ème* régresseur.

$\bar{x}_j$: Moyenne du *j ème* régresseur

x_{ij} : Valeur du régresseur *j* pour la *i ème* observation.

n : Nombre d'observations.

Pour les i = 1, 2, 3… n et j=1, 2, 3… k ; les éléments de la matrice $\boldsymbol{Z}$ sont ainsi calculés.

2. On calcul la matrice symétrique définie positive $\boldsymbol{A=Z'Z}$.
3. La factorisation de Cholesky peut être maintenant faite pour le calcul de la matrice $\boldsymbol{B}$ tel que $\boldsymbol{A=B'B}$.

4. Les nouvelles coordonnées des n observations c'est-à-dire les $\boldsymbol{w}$ sont obtenues par le calcul de la matrice $\boldsymbol{W}$:

$$\boldsymbol{W=ZB^{-1}}$$

Une fois la transformation faite ont utilise les points orthonormalisés pour calculer les distances euclidiennes entre les paires possibles de points. La distribution des points ou leur éclatement en un ensemble de prédiction et un second d'estimation se fait en utilisant ces distances de la manière suivante :

i. La distance la plus grande correspond à la paire de points que l'algorithme classe comme points d'estimations, ces points (E1 et E2) sont ensuite éliminés.

ii. Les deux points (P1 et P2) les plus éloignés, dans les $(n-2)$ restants, sont placés dans l'ensemble de prédiction puis automatiquement éliminés des $(n-2)$ points.

iii. L'algorithme tient compte seulement des distances des $(n-4)$ points par rapport aux points d'estimation préalablement choisis (E1 et E2).Chaque point M à deux distances une par rapport à E1, ME1, l'autre, ME2, est celle à E2 ; mais M sera caractérisé par la plus petite distance entre elles. La plus grande distance entre ces $(n-4)$ petites distances, qui caractérisent ces $(n-4)$ points, identifie le point E3 qui est par conséquent le plus éloigné de la paire (E1, E2) et est, subséquemment à cela, classé avec eux dans l'ensemble d'estimation et éliminé des $(n-4)$ point.

iv. Pour le choix du point P3 de l'ensemble de prédiction, qui sera éliminé des $(n-5)$ observations avant de procéder à l'étape (v); l'algorithme continue et refait l'étape (iii) mais en utilisant cette fois la paire (P1 et P2) comme référence pour les $(n-5)$ points restants.

v. Le processus computationnel se poursuit en plaçant les points alternativement dans un ensemble ou dans l'autre. Chaque point (Ei ou Pi) assigné à un ensemble sera éliminé et contribuera au choix du point suivant, car la distance de ce point (Ei ou Pi) au ($\big(n-(i+3)\big)$ ou $\big(n-\big(i+4\big)\big)$ respectivement) points restants sera prise en compte

pour caractériser ces derniers. Par exemple ; le point P3 jouera un rôle, avec P1 et P2 bien sure, pour trouver le point P4 parmi les $(n-7)$ points restants.

Les données en main peuvent être ainsi fractionnées, et dans n'importe quel ratio, en un ensemble de prédiction et un autre d'estimation en spécifiant le nombre de point requis que nous jugerons convenable à notre étude.

Pour Snee des précautions particulières doivent être prises avant de scinder les données par cette procédure :

- Le nombre d'observations n doit être supérieur ou égale à $(2k+26)$ si l'ont veut des ensembles d'égales dimensions.
- L'ensemble de prédiction doit contenir au moins 15 observations pour un contrôle rigoureux du pouvoir prédictif du modèle par les statistiques usuelles (Q^2_{ext}, $SDEP_{ext}$).
- Des points répliques d'autres points, ou des points étant des proches voisins, doivent en être purgée nos données d'origines avant tout traitement par DUPLEX.

I- MATERIEL ET METHODE

Le chromatographe utilisé est un système Philips (Pye Unicam) constitué d'une pompe PU 4010, d'un injecteur Rhéodyne 7125, d'un détecteur UV / visible à longueur d'onde variable PU 4200 et d'un enregistreur PM 8251. Une boucle de 20 µL est remplie avec une seringue Hamilton de 25 µL.

Une colonne de remplissage en inox (L : 25 cm ; diam. Int. = 4,6 mm) contenant des radicaux octadécyl-silyle (partisil ODS), dont le diamètre des particules support est 10 µm, a été utilisée pour les analyses. La colonne est placée dans une enveloppe en verre où circule de l'eau thermostatée, ce qui permet de fixer la température désirée à 0,1 °C près.

Les analyses ont été réalisées en régime isochratique avec des phases mobiles constituées de mélanges (méthanol + eau) dans les rapports (V : V) 15 : 85, 25 : 75, 50 : 50, 70 : 30 et 85 : 15, pour un débit réglé à 2 mL / min.

L'effet de la température sur la rétention a été examiné pour l'intervalle 12 – 52 °C (285 – 325 K) en faisant varier la température par bonds de 10 °, la phase mobile étant un mélange volume à volume méthanol – eau.

Nous avons pris pour t_0, le temps de rétention de ^2H_2O qui absorbe à 190 nm.

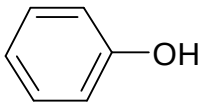
Le facteur de capacité k' des phénols est calculé à partir de la relation :

$$k' = \frac{t_R - t_0}{t_0} = \frac{t_R}{t_0} - 1 \qquad (18)$$

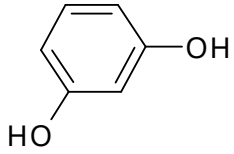
t_R étant le temps de rétention du soluté et t_0 le temps de rétention nulle .

Les structures des 10 phénols considérés sont représentées dans la figure 4.

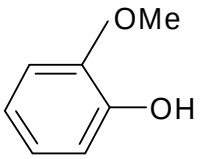
Les valeurs de k' mesurées pour différents ratios volumiques des constituants de la phase mobile et pour différentes températures (Tableau 2). Le tableau 3 réunis les valeurs mesurées des cœfficients de partage des phénols considérés, et calculées selon les deux méthodes développées dans la partie théorique.



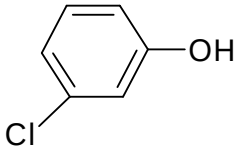
A/ Phénol



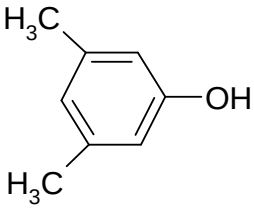
B/Résorcinol



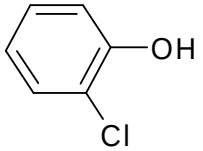
C/Gaïacol



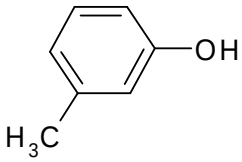
D/3-Chlorophénol



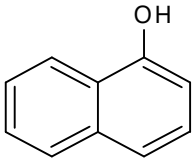
E/*m*-Xylénol



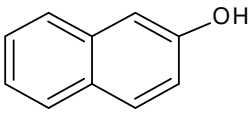
F/ 2-Chlorophénol



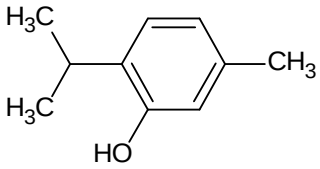
G/ *m*-Crésol



H/ naphthalen-1-ol



I/naphthalen-2-ol



J/ Thymol

Figure 4 – *Noms et structures des phénols étudiés*

Tableau 3 : Valeurs de k' mesurées en fonction de la température (t) et de la fraction volumique (x) du méthanol.

	t = 22 °C ; x(méthanol)					**x = 0,50 ; t (°C)**			
	0,15	**0,25**	**0,50**	**0,70**	**0,85**	**12**	**32**	**42**	**52**
A/ Phénol	3,0568	2,3616	1,1089	0,7391	0,5019	1,3940	0,9976	0,8754	0,8262
B/Résorcinol	1,6876	1,2624	0,7174	0,5067	0,4799	0,7011	0,6373	0,5863	0,5435
C/Gaïacol	5,604	3,4071	1,2433	0,7386	0,5909	1,3463	1,1047	0,9847	0,9137
D/3-Chlorophénol	9,9204	6,5517	2,0948	1,0122	0,5432	2,4284	1,7895	1,4648	1,3192
E/m-Xylénol	-	7,78	-	1,036	0,529	2,556	1,939	1,595	1,451
F/2-Chlorophénol	6,6333	5,3026	1,6425	0,8191	0,5946	2,1673	1,4484	1,2204	1,1161
G/ m-Crésol	6,9065	4,6444	1,5244	0,7955	0,5896	1,7142	1,3514	1,1562	1,0629
H/naphthalen-1-ol	-	-	3,0188	-	-	3,3968	2,3877	1,8902	1,607
I/naphthalen-2-ol	-	-	2,7749	-	-	3,3064	2,2553	1,8912	1,649
J/ Thymol	-	-	5.0688	-	-	6.1097	4.2742	3.3963	3.0279

Tableau 4 - Valeurs de MLOGP et ALOGP calculées.

	MLOGP	ALOGP
A) Phénol	1,506	1,563
B) Résorcinol	0,893	1,295
C) Gaïacol	1,246	1,546
D) 3-Chlorophénol	2,127	2,227
E) *m*-Xylénol	2,193	2,535
F) 2-Chlorophénol	2,127	2,227
G) *m*-Crésol	1,859	2,049
H) naphthalen-1-ol	2,637	2,471
I) naphthalen-2-ol	2,637	2,471
J) J/ Thymol	2,813	3,243

Pour les deux modèles, que nous allons développer et discuter par la suite ; des ensembles de validation de 16 observations ont été, choisis de manières différentes : un choix aléatoire et un choix par DUPLEX. La représentation, sur le même graphique, des deux

ensembles (calibration et validation) des deux éclatements (Aléatoire et par DUPLEX) à chaque fois permet d'apprécier les représentations des observations et d'en tirer, certaines remarques et conclusions. Les valeurs des w1, w2, w3 des différents descripteurs (MLOGP et ALOGP) ainsi que les valeurs des variables d'origines (T, X, MLOGP, ALOGP) figurent dans l'annexe de ce mémoire.

II- Modélisation du facteur de capacité par T, X et ALOGP :

II-1 Représentation des répartitions :

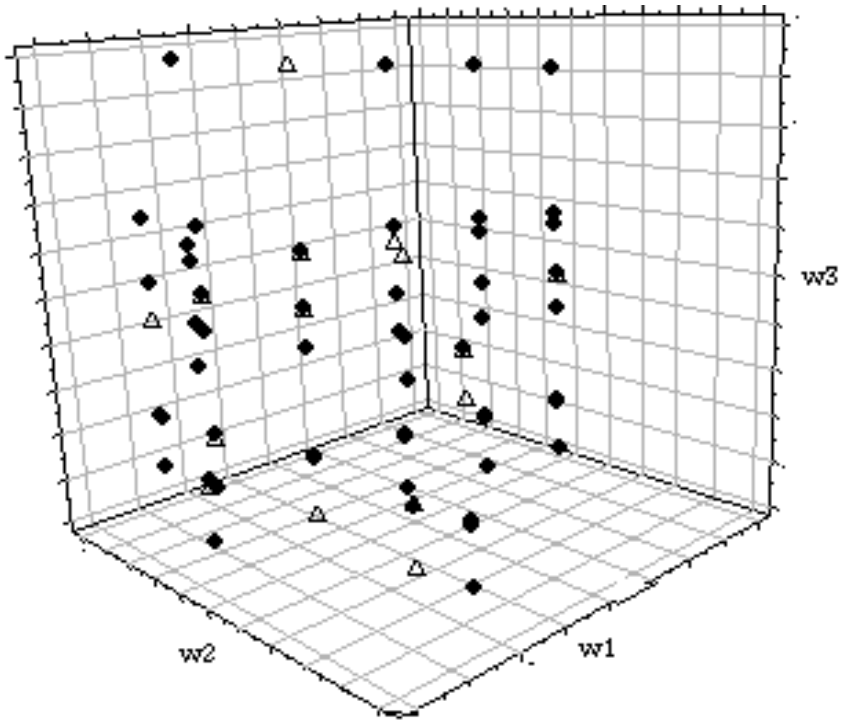


Figure 5 : *Représentation du choix aléatoire*

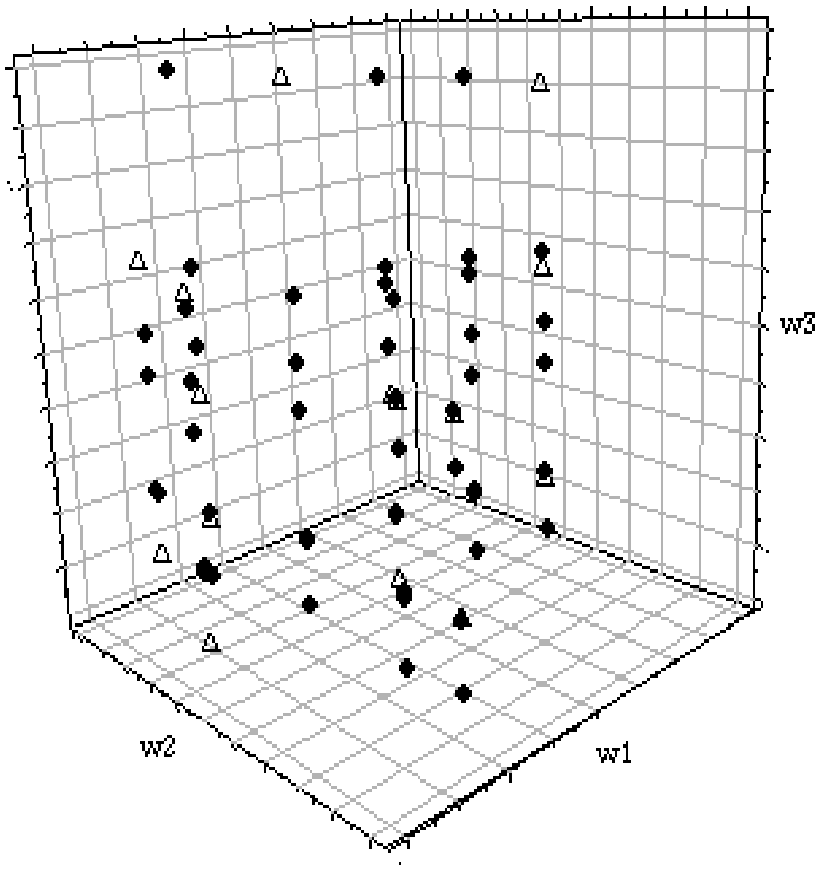


Figure 6 : *Représentation du choix par DUPLEX*

Dans les deux figures les observations de l'ensemble de calibration sont représentées par des cercles, et ceux de validations le sont par des triangles.

Tableau 5: Corrélations entre les variables pour les deux choix (cas de ALOGP)**,** obtenus avec le logiciel MINITAB [23]

	Choix aléatoire	**Choix par DUPLEX**
X	-0,742 (0,000)	-0,733 (0,000)
T	-0,196 (0,133)	-0,267 (0,039)
ALOGP	0,520 (0,000)	0,508 (0,000)

II-2 Analyse de régression :

La régression linéaire multiple (RLM) a été utilisée pour calculer les modèles sur les ensembles de calibration (de 60 observations, différents selon le choix); les équations suivantes résument ces modèles:

L'équation de régression du choix par DUPLEX

$$\log k = 2{,}40(\pm 0{,}2752) - 0{,}00746(\pm 0{,}0009056)\, T - 1{,}44(\pm 0{,}06579)\, X + 0{,}375(\pm 0{,}02283)\, ALOGP \qquad (19)$$

L'équation de régression du choix aléatoire

$$\log k = 2{,}28(\pm 0{,}2784) - 0{,}00688(\pm 0{,}0009105)\, T - 1{,}45(\pm 0{,}07116)\, X + 0{,}348(\pm 0{,}02317)\, ALOGP \qquad (20)$$

Les 16 observation écartées lors du calcul ont servi à la validation statistique externe par application de ces modèles à la prédiction des 16 valeurs de log k. Les paramètres statistiques de cette validation ($Q^2_{ext,}$ $EQMP_{ext}$) ainsi que les valeurs des statistiques pour l'ensemble de calibration sont regroupés dans le tableau 2 pour les deux répartitions.

Tableau 6 : Paramètres statistiques pour ALOGP

		Choix aléatoire	Choix par DUPLEX
n_{tr}		60	
n_{ext}		16	
F		217,28	255,06
R^2		92,09	93,18
R^2_{adj}		91,67	92,82
Q^2_{LOO}		90,37	91,8
Q^2_{LMO}	10%	90,27	91,71
	20%	90,15	91,60
	30%	90,98	92,34
	40%	91,77	93,03
	50%	89,37	91,03
Q^2_{BOOT}		89,4	90,99
Q^2_{ext}		91,8	87,69
S		0,09	0,08
EQMC		0,09	0,08
EQMP		0,10	0,09
$EQMP_{ext}$		0,09	0,11

II-3 Analyse des résidus :

Le tableau 7 réuni les valeurs du log k expérimentales, calculées ou prédites des deux modèles ainsi que les leviers, et les résidus de prédiction standardisés.

Tableau 7 : Compositions des ensembles de calibration (CAL) et de test (VAL) obtenues selon un choix aléatoire, ou à l'aide de l'algorithme DUPLEX et certaines mesures statistiques des deux modèles. (Cas de ALOGP)

			CHOIX ALEATOIRE				CHOIX PAR DUPLEX			
i	Code	$logk_{exp}$	Statut	$logk_{calc-pred}$	h_{ii}	e_{istd}	Statut	$logk_{calc-pred}$	e_{istd}	h_{ii}
1	Ap50	0,1443	CAL	0,1398	0,056	-0,052	CAL	0,1365	-0,0981	0,058
2	Aq15	0,4853	CAL	0,5926	0,113	1,3749	VAL	0,5817	1,1768	0,102
3	Aq25	0,3732	VAL	0,4475	0,076	0,8274	CAL	0,4375	0,8289	0,07
4	Aq50	0,0449	CAL	0,071	0,035	0,2949	CAL	0,0619	0,208	0,036
5	Aq70	-0,1313	CAL	-0,2054	0,058	-0,868	CAL	-0,2117	-1,0293	0,066
6	Aq85	-0,2994	CAL	-0,4231	0,104	-1,5615	CAL	-0,4281	-1,7913	0,116
7	Ar50	-0,001	CAL	0,0023	0,033	0,0371	CAL	-0,0127	-0,1422	0,036
8	As50	-0,0578	CAL	-0,0665	0,05	-0,1009	CAL	-0,0873	-0,3729	0,058
9	At50	-0,0829	CAL	-0,1353	0,086	-0,642	CAL	-0,1619	-1,0727	0,102
10	Bp50	-0,1542	CAL	0,0464	0,074	2,4094	VAL	0,0361	2,2927	0,079
11	Bq15	0,2273	CAL	0,4993	0,131	3,5939	CAL	0,4813	3,5781	0,123
12	Bq25	0,1012	VAL	0,3542	0,094	2,8442	CAL	0,3371	3,1475	0,091
13	Bq50	-0,1442	VAL	-0,0224	0,054	1,341	CAL	-0,0385	1,3393	0,059
14	Bq70	-0,2952	CAL	-0,2988	0,078	-0,043	CAL	-0,3121	-0,2247	0,09
15	Bq85	-0,3188	CAL	-0,5165	0,125	-2,5823	VAL	-0,5285	-2,613	0,139
16	Br50	-0,1957	CAL	-0,0911	0,054	1,2159	VAL	-0,1131	0,9839	0,059
17	Bs50	-0,2319	CAL	-0,1599	0,073	0,8626	CAL	-0,1877	0,5814	0,083
18	Bt50	-0,2648	CAL	-0,2287	0,11	0,4606	CAL	-0,2623	0,0355	0,127
19	Cp50	0,1291	VAL	0,1339	0,056	0,0521	CAL	0,1301	0,0126	0,059
20	Cq15	0,7485	CAL	0,5867	0,114	-2,0746	CAL	0,5754	-2,36	0,104
21	Cq25	0,5324	CAL	0,4416	0,078	-1,0965	CAL	0,4311	-1,3081	0,071
22	Cq50	0,0946	CAL	0,0651	0,036	-0,3331	CAL	0,0555	-0,4778	0,037
23	Cq70	-0,1316	VAL	-0,2114	0,059	-0,8797	VAL	-0,2181	-1,0345	0,066
24	Cq85	-0,2285	CAL	-0,429	0,105	-2,5355	CAL	-0,4345	-2,8731	0,118
25	Cr50	0,0432	CAL	-0,0037	0,034	-0,5289	CAL	-0,019	-0,7626	0,037
26	Cs50	-0,0067	CAL	-0,0724	0,051	-0,7616	CAL	-0,0936	-1,102	0,059
27	Ct50	-0,0392	CAL	-0,1412	0,088	-1,253	VAL	-0,1682	-1,5742	0,102
28	Dp50	0,3211	CAL	0,3024	0,024	-0,2084	CAL	0,3107	-0,1244	0,023
29	Dp50	0,3853	CAL	0,3711	0,048	-0,1633	CAL	0,3853	0,0002	0,046
30	Dq15	0,9965	CAL	0,824	0,105	-2,1815	CAL	0,8306	-2,2317	0,096
31	Dq25	0,8164	CAL	0,6789	0,068	-1,6363	VAL	0,6863	-1,5512	0,061
32	Dq70	0,0053	CAL	0,0259	0,047	0,2375	CAL	0,0371	0,3974	0,049
33	Dq85	-0,265	CAL	-0,1917	0,092	0,9061	CAL	-0,1793	1,1559	0,097
34	Dr50	0,2527	CAL	0,2336	0,019	-0,2107	CAL	0,2361	-0,1978	0,021
35	Ds50	0,1658	CAL	0,1648	0,033	-0,0107	CAL	0,1616	-0,0519	0,04
36	Dt50	0,1203	VAL	0,0961	0,065	-0,2685	CAL	0,087	-0,4386	0,082
37	Ep50	0,4076	CAL	0,4785	0,063	0,8367	VAL	0,5008	1,112	0,061
38	Eq25	0,891	VAL	0,7862	0,082	-1,1703	CAL	0,8017	-1,1656	0,078
39	Eq70	0,0154	CAL	0,1332	0,06	1,3836	CAL	0,1526	1,7465	0,062

Tableau 7 : suite et fin.

			CHOIX ALEATOIRE				CHOIX PAR DUPLEX			
i	Code	$logk_{exp}$	Statut	$logk_{calc-pred}$	h_{ii}	e_{istd}	Statut	$logk_{calc-pred}$	e_{istd}	h_{ii}
40	Eq85	-0,2765	CAL	-0,0844	0,104	2,4249	VAL	-0,0638	2,603	0,107
41	Er50	0,2876	CAL	0,3409	0,031	0,5982	CAL	0,3516	0,7794	0,034
42	Es50	0,2028	CAL	0,2721	0,043	0,7933	CAL	0,277	0,9313	0,053
43	Et50	0,1617	CAL	0,2034	0,075	0,5012	CAL	0,2024	0,5459	0,094
44	Fp50	0,3359	VAL	0,3711	0,048	0,3863	CAL	0,3853	0,6136	0,046
45	Fq15	0,8217	CAL	0,824	0,105	0,0286	CAL	0,8306	0,1189	0,096
46	Fq25	0,7245	VAL	0,6789	0,068	-0,5054	VAL	0,6863	-0,4554	0,061
47	Fq50	0,2155	CAL	0,3024	0,024	0,9643	CAL	0,3107	1,1394	0,023
48	Fq70	-0,0867	VAL	0,0259	0,046	1,2336	CAL	0,0371	1,5435	0,049
49	Fq85	-0,2258	CAL	-0,1917	0,092	0,4207	CAL	-0,1793	0,6269	0,097
50	Fr50	0,1609	CAL	0,2336	0,019	0,8008	CAL	0,2361	0,8976	0,021
51	Fs50	0,0865	CAL	0,1648	0,033	0,8813	CAL	0,1616	0,9232	0,04
52	Ft50	0,0477	CAL	0,0961	0,066	0,573	CAL	0,087	0,5163	0,082
53	Gp50	0,2341	CAL	0,3091	0,045	0,8606	CAL	0,3186	1,0453	0,044
54	Gq15	0,8393	VAL	0,762	0,1	-0,872	CAL	0,7639	-1,0069	0,092
55	Gq25	0,6669	CAL	0,6169	0,065	-0,5929	CAL	0,6196	-0,5981	0,058
56	Gq50	0,1831	CAL	0,2404	0,022	0,6332	CAL	0,244	0,7264	0,02
57	Gq70	-0,0994	CAL	-0,0361	0,044	0,7248	VAL	-0,0296	0,8266	0,047
58	Gq85	-0,2294	VAL	-0,2538	0,088	-0,2725	CAL	-0,246	-0,2222	0,096
59	Gr50	0,1308	CAL	0,1716	0,017	0,4483	VAL	0,1694	0,4512	0,019
60	Gs50	0,063	CAL	0,1028	0,032	0,4471	CAL	0,0949	0,3906	0,039
61	Gt50	0,0265	CAL	0,034	0,066	0,0894	CAL	0,0203	-0,0819	0,082
62	Hp50	0,4798	CAL	0,3874	0,034	-1,0417	CAL	0,4768	-0,6862	0,058
63	Hp50	0,5311	CAL	0,4562	0,059	-0,8783	CAL	0,4022	-0,9441	0,033
64	Hr50	0,378	VAL	0,3186	0,027	-0,6442	CAL	0,3276	-0,6101	0,03
65	Hs50	0,2765	CAL	0,2498	0,04	-0,3035	CAL	0,253	-0,2933	0,049
66	Ht50	0,206	CAL	0,1811	0,072	-0,2986	VAL	0,1784	-0,3346	0,089
67	Ip50	0,5194	CAL	0,4562	0,059	-0,741	CAL	0,4768	-0,5382	0,058
68	Iq50	0,4432	VAL	0,3874	0,034	-0,6081	CAL	0,4022	-0,4993	0,033
69	Ir50	0,3532	VAL	0,3186	0,027	-0,3753	CAL	0,3276	-0,3101	0,03
70	Is50	0,2767	CAL	0,2498	0,04	-0,3061	VAL	0,1784	-0,4703	0,089
71	It50	0,2172	CAL	0,1811	0,072	-0,4327	CAL	0,253	-0,2962	0,049
72	Jp50	0,786	CAL	0,7251	0,142	-0,8196	CAL	0,7661	-0,2931	0,148
73	Jq50	0,7049	VAL	0,6564	0,111	-0,551	VAL	0,6915	-0,1654	0,119
74	Jr50	0,6309	CAL	0,5876	0,102	-0,5445	CAL	0,6169	-0,1942	0,116
75	Js50	0,531	CAL	0,5188	0,111	-0,1557	CAL	0,5423	0,1615	0,132
76	Jt50	0,4811	CAL	0,45	0,139	-0,4167	VAL	0,4677	-0,1704	0,169

II-4 Diagrammes de Williams :

Le diagramme de Williams permet de visualiser les valeurs des résidus de prédiction standardisés des observations en fonction de leurs influences, ces derniers dont tous inférieurs à la valeur critique $h^* = \frac{3\times(p+1)}{n} = \frac{3\times(3+1)}{60} = 0,2$. Dans les deux répartitions on n'a pas de points influents. Pour le choix aléatoire l'observation Bq15 est aberrante, et avec Bq25, l'est aussi pour le choix fait avec DUPLEX. Dans les deux cas ces observations ont des valeurs de e_{istd} supérieurs en valeur absolue à 3.

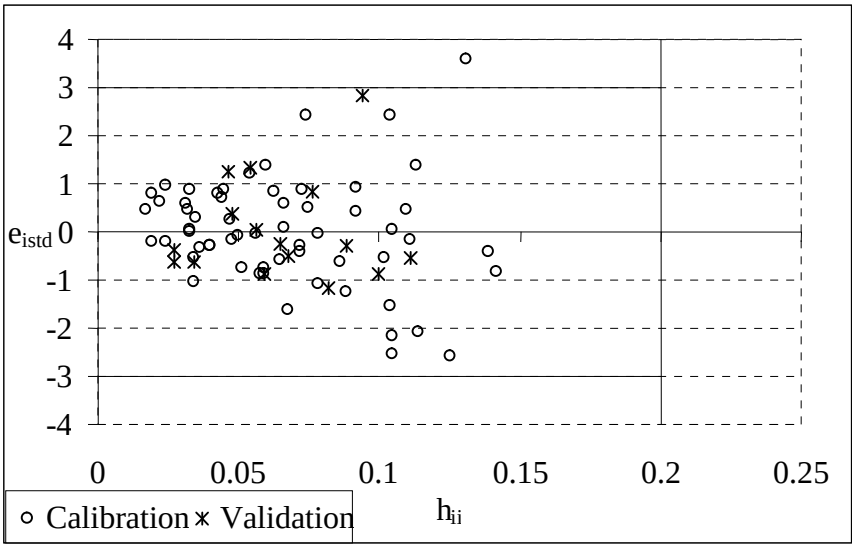


Figure 7 : *Diagramme de Williams (Choix aléatoire, ALOGP).*

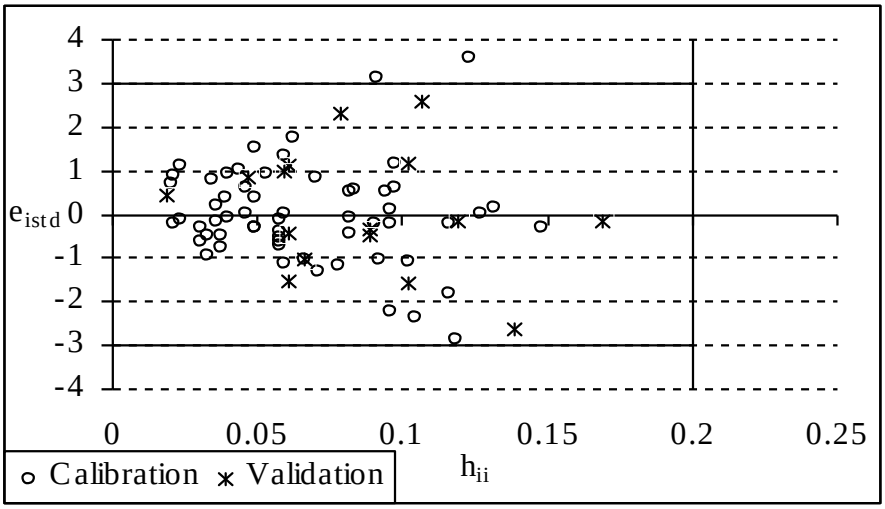


Figure 8 : *Diagramme de Williams (Choix par DUPLEX, ALOGP)*

II-5 Qualité de l'ajustement :

$$\log k_{exp} = -0.0246067 + 1.07144 \log k_{pred} \quad (21)$$

S = 0.0952093 R-carré = 93.4 % R-carré(ajust) = 93.0 %

$$\log k_{exp} = 0.0000021 + 1.00000 \log k_{calc} \quad (22)$$

S = 0.0918179 R-carré = 92.1 % R-carré(ajust) = 92.0 %

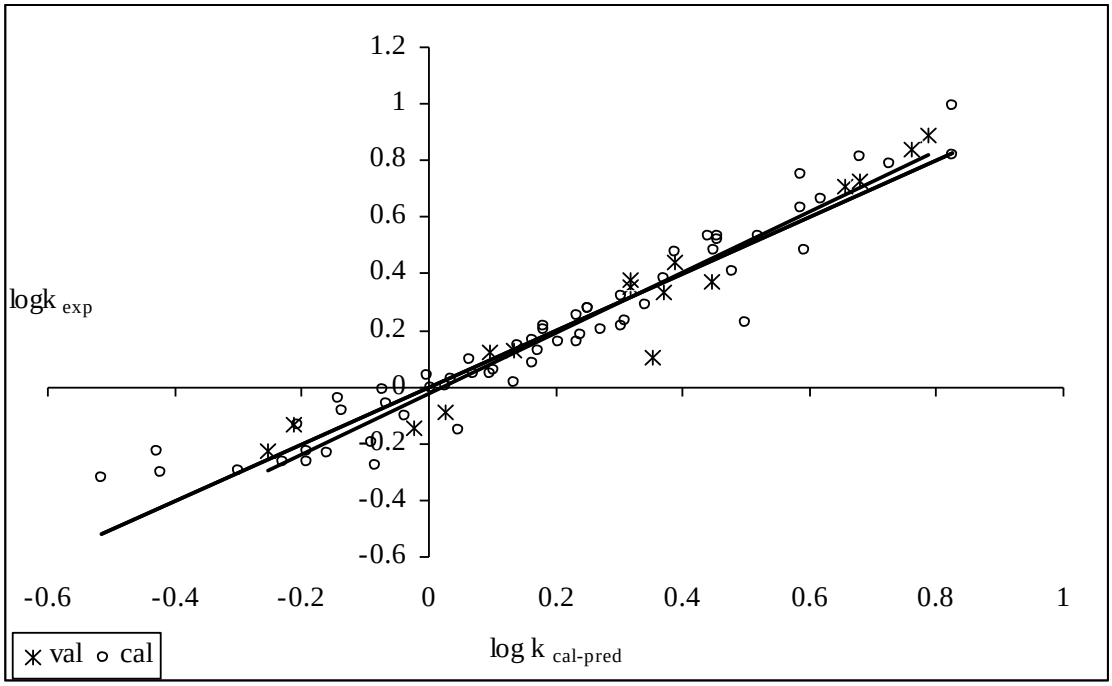


Figure 9 : *Droites d'ajustements des deux ensembles ; cas du choix aléatoire*

$$\log k_{exp} = 0.0012866 + 0.961547 \log k_{pred} \quad (23)$$

S = 0.118930 R-carré = 90.8 % R-carré(ajust) = 90.1 %

$$\log k_{exp} = 0.0000104 + 0.999966 \log k_{calc} \quad (24)$$

S = 0.0849857 R-carré = 93.2 % R-carré(ajust) = 93.1 %

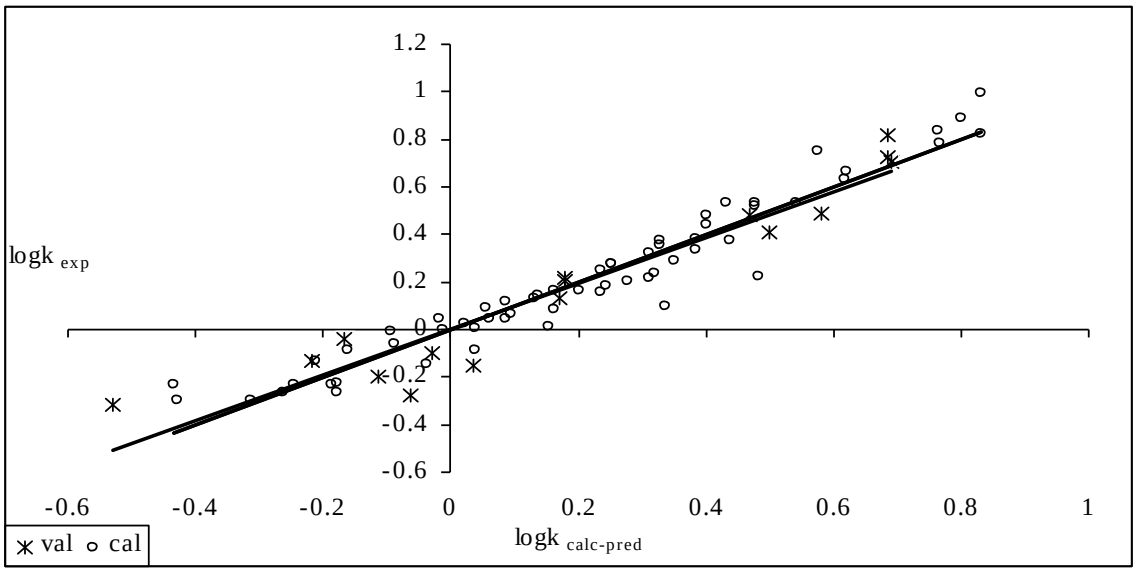


Figure 10: *Droites d'ajustements des deux ensembles ; cas du choix par DUPLEX.*

Les figure 9 et 10 (Choix aléatoire et par DUPLEX, respectivement) nous permettent d'apprécier la qualité de l'ajustement pour les ensembles de calibration et de validation.

II-6 Tests de randomisations :

Les figures suivant 11 et 12 permettent de comparer les résultats obtenus pour les modèles randomisés (cercle) au modèle réel de départ (triangle noir).

Il est clair que les statistiques obtenues pour les vecteurs modifiés du logarithme du coefficient de partage sont plus petites que celles du modèle QSRR réel. On obtient aussi des $Q^2 < 0$.

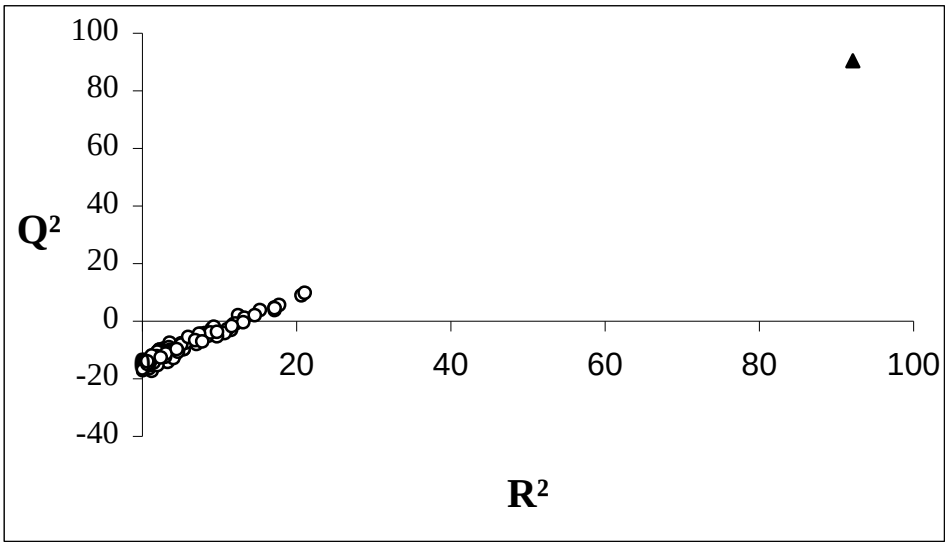


Figure 11 : *Test de randomisation relatif à ALOGP (Choix aléatoire)*

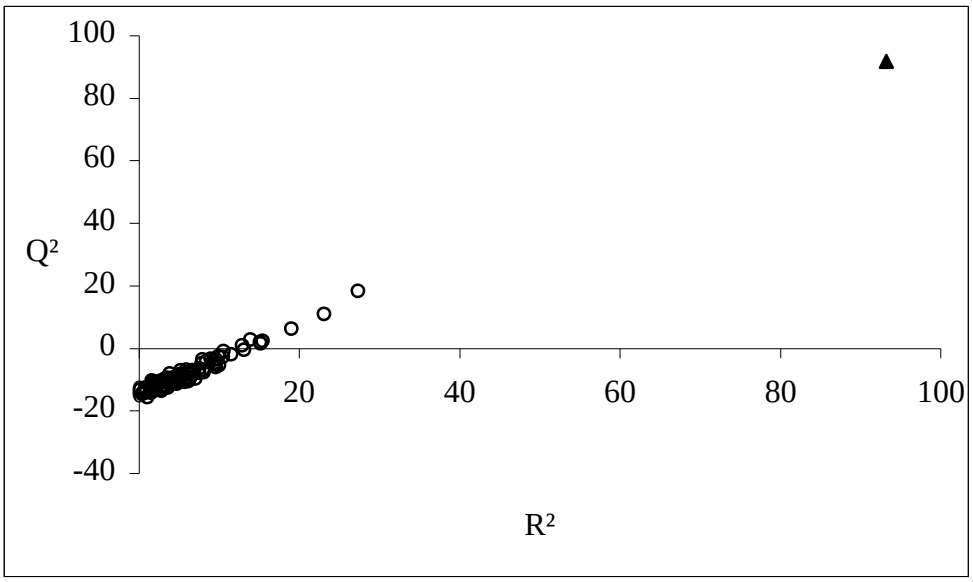


Figure 12 : *Test de randomisation relatif à ALOGP (Choix par DUPLEX)*

III- Modélisation du facteur de capacité par T, X et MLOGP :

III-1 Représentation des répartitions :

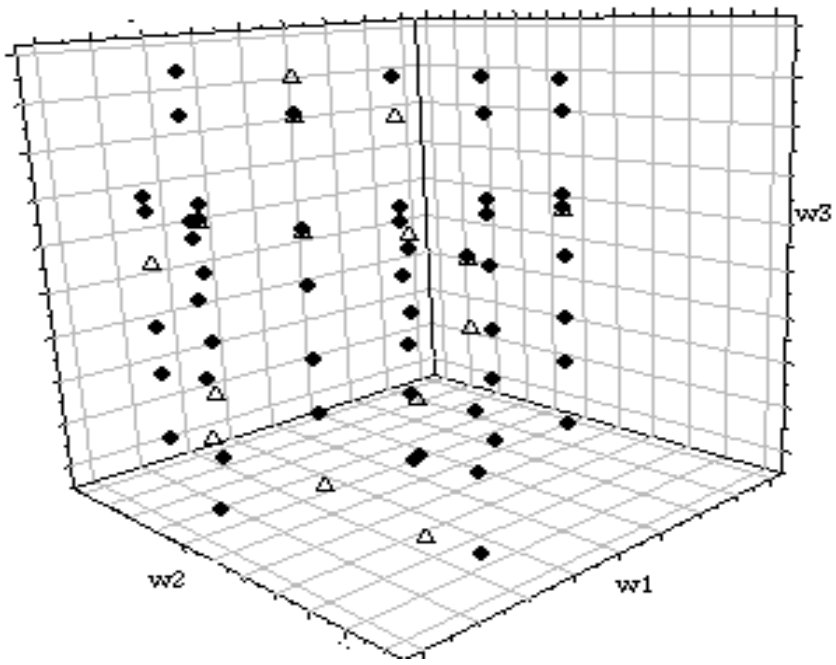


Figure 13 *Représentation du choix aléatoire*

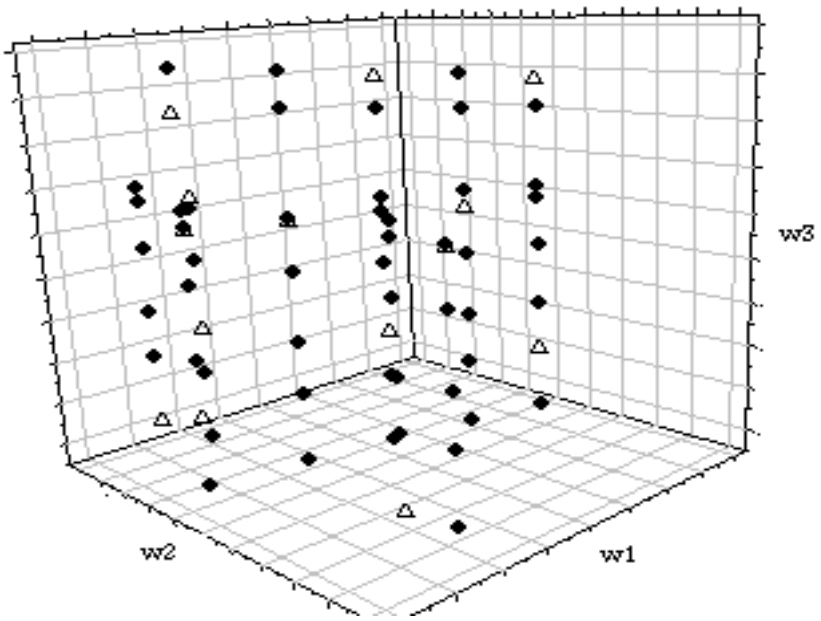


Figure 14 : *Représentation du choix par DUPLEX*

Tableau 8 : Corrélations entre les variables pour les deux choix (cas de MLOGP), obtenus avec le logiciel MINITAB [23]

	Choix aléatoire	**Choix par DUPLEX**
X	-0,742 (0,000)	-0,746 (0,000)
T	-0,196 (0,133)	-0,205 (0,116)
MLOGP	0,526 (0,000)	0,472 (0,000)

III-2 Analyse de régression

L'équation (25) de régression est relative au choix aléatoire et l'équation (26) est celle du choix par DUPLEX

$$\log k = 2{,}37(\pm 0{,}2949) - 0{,}00676(\pm 0{,}0009650)\,T - 1{,}43(\pm 0{,}07544)X + 0{,}313(\pm 0{,}02244)\ \text{MLOGP} \qquad (25)$$

$$\log k = 2{,}51(\pm 0{,}2924) - 0{,}00727(\pm 0{,}0009695)\,T - 1{,}48(\pm 0{,}07092)\,X + 0{,}325(\pm 0{,}02189)\ \text{MLOGP} \qquad (26)$$

Les paramètres statistiques de l'ensemble de calibration et de l'ensemble de validation, et dans les deux cas (choix aléatoire et choix par DUPLEX) sont réunis dans le tableau 9 :

Tableau 9 : Paramètres statistiques des deux régressions et validations (Cas de MLOGP)

		Choix aléatoire	Choix par DUPLEX
n_{tr}		60	
n_{ext}		16	
F		191,10	212,73
R^2		91,10	91,93
R^2_{adj}		90,62	91,5
Q^2_{LOO}		89,11	90,34
Q^2_{LMO}	10%	88,99	90,25
	20%	88,85	90,14
	30%	89,80	90,96
	40%	90,71	91,82
	50%	88,08	89,48
Q^2_{BOOT}		88,08	89,47
Q^2_{ext}		89,72	86,69
S		0,09	0,09
EQMC		0,09	0,09
EQMP		0,10	0,10
$EQMP_{ext}$		0,10	0,11

III-3 Analyse des résidus :

Tableau 10 : Compositions des ensembles, valeurs des log k (calculées, prédites et expérimentales), résidus de prédiction standardisés et valeurs des leviers (h_{ii}) des observations. (Cas de MLOGP)

			CHOIX ALEATOIRE				CHOIX PAR DUPLEX			
i	Code	$logk_{exp}$	Statut	$logk_{calc-pred}$	h_{ii}	e_{istd}	Statut	$logk_{calc-pred}$	e_{istd}	h_{ii}
1	Ap50	0,1443	CAL	0,1931	0,049	0,5313	CAL	0,1889	0,5063	0,048
2	Aq15	0,4853	CAL	0,6405	0,107	1,8558	CAL	0,6478	1,9958	0,096
3	Aq25	0,3732	VAL	0,4972	0,07	1,2972	CAL	0,5001	1,4758	0,063
4	Aq50	0,0449	CAL	0,1255	0,028	0,8478	CAL	0,1162	0,7833	0,027
5	Aq70	-0,1313	CAL	-0,1476	0,05	-0,1778	VAL	-0,1647	-0,3619	0,052
6	Aq85	-0,2994	CAL	-0,3625	0,096	-0,7412	CAL	-0,3863	-1,0719	0,099
7	Ar50	-0,001	CAL	0,0578	0,025	0,6171	VAL	0,0436	0,4768	0,026
8	As50	-0,0578	CAL	-0,0098	0,041	0,5161	CAL	-0,0291	0,3254	0,047
9	At50	-0,0829	CAL	-0,0774	0,077	0,0627	CAL	-0,1018	-0,2285	0,089
10	Bp50	-0,1542	CAL	0,0011	0,087	1,7976	CAL	-0,0103	1,7358	0,086
11	Bq15	0,2273	CAL	0,4485	0,146	2,8298	CAL	0,4486	2,8844	0,132
12	Bq25	0,1012	VAL	0,3052	0,109	2,1806	VAL	0,3009	2,2181	0,099
13	Bq50	-0,1442	VAL	-0,0665	0,067	0,8122	CAL	-0,083	0,7182	0,068
14	Bq70	-0,2952	CAL	-0,3396	0,091	-0,5162	CAL	-0,364	-0,8439	0,097
15	Bq85	-0,3188	CAL	-0,5545	0,137	-2,9661	VAL	-0,5856	-3,0398	0,144
16	Br50	-0,1957	CAL	-0,1341	0,068	0,6903	CAL	-0,1557	0,4718	0,072
17	Bs50	-0,2319	CAL	-0,2018	0,087	0,3486	CAL	-0,2283	0,0436	0,096
18	Bt50	-0,2648	CAL	-0,2694	0,125	-0,0564	CAL	-0,301	-0,4801	0,142
19	Cp50	0,1291	VAL	0,1117	0,06	-0,182	VAL	0,1044	-0,269	0,058
20	Cq15	0,7485	CAL	0,5591	0,119	-2,3108	CAL	0,5633	-2,3115	0,106
21	Cq25	0,5324	CAL	0,4158	0,083	-1,339	CAL	0,4156	-1,3819	0,074
22	Cq50	0,0946	CAL	0,044	0,04	-0,5423	CAL	0,0317	-0,7041	0,039
23	Cq70	-0,1316	VAL	-0,229	0,062	-1,0152	CAL	-0,2492	-1,3747	0,066
24	Cq85	-0,2285	CAL	-0,444	0,109	-2,5835	CAL	-0,4708	-3,0643	0,114
25	Cr50	0,0432	CAL	-0,0236	0,039	-0,7154	CAL	-0,0409	-0,9451	0,041
26	Cs50	-0,0067	CAL	-0,0912	0,056	-0,9298	CAL	-0,1136	-1,2434	0,063
27	Ct50	-0,0392	CAL	-0,1588	0,092	-1,3962	VAL	-0,1863	-1,6393	0,105
28	Dp50	0,3211	CAL	0,3199	0,026	-0,0127	VAL	0,3181	-0,0329	0,025
29	Dp50	0,3853	CAL	0,3876	0,05	0,0244	CAL	0,3907	0,0615	0,05
30	Dq15	0,9965	CAL	0,835	0,106	-1,9296	VAL	0,8497	-1,6325	0,1
31	Dq25	0,8164	CAL	0,6917	0,07	-1,4021	CAL	0,7019	-1,3379	0,067
32	Dq70	0,0053	CAL	0,0469	0,048	0,4524	CAL	0,0371	0,3615	0,048
33	Dq85	-0,265	CAL	-0,1681	0,094	1,1339	CAL	-0,1845	0,9818	0,092
34	Dr50	0,2527	CAL	0,2523	0,02	-0,0043	CAL	0,2454	-0,0798	0,021
35	Ds50	0,1658	CAL	0,1847	0,034	0,201	VAL	0,1727	0,0748	0,038
36	Dt50	0,1203	VAL	0,1171	0,066	-0,0338	CAL	0,1001	-0,2402	0,076
37	Ep50	0,4076	CAL	0,4082	0,052	0,0073	CAL	0,4122	0,0527	0,052
38	Eq25	0,891	VAL	0,7124	0,071	-1,8701	CAL	0,7234	-1,9684	0,069
39	Eq70	0,0154	CAL	0,0675	0,051	0,5692	VAL	0,0586	0,4673	0,049
40	Eq85	-0,2765	CAL	-0,1474	0,096	1,5152	CAL	-0,163	1,3874	0,094
41	Er50	0,2876	CAL	0,273	0,022	-0,1523	CAL	0,2668	-0,2263	0,023
42	Es50	0,2028	CAL	0,2054	0,036	0,0277	CAL	0,1942	-0,096	0,039

Tableau 10 : suite et fin

			CHOIX ALEATOIRE				CHOIX PAR DUPLEX			
i	Code	$logk_{exp}$	Statut	$logk_{calc-pred}$	h_{ii}	e_{istd}	Statut	$logk_{calc-pred}$	e_{istd}	h_{ii}
43	Et50	0,1617	CAL	0,1377	0,068	-0,2682	CAL	0,1215	-0,4774	0,077
44	Fp50	0,3359	VAL	0,3876	0,049	0,5344	CAL	0,3907	0,6239	0,05
45	Fq15	0,8217	CAL	0,835	0,106	0,1582	VAL	0,8497	0,3107	0,1
46	Fq25	0,7245	VAL	0,6917	0,069	-0,343	CAL	0,7019	-0,2637	0,067
47	Fq50	0,2155	CAL	0,3199	0,026	1,0956	CAL	0,3181	1,1233	0,025
48	Fq70	-0,0867	VAL	0,0469	0,048	1,381	CAL	0,0371	1,4051	0,048
49	Fq85	-0,2258	CAL	-0,1681	0,094	0,6748	CAL	-0,1845	0,5032	0,092
50	Fr50	0,1609	CAL	0,2523	0,02	0,9514	CAL	0,2454	0,9201	0,021
51	Fs50	0,0865	CAL	0,1847	0,034	1,0436	VAL	0,1727	0,9271	0,038
52	Ft50	0,0477	CAL	0,1171	0,067	0,7762	CAL	0,1001	0,6218	0,076
53	Gp50	0,2341	CAL	0,3036	0,045	0,7518	CAL	0,3036	0,7845	0,044
54	Gq15	0,8393	VAL	0,751	0,1	-0,9384	CAL	0,7626	-0,9381	0,094
55	Gq25	0,6669	CAL	0,6078	0,065	-0,6607	CAL	0,6148	-0,6028	0,06
56	Gq50	0,1831	CAL	0,236	0,022	0,5517	CAL	0,231	0,5207	0,021
57	Gq70	-0,0994	CAL	-0,0371	0,044	0,6731	CAL	-0,05	0,5576	0,045
58	Gq85	-0,2294	VAL	-0,252	0,088	-0,2382	CAL	-0,2716	-0,5121	0,09
59	Gr50	0,1308	CAL	0,1684	0,017	0,3896	CAL	0,1583	0,2983	0,018
60	Gs50	0,063	CAL	0,1008	0,032	0,3999	CAL	0,0856	0,2522	0,037
61	Gt50	0,0265	CAL	0,0331	0,066	0,0744	CAL	0,013	-0,1606	0,076
62	Hp50	0,5311	CAL	0,5473	0,08	0,1852	VAL	0,5565	0,2795	0,081
63	Hp50	0,4798	CAL	0,4796	0,054	-0,002	CAL	0,4838	0,0457	0,054
64	Hr50	0,378	VAL	0,412	0,046	0,3517	CAL	0,4112	0,3763	0,047
65	Hs50	0,2765	CAL	0,3444	0,058	0,7489	CAL	0,3385	0,7188	0,061
66	Ht50	0,206	CAL	0,2768	0,088	0,82	CAL	0,2658	0,7341	0,096
67	Ip50	0,5194	CAL	0,5473	0,08	0,3192	VAL	0,5565	0,4084	0,081
68	Iq50	0,4432	VAL	0,4796	0,053	0,3774	CAL	0,4838	0,4652	0,054
69	Ir50	0,3532	VAL	0,412	0,046	0,6076	CAL	0,4112	0,6573	0,047
70	It50	0,2767	CAL	0,3444	0,058	0,7464	CAL	0,3385	0,7161	0,061
71	It50	0,2172	CAL	0,2768	0,088	0,6902	CAL	0,2658	0,5966	0,096
72	Jp50	0,786	CAL	0,6024	0,097	-2,1579	CAL	0,6137	-2,1283	0,1
73	Jq50	0,7049	VAL	0,5348	0,069	-1,7788	CAL	0,541	-1,93	0,071
74	Jr50	0,6309	CAL	0,4671	0,061	-1,8159	VAL	0,4684	-1,7696	0,062
75	Js50	0,531	CAL	0,3995	0,072	-1,4838	CAL	0,3957	-1,6059	0,076
76	Jt50	0,4811	CAL	0,3319	0,102	-1,7684	VAL	0,323	-1,7655	0,108

III-4 Diagrammes de Williams :

Le diagramme de Williams pour les régressions multilinéaires (RLM) est reproduit dans les deux figures 15 et 16. Dans les deux répartitions on n'a pas de point influent car les leviers sont tous inférieurs à h^{*}=0.2.

Les observations Bq85 (Validation) et Cq85 (calibration) pour le choix fait avec DUPLEX sont aberrantes. Pour le choix aléatoire tous les résidus de prédiction standardisés sont compris entre les bornes ±3.

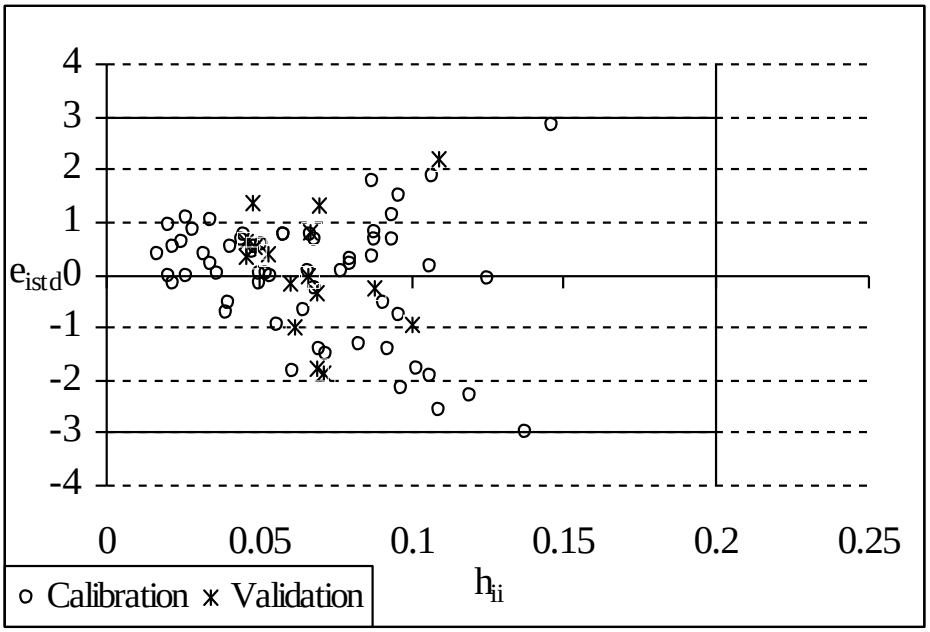


Figure 15 : *Diagramme de Williams (Choix aléatoire, MLOGP).*

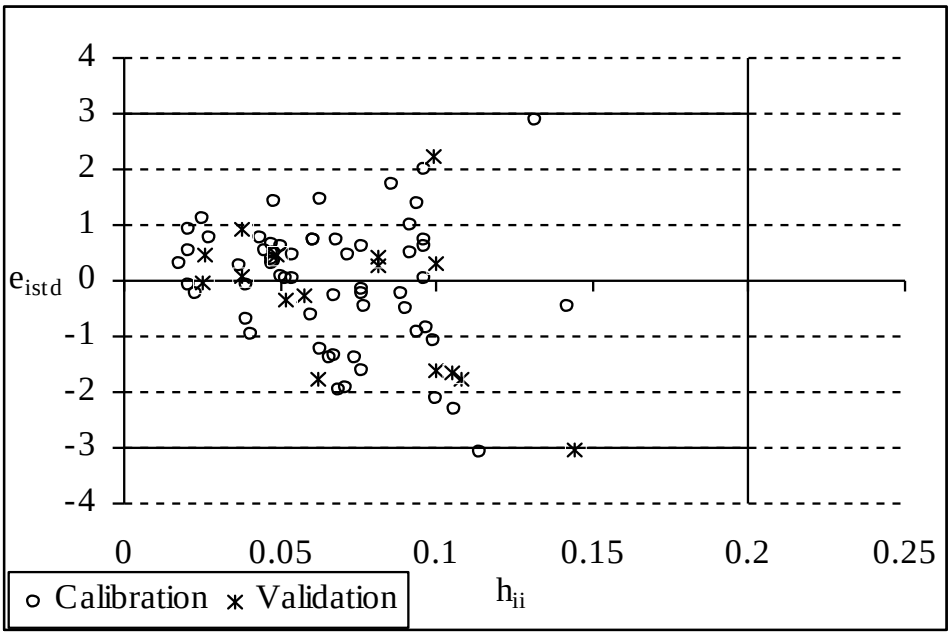


Figure 16 : *Diagramme de Williams (Choix par DUPLEX, MLOGP)*

III-5 Qualité de l'ajustement :

La qualité de l'ajustement peut être vérifié en procédant à la représentation des valeurs expérimentales en fonction de celles calculées du log k, pour l'ensemble de calibration, et celles prédites pour l'ensemble de validation, dans les figures 17 et 18 par les deux droites d'ajustements dont les équations sont :

- Pour l'ensemble de calibration

$$\log k_{exp} = 0{,}0000137 + 0{,}999970 \log k_{calc} \qquad (27)$$

S = 0,0973766 R-carré = 91,1 % R-carré(ajust) = 90,9 %

- Pour l'ensemble de validation

$$\log k_{exp} = -0{,}0297285 + 1{,}07449 \log k_{pred} \qquad (28)$$

S = 0,106979 R-carré = 91,7 % R-carré(ajust) = 91,1 %

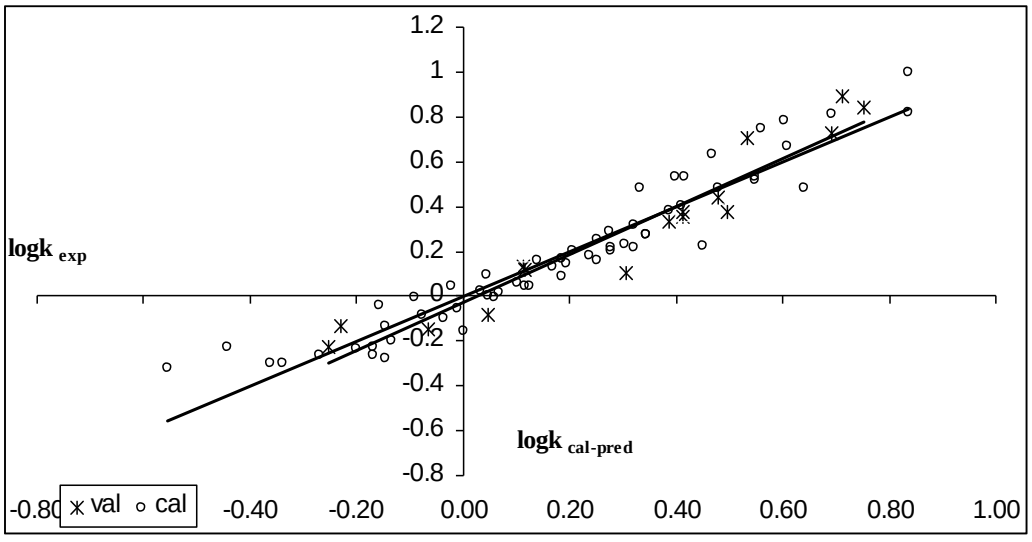


Figure 17 : *Droites d'ajustements des deux ensembles ; cas du choix aléatoire*

- Pour l'ensemble de calibration

$$\log k_{exp} = 0{,}0000019 + 1{,}00002 \log k_{calc} \qquad (29)$$

S = 0,0931705 R-carré = 91,9 % R-carré(ajust) = 91,8 %

- Pour l'ensemble de validation

$$\log k_{exp,} = 0{,}0503152 + 0{,}913047 \log k_{pred} \qquad (30)$$

S = 0.116947 R-carré = 90.3 % R-carré(ajust) = 89.6 %

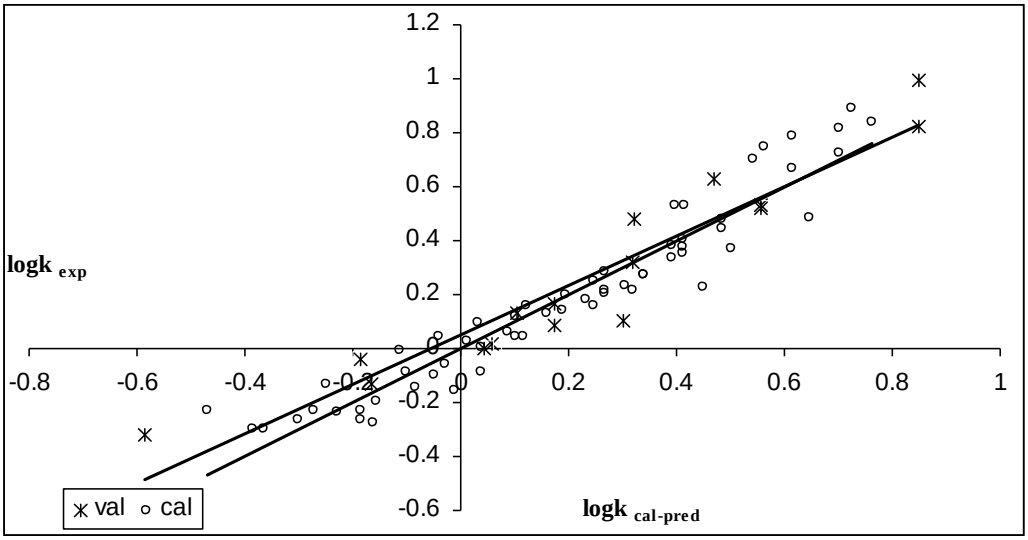


Figure 18 : *Droites d'ajustements des deux ensembles ; cas du choix par DUPLEX.*

III-6 Tests de randomisations :

Les 100 modèles pour lesquels nous avons randomisé les valeurs du logarithme du coefficient de partage ont des valeurs de Q^2 ou faibles ou négatives, et des valeurs du coefficient de corrélation multiple (R^2) petites. Seuls les modèles (triangles) avec les vecteurs réels offrent des valeurs élevées pour les deux statistiques représentées dans les figures 19 et 20.

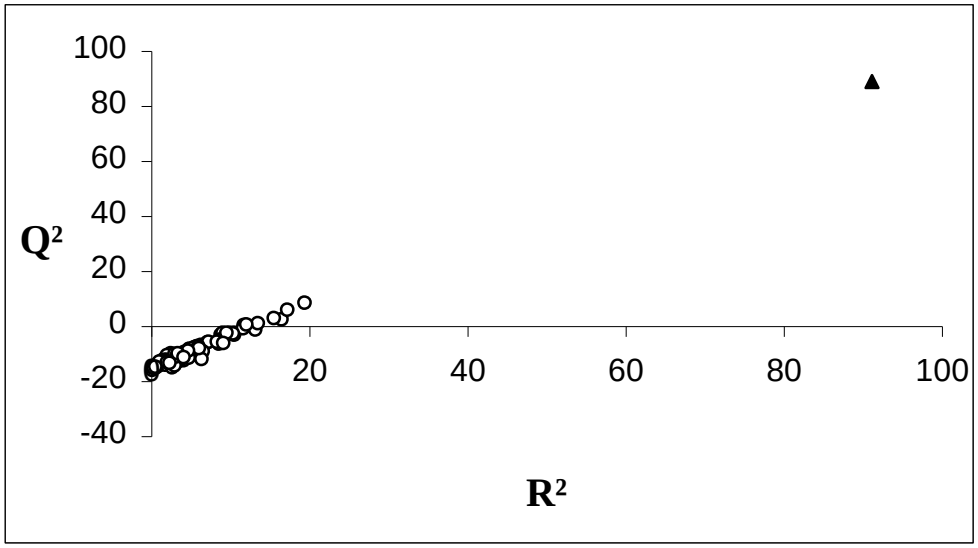


Figure19 : ***Représentation du test de randomisation (****Choix aléatoire, MLOGP)*

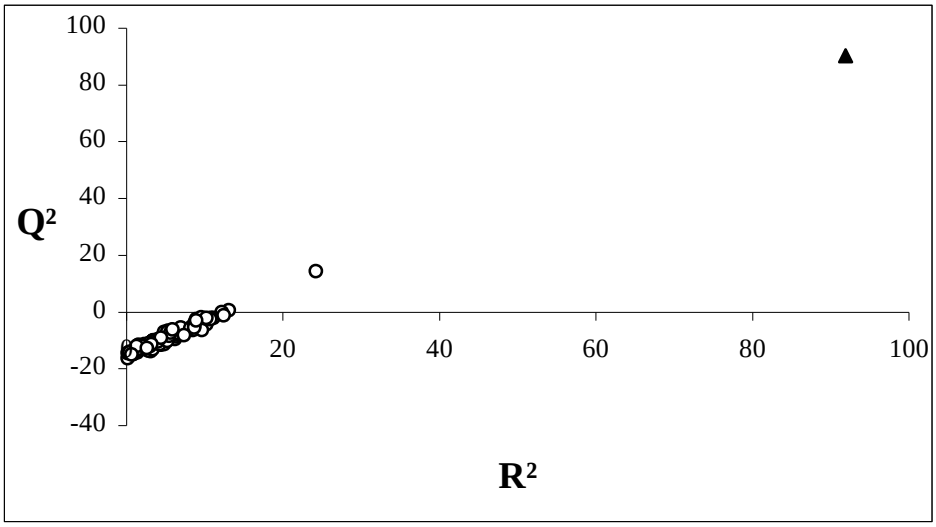


Figure 20 : ***Représentation du test de randomisation (****Choix par DUPLEX, MLOGP)*

Représentation de la répartition :cas de ALOGP

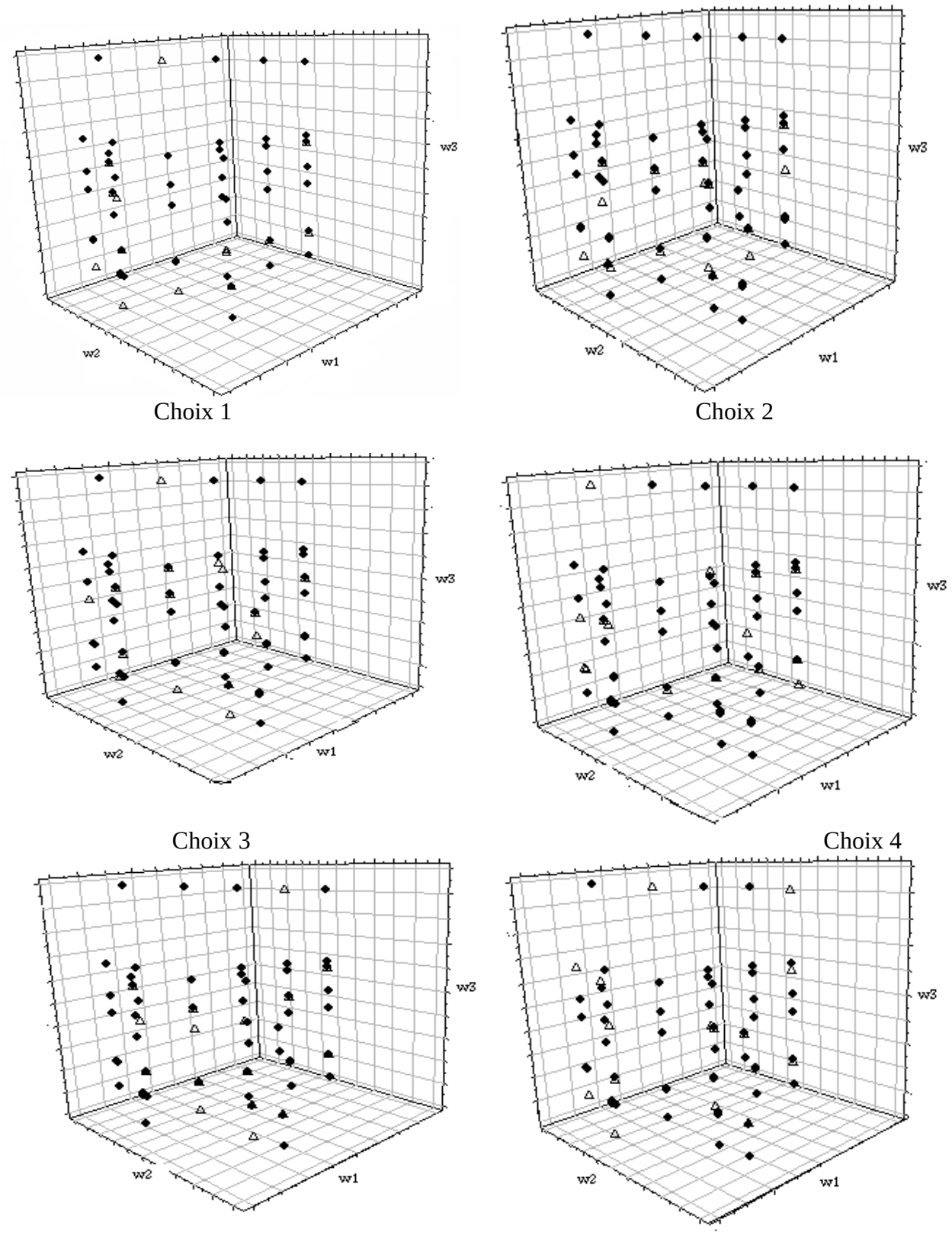
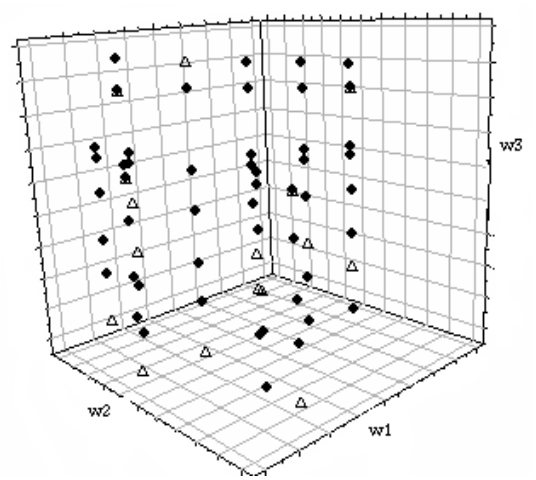


Figure 21 : *Représentations des 5 choix aléatoires et du choix par DUPLEX*

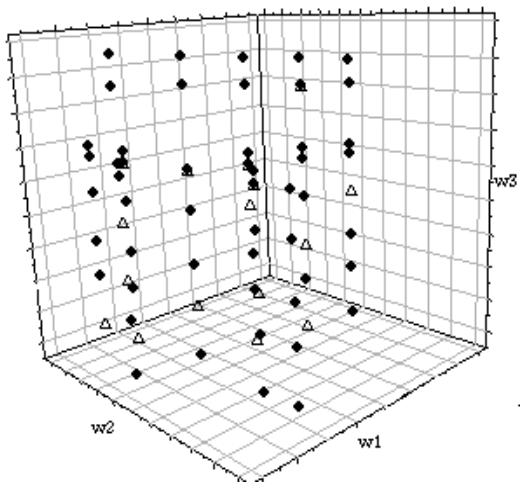
Tableau 11 : Paramètres statistiques des différents choix (Cas ALOGP)

		Choix 1	Choix 2	Choix3	Choix 4	Choix 5	Choix par DUPLEX
n_{tr}		60					
n_{ext}		16					
F		219,28	207,72	217,28	166,42	229,75	255,06
R^2		92,16	91,75	92,09	89,92	92,49	93,18
R^2_{adj}		91,74	91,31	91,67	89,37	92,08	92,82
Q^2_{LOO}		90,56	90,17	90,37	87,67	90,93	91,8
Q^2_{LMO}	10%	90,47	90,08	90,27	87,58	90,85	91,71
	20%	90,36	89,97	90,15	87,41	90,73	91,60
	30%	91,17	90,80	90,98	88,45	91,52	92,34
	40%	92,00	91,67	91,77	89,50	92,31	93,03
	50%	89,65	89,25	89,37	86,47	90,09	91,03
Q^2_{BOOT}		89,69	89,32	89,4	86,37	90,13	90,99
Q^2_{ext}		83,44	92,46	91,80	88,33	90,01	87,69
S		0,09	0,10	0,09	0,09	0,10	0,08
EQMC		0,08	0,09	0,09	0,09	0,09	0,08
EQMP		0,09	0,10	0,10	0,10	0,10	0,09
$EQMP_{ext}$		0,12	0,09	0,09	0,10	0,11	0,11

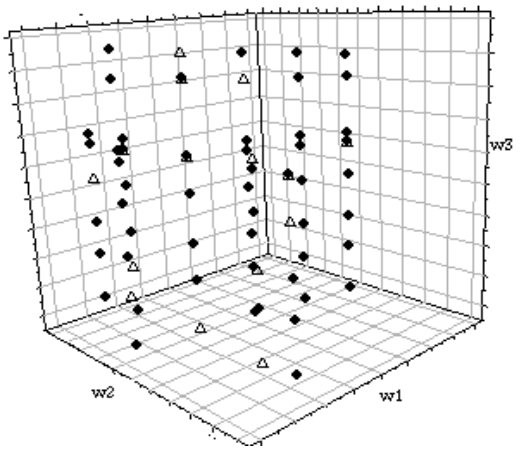
Représentation de la répartition : cas de MLOGP



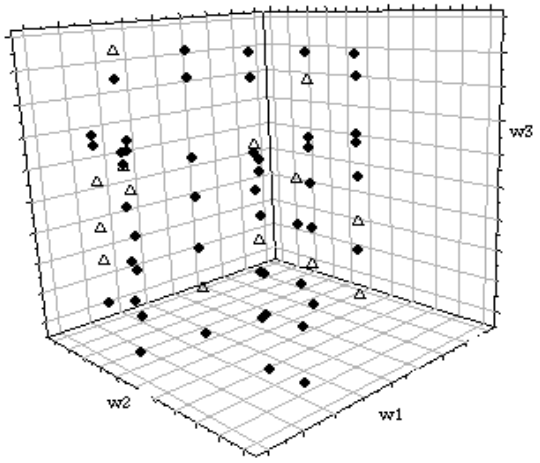
Choix 1



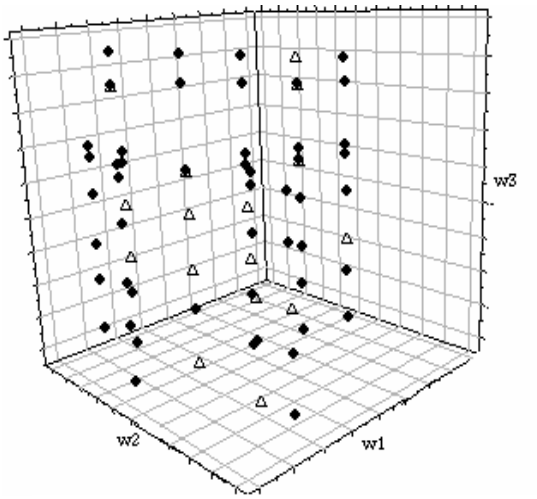
Choix 2



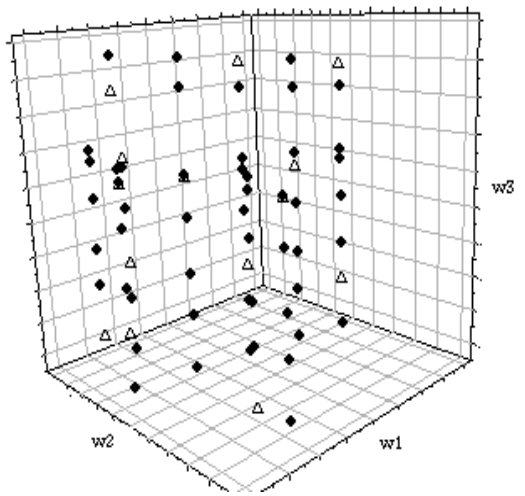
Choix 3



Choix 4



Choix 5



Choix par DUPLEX

Figure 22 : *Représentations des 5 choix aléatoires et du choix par DUPLEX*

Tableau 12 : Paramètres statistiques des diffèrent choix (cas MLOGP)

		Choix 1	Choix 2	Choix3	Choix 4	Choix 5	Choix par DUPLEX
n_{tr}		60					
n_{ext}		16					
F		219,28	207,72	191,10	166,42	229,75	212,73
R^2		92,16	91,75	91,10	89,92	92,49	91,93
R^2_{adj}		91,74	91,31	90,62	89,37	92,08	91,5
Q^2_{LOO}		90,56	90,17	89,11	87,67	90,93	90,34
Q^2_{LMO}	10%	90,47	90,08	88,99	87,58	90,85	90,25
	20%	90,36	89,97	88,85	87,41	90,73	90,14
	30%	91,17	90,80	89,80	88,45	91,52	90,96
	40%	92,00	91,67	90,71	89,50	92,31	91,82
	50%	89,65	89,25	88,08	86,47	90,09	89,48
Q^2_{BOOT}		89,69	89,32	88,08	86,37	90,13	89,47
Q^2_{ext}		83,44	92,46	89,72	88,33	90,01	86,69
S		0,0915	0,1024	0,09	0,0999	0,1	0,09
EQMC		0,088	0,099	0,09	0,097	0,097	0,09
EQMP		0,097	0,108	0,10	0,107	0,106	0,10
$EQMP_{ext}$		0,128	0,095	0,10	0,104	0,111	0,11

Dans les figures 21et 22 (les deux pages suivantes), représentant les différents choix (5 obtenues) aléatoirement faits et un avec l'algorithme DUPLEX) on peut faire les remarques suivantes qui s'appliquent pour les deux coefficient de partage utilisés :

- Le choix par DUPLEX donne une représentation homogène des observations qui se vérifie dans la distribution des points de la périphérie. On remarque que ces points sont également partagés entre les ensembles de calibration et de validation. De cela on peut supposer ou prévoir (ce qui sera vérifié par la suite par les paramètres statistiques des différents modèles), que le modèle obtenu par DUPLEX devra faire un effort, si l'on puit dire, pour prédire ces points périphériques écartés pour la validation.
- Dans certains choix, quelques points de validation ont une tendance particulière de regroupement dans une région donnée. Le premier choix en particulier, est très parlant car on y perçoit les points de validation, pour la majorité, situé dans la périphérie.

Les tableaux 11 et 12 condensent les paramètres statistiques des différents choix et pour les deux mesures d'hydrophobicité, on y remarque la prédominance des modèles obtenus par la séparation avec la l'algorithme DUPLEX. Excepté pour la validation statistique externe, les modèles sont, à tout point de vue, meilleurs que ceux obtenus en choisissant aléatoirement les deux ensembles, d'estimation et de validation.

Dans le ces ALOGP-DUPLEX, la primauté du modèle est nettement perceptible et si l'on compare les paramètres statistiques du modèles MLOGP-DUPLEX aux valeurs moyennes obtenues pour les 5 choix (colonne 6 du tableau 12), la supériorité du modèle LOGP-DUPLEX est clairement établie.

La déficience qu'on remarque pour les paramètres statistiques de la validation (Q^2_{ext} ,$SDEP_{ext}$) des modèles obtenus par DUPLEX par rapport à ceux obtenus après division aléatoire des données, est une conséquence du fait que l'on teste, aussi sévèrement que le permettent les données, les performances des modèles et leur capacité à l'extrapolation (prédiction des points périphériques mentionnés plus haut).

Tableau : Récapitulatif des paramètres statistiques des différents modèles développés.

	R^2	R^2_{adj}	Q^2_{LOO}	Q^2_{LMO}					Q^2_{BOOT}	Q^2_{ext}
				10%	20%	30%	40%	50%		
ALOGP-DUPLEX	93.18	92.82	91.80	91.71	91.60	92.34	93.03	91.03	90.99	87.69
ALOGP-Aléatoire	92.09	91.67	90.37	90.27	90.15	90.98	91.77	89.37	89.40	91.80
MLOGP-DUPLEX	91.93	91.50	90.34	90.25	90.14	90.96	91.82	89.48	89.47	86.69
MLOGP-Aléatoire	91.10	90.62	89.11	88.99	88.85	89.80	90.71	88.08	88.08	89.72

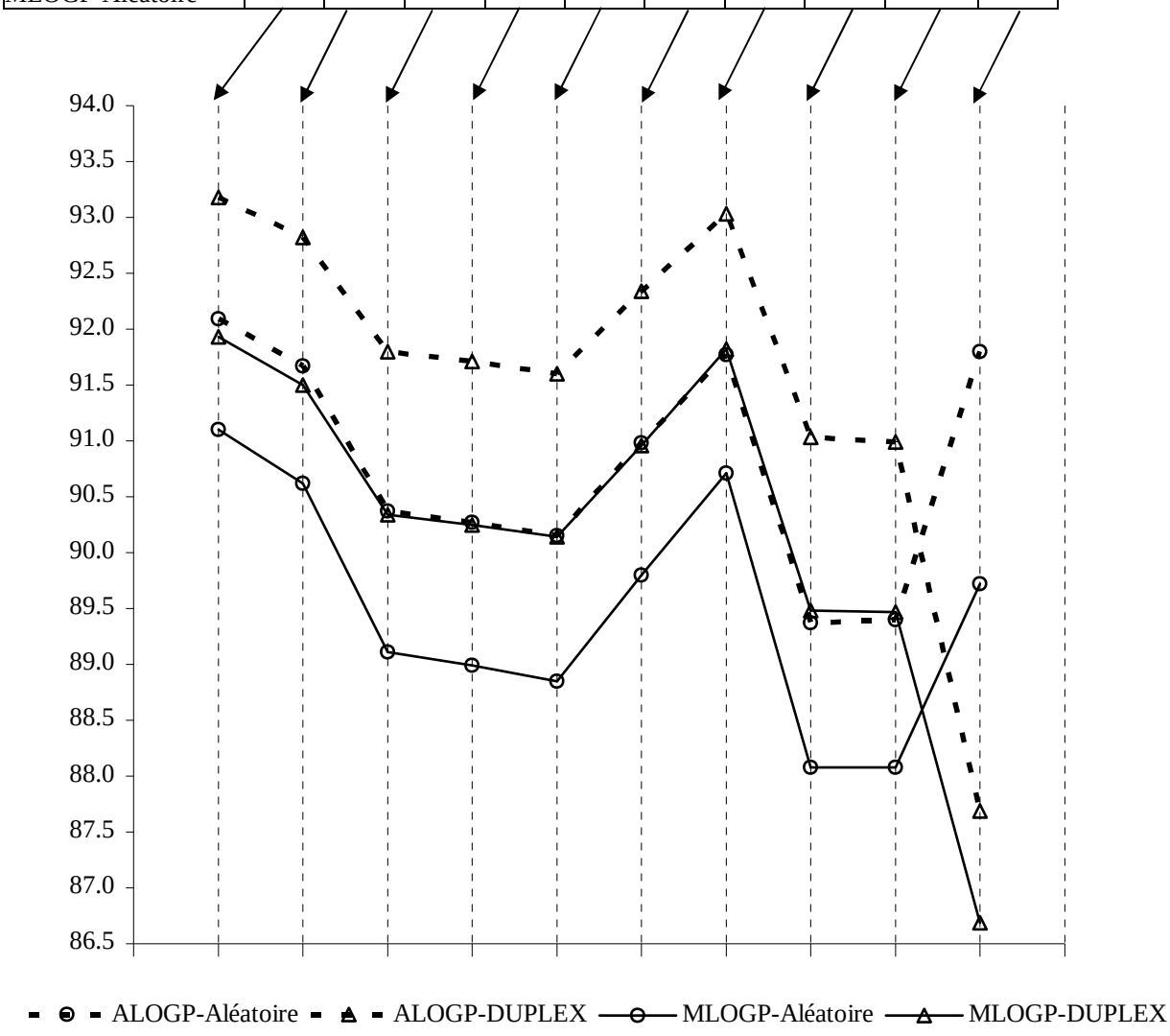


Figure : *Illustration des paramètres statistiques du tableau*

CONCLUSION

Dans notre travaill nous avons utilisé la méthodologie QSRR pour relier le facteur de capacité des 10 phénols, à des descripteurs moléculaires théoriques caractéristiques de la molécule.

Les conditions expérimentales obtenus dans différent conditions de séparation s'imposent d'elles mêmes dans le choix de T (température de la colonne) ; x (la fraction volumique) du méthanol des phases mobiles hydro -organiques, en plus du descripteur log P (coefficient de partage phase organique /aqueuse) est introduit dons le modèle QSRR, parce que l'hydrophobicité encode la plupart des forces d'intermoléculaire ente le soluté et le solvant.

Nous avons appliqué la RLM pour recherché des corrélations linéaires entre la variable dépendante et les variables explicatives imposées.

Les 76 observation ont été séparées une fois aléatoirement, et autre fois par l'algorithme DUPLEX, en ensemble de calibration de 60 observations, et ensemble de validation externe de 16 éléments.

Les deux modèle obtenus sont bons à tous les points de vue : qualité de l'ajustement, robustesse interne et externe, capacité prédictive, et la stabilité ce qui prouve que les corrélations variable dépendante –variables indépendantes (température de la colonne, fraction volumique, et LOGP) sont linéaires.

[1]- G,G. kirkland, L. R Snyder. A.Wiley. Introduction to Modern Liquid Chromatography, New York, Chichester, Brisbane, Tornto, Singapore, (1978).

[2]- G. Mahuzier, M. Hamon, Abrégé de chimie analytique (Tome II), Masson Pairis .134,(1978).

[3]- G. Mahuzier, M. Hamon, D. Ferrier, P. Prognon, chimie analytique (Tome II),Masson Paris, 197-201, (1999).

[4]- M.Karelson. Molecular descriptors in QSAR/QSPR. Wiley- Interscience, . 385, (2000).

[5]- T. Aoyama, Y. Suzuki, H. Ichikawa. J. Med. Chem., 33, 2583 (1990).

[6]-A.K.Ghose and G.M.Crippen ,J.Comput.chim.7,565-577,(1986).

[7]- V.N.Vis wanadhan et al.,J. Comput.chim.,14,1019-1026,(1993).

[8]-I.Morighuchi,S.Hirono,Q.Lui,I.Nakagome,and Y.Mastushita, Chem.Pharm.bull., 40,127-130,(1992).

[9]-I.Morighuchi,S.Hirono,I.Nakagome,H.Hirono,Chem.Pharm.Bull, 42,976-978,(1994).

[10]- B. Kowalski, R. Gerlach, H. Wold. Systems under Indirect Observation (K. Jörenskog et H. Wold, eds.), North Holland, Amsterdam, 191-206, (1982).

[11]- L. Eriksson, E. Vohannson, N. Kettaneh- Wold. Multi and Megavariate Data Analysis- Principles and Applications. Umetricsacademy, Umeå (2001).

[12]- S. Wold, A. Ruhe, H.Wold, W. Dunn. SIAMJ. Sci. Stat. Comput., 5, 735 (1984).

[13]- S. Wold. Chemometrics: Mathematics and Statistics in Chemistry. Reidel, Dordrecht, The Netherlands (1984).

[14]- P. Gelada, B. R. Kowalski, Anal. Chim. Acta, 185, 1 (1986).

[15]- A. Höskuldsson, J. Chemometrics, 2, 211 (1988).

[16]- Hyperchem ^{TM} Release 6.03 for windows, Molecular Modeling System (2000).

[17]- R. Todeschini, V. Consonni, M. Pavan. DRAGON, Software for the Calculation of Molecular Descriptors. Release 5.3 for windows, Milano (2005).

[18]- R. Todeschini, D. Ballabio, V. Consonni, A. Mauri, M. Pavan. MOBYDIGS Software for Multilinear Regression Analysis and variable Subset Selection by Genetic Algorithm. Release I.1 for Windows, Milano (2009)

[19]-D. M. Allen. The relationship between variable selection and data augmentation and a method for prection. Technometries

[20]- B. Efron, R.J. Tibshirami. An introduction to the Bootstrap. Chapman & Hall (1993) – R. Wehrens, H. Putter, L.M.C. Buydens. The Bootstrap. A Tutorial. Chem. Int. Lab. Syst. 54 : 35-52 (2000).

[21]-L. Erikson, J. Jaworska, A. Worth, M. Cromin, R.M. Mc Dowell, P. Gramatica. Methods for reliability, uncertainty assessment, and applicability evaluations of regression based and classification QSPRs. Environmental Health Perspective. 111 (10) : 1361-1375 (2003).

[22]- R. Todeschini, V. Consonni. Handbook of Molecular Descriptors. R. Mannhold, H. Kubinyi, H. Timmerman eds. Wiley – VCH Verlag GmbH, Weinheim (2000)

[23]- Ronald D .Snee *Technometrics*, 19, 415-428,(1977).

[24]- Matlab Version 7.0.0.19920 (Release 14) The Language of Technical Computing The MathWorks, Inc. May 06, 2004.

ANNEXE : Variable orthonormalisés des différents cœfficients de partages utilisés

			ALOGP			MLOGP		
i	**Code**	**log k**	**w1**	**w2**	**w3**	**w1**	**w2**	**w3**
1	Ap50	0,144263	-0,1565	-0,001	-0,0833	-0,1565	-0,001	-0,0442
2	Aq15	0,485267	-0,0768	-0,2506	-0,0833	-0,0768	-0,2506	-0,0462
3	Aq25	0,373206	-0,0768	-0,1792	-0,086	-0,0768	-0,1792	-0,0483
4	Aq50	0,044892	-0,0569	-0,0007	-0,095	-0,0569	-0,0007	-0,0559
5	Aq70	-0,131297	-0,0768	0,142	-0,098	-0,0768	0,142	-0,0577
6	Aq85	-0,299383	-0,0768	0,2491	-0,1019	-0,0768	0,2491	-0,0608
7	Ar50	-0,001044	0,0428	-0,0005	-0,1066	0,0428	-0,0005	-0,0675
8	As50	-0,057793	0,1424	-0,0002	-0,1183	0,1424	-0,0002	-0,0791
9	At50	-0,082915	0,242	0,0001	-0,13	0,242	0,0001	-0,0908
10	Bp50	-0,15422	-0,1565	-0,001	-0,1446	-0,1565	-0,001	-0,1769
11	Bq15	0,22727	-0,0768	-0,2506	-0,1447	-0,0768	-0,2506	-0,1789
12	Bq25	0,101197	-0,0768	-0,1792	-0,1473	-0,0768	-0,1792	-0,181
13	Bq50	-0,144239	-0,0569	-0,0007	-0,1563	-0,0569	-0,0007	-0,1886
14	Bq70	-0,295249	-0,0768	0,142	-0,1593	-0,0768	0,142	-0,1904
15	Bq85	-0,318849	-0,0768	0,2491	-0,1633	-0,0768	0,2491	-0,1935
16	Br50	-0,195656	0,0428	-0,0005	-0,168	0,0428	-0,0005	-0,2002
17	Bs50	-0,23188	0,1424	-0,0002	-0,1797	0,1424	-0,0002	-0,2118
18	Bt50	-0,2648	0,242	0,0001	-0,1913	0,242	0,0001	-0,2235
19	Cp50	0,129142	-0,1565	-0,001	-0,0872	-0,1565	-0,001	-0,1005
20	Cq15	0,748498	-0,0768	-0,2506	-0,0872	-0,0768	-0,2506	-0,1025
21	Cq25	0,532385	-0,0768	-0,1792	-0,0899	-0,0768	-0,1792	-0,1046
22	Cq50	0,094576	-0,0569	-0,0007	-0,0989	-0,0569	-0,0007	-0,1121
23	Cq70	-0,131591	-0,0768	0,142	-0,1018	-0,0768	0,142	-0,114
24	Cq85	-0,228486	-0,0768	0,2491	-0,1058	-0,0768	0,2491	-0,1171
25	Cr50	0,043244	0,0428	-0,0005	-0,1105	0,0428	-0,0005	-0,1238
26	Cs50	-0,006696	0,1424	-0,0002	-0,1222	0,1424	-0,0002	-0,1354
27	Ct50	-0,039196	0,242	0,0001	-0,1339	0,242	0,0001	-0,1471
28	Dp50	0,321143	-0,0569	-0,0007	0,0571	-0,0569	-0,0007	0,0786
29	Dp50	0,38532	-0,1565	-0,001	0,0687	-0,1565	-0,001	0,0902
30	Dq15	0,996529	-0,0768	-0,2506	0,0687	-0,0768	-0,2506	0,0882
31	Dq25	0,816354	-0,0768	-0,1792	0,066	-0,0768	-0,1792	0,0861
32	Dq70	0,005266	-0,0768	0,142	0,0541	-0,0768	0,142	0,0767
33	Dq85	-0,26504	-0,0768	0,2491	0,0501	-0,0768	0,2491	0,0736
34	Dr50	0,252732	0,0428	-0,0005	0,0454	0,0428	-0,0005	0,0669
35	Ds50	0,165778	0,1424	-0,0002	0,0337	0,1424	-0,0002	0,0553
36	Dt50	0,120311	0,242	0,0001	0,022	0,242	0,0001	0,0436
37	Ep50	0,407561	-0,1565	-0,001	0,1392	-0,1565	-0,001	0,1045
38	Eq25	0,89098	-0,0768	-0,1792	0,1366	-0,0768	-0,1792	0,1004
39	Eq70	0,01536	-0,0768	0,142	0,1246	-0,0768	0,142	0,091

ANNEXE : suite et fin

			ALOGP			MLOGP		
i	**code**	**log k**	**w1**	**w2**	**w3**	**w1**	**w2**	**w3**
40	Eq85	-0,276544	-0,0768	0,2491	0,1206	-0,0768	0,2491	0,0879
41	Er50	0,287578	0,0428	-0,0005	0,1159	0,0428	-0,0005	0,0812
42	Es50	0,202761	0,1424	-0,0002	0,1042	0,1424	-0,0002	0,0696
43	Et50	0,161667	0,242	0,0001	0,0925	0,242	0,0001	0,0579
44	Fp50	0,335919	-0,0569	-0,0007	0,0571	-0,0569	-0,0007	0,0786
45	Fq15	0,82173	-0,1565	-0,001	0,0687	-0,1565	-0,001	0,0902
47	Fq50	0,215505	-0,0768	-0,1792	0,066	-0,0768	-0,1792	0,0861
48	Fq70	-0,086663	-0,0768	0,142	0,0541	-0,0768	0,142	0,0767
49	Fq85	-0,225775	-0,0768	0,2491	0,0501	-0,0768	0,2491	0,0736
50	Fr50	0,160889	0,0428	-0,0005	0,0454	0,0428	-0,0005	0,0669
51	Fs50	0,086502	0,1424	-0,0002	0,0337	0,1424	-0,0002	0,0553
52	Ft50	0,047703	0,242	0,0001	0,022	0,242	0,0001	0,0436
53	Gp50	0,234061	-0,1565	-0,001	0,028	-0,1565	-0,001	0,0322
54	Gq15	0,839258	-0,0768	-0,2506	0,0279	-0,0768	-0,2506	0,0302
55	Gq25	0,66693	-0,0768	-0,1792	0,0253	-0,0768	-0,1792	0,0281
56	Gq50	0,183099	-0,0569	-0,0007	0,0163	-0,0569	-0,0007	0,0206
57	Gq70	-0,09936	-0,0768	0,142	0,0133	-0,0768	0,142	0,0187
58	Gq85	-0,229443	-0,0768	0,2491	0,0093	-0,0768	0,2491	0,0156
59	Gr50	0,130784	0,0428	-0,0005	0,0046	0,0428	-0,0005	0,0089
60	Gs50	0,063033	0,1424	-0,0002	-0,0071	0,1424	-0,0002	-0,0027
61	Gt50	0,026492	0,242	0,0001	-0,0187	0,242	0,0001	-0,0144
62	Hp50	0,479834	-0,1565	-0,001	0,1246	-0,1565	-0,001	0,2006
63	Hp50	0,53107	-0,0569	-0,0007	0,1129	-0,0569	-0,0007	0,189
64	Hr50	0,37798	0,0428	-0,0005	0,1012	0,0428	-0,0005	0,1773
65	Hs50	0,276508	0,1424	-0,0002	0,0896	0,1424	-0,0002	0,1657
66	Ht50	0,206016	0,242	0,0001	0,0779	0,242	0,0001	0,154
67	Ip50	0,519355	-0,1565	-0,001	0,1246	-0,1565	-0,001	0,2006
68	Iq50	0,443247	-0,0569	-0,0007	0,1129	-0,0569	-0,0007	0,189
69	Ir50	0,353204	0,0428	-0,0005	0,1012	0,0428	-0,0005	0,1773
70	Is50	0,276737	0,1424	-0,0002	0,0896	0,1424	-0,0002	0,1657
71	It50	0,217221	0,242	0,0001	0,0779	0,242	0,0001	0,154
72	Jp50	0,78602	-0,1565	-0,001	0,3013	-0,1565	-0,001	0,2387
73	Jq50	0,704905	-0,0569	-0,0007	0,2897	-0,0569	-0,0007	0,2271
74	Jr50	0,630855	0,0428	-0,0005	0,278	0,0428	-0,0005	0,2154
75	Js50	0,531006	0,1424	-0,0002	0,2663	0,1424	-0,0002	0,2038
76	Jt50	0,481142	0,242	0,0001	0,2546	0,242	0,0001	0,1921