

UNIVERSITE KASDI MARBAH OUARGLA

Faculté des Sciences et de la Technologie et Science de la matière

Département de science de matière



Mémoire

MASTER ACADEMIQUE

Domaine Sciences de matière

Filière Chimie

Spécialité chimie appliqué

Présenté Par : REGADI Dahmane

Thème

***Développement de modèle QSPR pour la
prédiction des propriétés physique des quelques
composés odorants***

Devant le jury :

Ms: RAHMANI Zehour	Maître Assistant (A) UKMO	Présidente
Ms: MEKADAM Khadra	Maître Assistant (A) UKMO	Examinatrice
Mr: KORICHI Mourad	Maître de conférence UKMO	Encadreur
Ms: TEKHA Mebarka	Maître Assistant (B) UKMO	Invité

Année Universitaire : 2012 /2013

CHAPITRE I

Généralités sur l'odeur

I-1- Introduction :

Jusqu'à la montée en puissance de la chimie organique à la fin du XIXe siècle, les parfums sont composés principalement à partir de produits d'origine végétale auxquels s'ajoutent quelques composants d'origine animale comme le musc ou l'ambre gris. Le traitement des plantes fournit des hydrolats (eau de rose, eau de fleur d'oranger...), des huiles essentielles et des alcoolats (alcoolat de romarin). Il existe aussi depuis très longtemps des compositions parfumées, faites de substances solides. Ainsi, l'encens, qui est probablement l'un des plus anciens parfums connus, utilise la technique de la combustion pour produire les substances odorantes. De même, la myrrhe, qui servait à la momification en Egypte, était utilisée dans la composition des parfums.

I-2- Définition d'une odeur :

Une odeur est une émanation transmise par un l'air et perçue par l'appareil olfactif. Les récepteurs olfactifs sont situés dans les fosses nasales et sont reliés au cerveau par le nerf olfactif. Il existe sept odeurs primaires qui correspondent aux sept types de récepteurs sensoriels situés sur les cils des cellules olfactives. Des substances ayant des odeurs semblables ont des molécules de forme similaire. La forme d'une molécule détermine la nature de son odeur. Ces molécules se fixent sur les récepteurs. Une même molécule peut se fixer sur plusieurs sites si elle possède la " clé " de récepteurs différents. Ce phénomène est le début d'une série d'événements : transmission de l'influx par le nerf olfactif et perception d'une odeur par le cerveau [1].

On peut distinguer sept odeurs de base : Camphrée, musquée, florale, mentholée, éthérée, piquante, putride.

I-3- l'olfaction:

Les molécules odorantes doivent être volatiles pour pouvoir atteindre l'épithélium olfactif mais elles doivent également être hydrosolubles pour pouvoir atteindre les récepteurs des cils qui baignent dans le mucus, Après la pénétration dans les fosses nasales, les molécules odorantes se dispersent dans le mucus (fine couche de glaire recouvrant la muqueuse des fosses nasales), puis vont se fixer sur les récepteurs des cils (**Figure I-1**). Cela déclenche une stimulation nerveuse puis un message qui sera transmis par l'intermédiaire des voies nerveuses olfactives jusqu'au cerveau, au niveau des lobes temporaux.

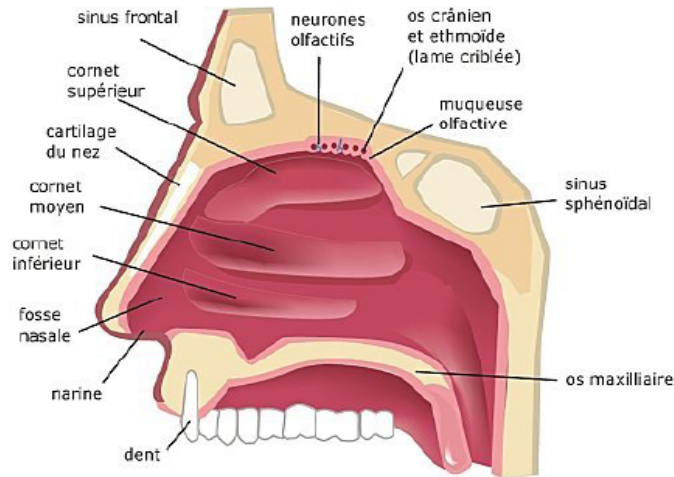


Figure I-1: représentation le nez (HARLE, 2009)

I-4- Mécanismes de l'olfaction:

L'olfaction mettant en jeu à la fois des mécanismes physico-chimiques (interaction des molécules odorantes avec les neurorécepteurs situés dans la muqueuse olfactive), physiologiques (codage et transport de l'information) et psychologique (interprétation), l'appréciation de cette sensation présente souvent une variabilité importante. La notion d'odeur est donc particulièrement subjective et il est délicat de caractériser de manière fiable une nuisance olfactive. De plus, on perçoit souvent des mélanges de composés odorants, ce qui complique l'étude car des phénomènes de synergie ou d'inhibition sont mis en jeu [Thomas-Danguin, 1997]. Il n'existe donc pas de corrélation entre la grandeur de la sensation et les concentrations des molécules odorantes [2].

I -5- Les propriétés physico-chimiques des composés odoriférants:

- La volatilité: une molécule odorante doit avoir une masse moléculaire moyenne (30 à 300) g/mol.
- La concentration: est importante pour la perception des odeurs par l'homme (seuil de détection à 10⁻¹⁷ ou 10⁻¹⁸ mol/L).
- La polarité: une molécule odorante est souvent apolaire:



Figure I-2: présentation de polarité des composés odoriférants.

- La solubilité: les molécules odorantes ont un caractère lipophile.
- La structure: les odeurs varient selon la structure de la molécule.
- La chiralité: les énantiomères donnent des odeurs différentes malgré le fait que les molécules soient identiques:

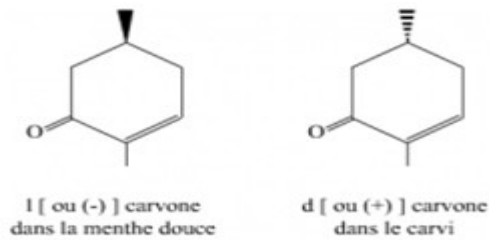


Figure I-3: Énantiomères de carvone

- L'isomérisie: les molécules qui ne possèdent la même formule brute mais pas la même formule semi-développée et développée donnent des odeurs différentes.

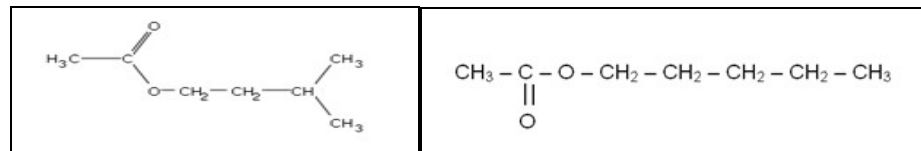


Figure I-4: Formule semi développée de l'acétate d'isoamyle [3,4].

I -5-1- Odeurs et propriétés physiques :

Les deux propriétés physiques qui jouent un rôle très important sont la tension de vapeur et la solubilité dans le mucus.

La tension de vapeur ou pression de vapeur saturante est la pression partielle. Maximale qu'on peut avoir en phase gazeuse au dessus du solide ou du liquide, pour une substance donnée à une température donnée. Cette constante physique est relativement facile à mesurer (il existe des valeurs expérimentales très nombreuses) et à calculer à partir de la structure. L'effet bien connu d'augmentation de la pression de vapeur avec la température ne sera pas considéré ici. La pression de vapeur saturante diminue lorsque la masse moléculaire et la température d'ébullition augmentent. On considère que la limite au delà de laquelle ce facteur

rend les molécules inodores se trouve vers une masse molaire de 300. La molécule de plus haute masse connue comme odorante est un dérivé du labdane de masse 294.

La solubilité dans le mucus pose un tout autre problème pour au moins deux raisons:

- La première: est la difficulté d'une utilisation reproductible du mucus.
- La seconde: est la complexité du mucus qui contient, en solution aqueuse, des protéines, des sucres et des sels minéraux. Contrairement à la grandeur précédente, la solubilité dans le mucus est difficile à mesurer et encore plus difficile à estimer. Une solution, peu satisfaisante, consiste à considérer le mucus comme de l'eau et à utiliser les nombreuses données sur la solubilité dans l'eau [1].

I -6- La perception des odeurs:

I -6-1- Seuil de détection olfactif et seuil de reconnaissance olfactif

Le seuil olfactif ou seuil de détection olfactive peut représenter :

La concentration de la substance la plus faible qui puisse être sentie par un groupe de sujets alors que le seuil de détection olfactive signifie que le sujet a une sensation olfactive sans pour autant être capable d'identifier ou de reconnaître la substance qui est présente dans l'air, le seuil de reconnaissance est la concentration dans l'air d'une substance qui permet au sujet de la reconnaître ou de l'identifier. Le seuil de reconnaissance représente en général une concentration plus importante que le seuil de détection olfactive. La distinction entre le seuil de détection olfactive et le seuil de reconnaissance olfactive n'est pas toujours explicite, ce qui peut mener à confondre.

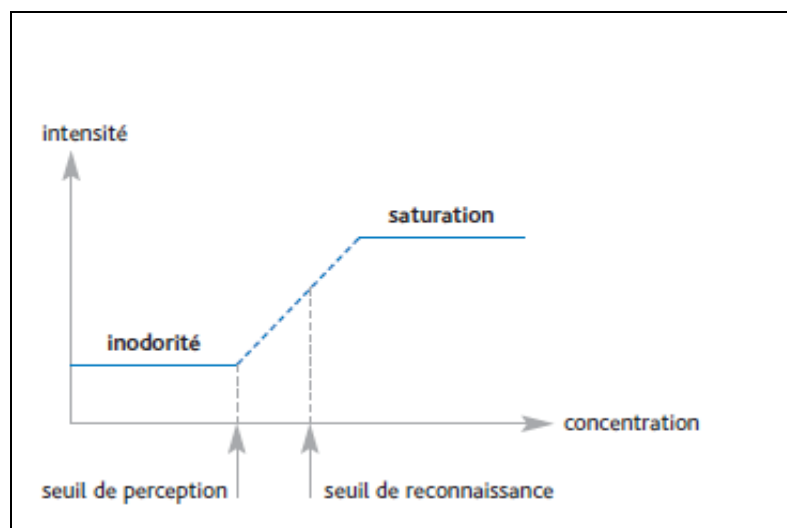


Figure I-5 : Explication de loi de Steven.

- Loi de Steven : relation entre la concentration de l'odeur et son intensité. Plus la concentration augmente, plus l'intensité augmente jusqu'à saturation.

Ces deux valeurs de seuils. Par conséquent, il se peut que les valeurs fournies pour le seuil olfactif dans les présentes fiches toxicologiques soient des valeurs représentant le seuil de détection olfactive ou le seuil de reconnaissance d'un produit donné. De plus, ces seuils peuvent avoir été définis par un individu ou par un groupe d'individus [5].

I -6-2-Sensibilité olfactive et seuils de perception

On peut distinguer deux valeurs de seuils: le seuil de détection, lorsque le sujet a une sensation olfactive sans pour autant affecter de label à cette odeur, et le seuil de reconnaissance, lorsque la qualité de l'odeur est accessible au sujet. Plusieurs méthodes sont utilisées pour mesurer la sensibilité olfactive chez l'homme. La plupart de ces procédures utilisent des présentations par paires, et elles tirent leur spécificité par le mode de détermination des seuils. La détermination des seuils, se fait généralement en utilisant une gamme de concentrations croissantes d'un odorant mis en solution dans un solvant [6].

Trois grandes méthodes peuvent être définies :

- La première technique consiste à présenter de manière croissante des concentrations, jusqu'à ce que le sujet détecte correctement le bon flacon. La valeur seuil est prise comme la première concentration des 3 ou 4 successivement correctes.
- Dans la seconde technique, les concentrations sont présentées de manière croissante et décroissante, jusqu'à ce que le sujet stabilise ses réponses selon un procédé bien défini.
- Enfin, la troisième technique, développée par Doty et al. (1986). Les seuils sont déterminés grâce à une procédure de choix forcés. L'expérience comporte plusieurs essais. Un essai consiste en la présentation successive rapide de 2 flacons : un contenant du solvant et l'autre contenant l'odorant étudié, dilué dans le solvant [7].

I-6-3- Solubilité des odeurs

Pour être odorante, une substance doit être légèrement soluble dans l'eau, donc posséder une partie hydrophile pour se dissoudre au niveau du mucus sécrété par l'épithélium olfactif recouvrant les cellules nerveuses sensorielles. Afin d'atteindre les récepteurs olfactifs, la substance doit ensuite pouvoir traverser la membrane cellulaire constituée de lipides (graisses). La solubilité dans les lipides doit être suffisante : elle doit donc présenter une partie lipophile [8].

I-6-4- Volatilité des odeurs :

Un corps qui ne possède pas de molécules volatiles n'émet pas d'odeur (ex: le verre). Pour qu'un corps soit odorant, il doit d'abord être assez volatil pour être facilement vaporisé à température ordinaire et atteindre les récepteurs olfactifs.

La volatilité des constituants d'un parfum est une contrainte chimique qui doit être maîtrisée par l'industrie des parfums. Ainsi, si les composés sont trop volatils, l'odeur va rapidement se dissiper.

C'est pourquoi, un parfum contient des molécules lourdes inodores. Non volatiles, elles exercent un effet fixateur du parfum lui conférant une plus grande ténacité [6].

I -7-Structures moléculaires odoriférantes :

Les composés odorants peuvent être classés selon la principale fonction qu'ils possèdent ; comme [9] :

- Les hydrocarbures ou essences, surtout, des alcènes (ex : limonène)
- les alcools, dont les chaînes carbonées comportent huit à douze carbones (ex : le menthol ou autre que l'on trouve dans les fleurs).

I -8-Les sources d'odeurs:

Les activités susceptibles de provoquer des problèmes d'odeurs sont relativement nombreuses. On peut citer par exemple [10] :

- Industries agroalimentaires
- Raffineries de pétrole.
- Industries chimiques.

I -9-Système olfactif:

Le système olfactif est le système sensoriel permettant de détecter et d'analyser les molécules odorantes présentes dans l'air. Le système olfactif est composé d'une structure périphérique permettant la chimioréception des molécules odorantes et de structures centrales assurant l'analyse et le traitement de l'information olfactive (Figure 1) [11].

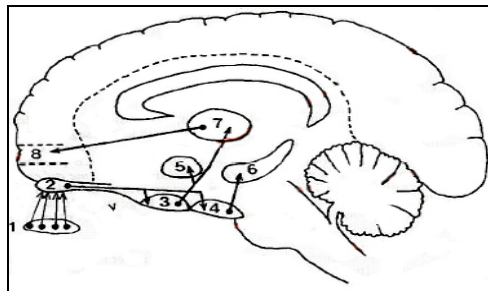


Figure I-6: Représentation schématique des projections du système olfactif dans le système nerveux central

I -10-La transduction du message olfactif:

Lors d'une inspiration, les molécules odorantes portées par l'air entrent et circulent dans la cavité nasale. Les odeurs sont des molécules volatiles et hydrophobes. Avant d'atteindre les récepteurs moléculaires aux odeurs portés par les cils des neurorécepteurs olfactifs les molécules odorantes doivent traverser une couche aqueuse de mucus. C'est dans ce mucus que sont sécrétées les « olfactory binding proteins » (OBPs), de petites protéines hydrophiles capables de se lier de manière réversible aux molécules odorantes et de les transporter dans le mucus aqueux afin de faciliter leur accès aux récepteurs moléculaires aux odeurs Les neurorécepteurs olfactifs expriment une protéine G spécifique à la voie olfactive [12].

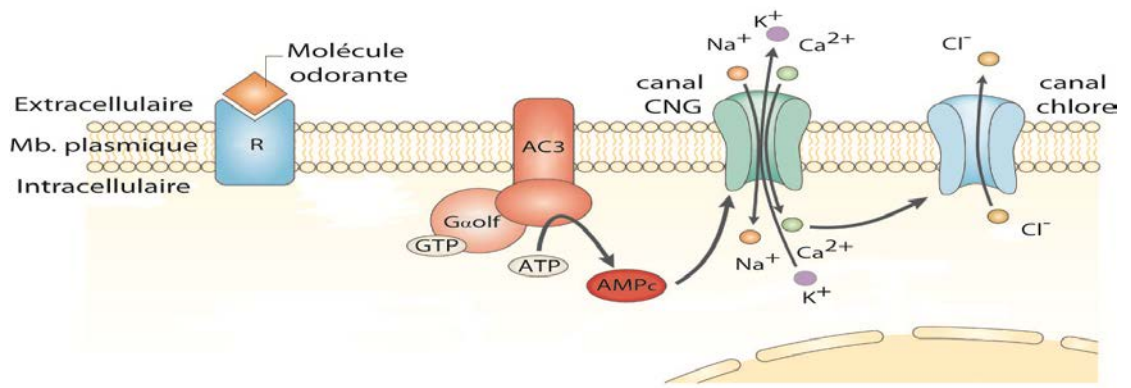


Figure I-7 : La voie principale de transduction chimio-électrique olfactive

I -11-Récepteur d'odeur:

Les RO appartiennent à la superfamille de récepteurs couplés aux protéines G (RCPG) . Il s'agit d'une large famille de récepteurs membranaires capable de reconnaître et d'effectuer la transduction de signaux extracellulaires aussi variés que des photons, des odorants, des phéromones, des oses ou des macromolécules.

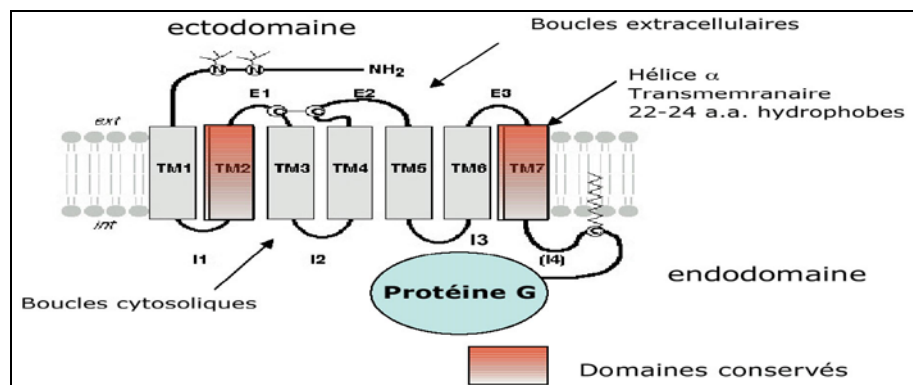


Figure I-8: Représentation schématique d'un récepteur couplé aux protéines G

I -12-Méthodes des mesures d'odeur:

Il ya trois approches complémentaires sont utilisées :

- l'analyse olfactométrique basée sur une évaluation sensorielle, qui cible l'odeur elle-même et en permet une appréciation globale ;
- l'analyse physico-chimique qui cible les molécules responsables de l'odeur et permet l'identification de sa composition tant qualitativement que quantitativement.
- Des appareils appelés « nez électroniques » et basés sur la reproduction du système olfactif par des capteurs sont également en développement depuis une vingtaine d'années. Néanmoins leur application dans le domaine de l'environnement reste limitée et cantonnée à des missions surveillance [13].

I -12-1-L'analyse physico-chimique

La difficulté de la mesure physico-chimique résulte du fait que l'analyste est potentiellement confronté à plusieurs centaines de composés et que la classification admet une multitude de familles. Il existe donc un nombre conséquent de techniques d'échantillonnage et d'analyse possibles. Les informations fournies par la mesure peuvent être globales (mesure représentative de toute une famille ou d'un marqueur spécifique) ou spécifiques (identification et quantification individuelle de chaque molécule).

I -12-1-L'analyse olfactométrique

Cette technique repose sur l'utilisation du nez humain pour mesurer à la fois le niveau et l'intensité d'une odeur. La perception d'une odeur varie d'un individu à l'autre mais aussi pour un même individu en fonction de son état physiologique et des conditions de flairage. Ainsi, un jury de nez représentatif de la population est utilisé pour ces analyses. Le jury est composé de 4 à 16 personnes selon la précision désirée.

Les membres sont formés à la technique et sélectionnés pour la répétabilité de leurs résultats [14].

Il existe deux types d'olfactométrie: l'olfactométrie dynamique, pour laquelle l'échantillon gazeux est dilué en continu dans un flux d'air inodore et l'olfactométrie statique, pour laquelle la dilution est obtenue dans une enceinte fermée en introduisant un volume connu de l'échantillon dans un volume connu de gaz inodore [13,15].

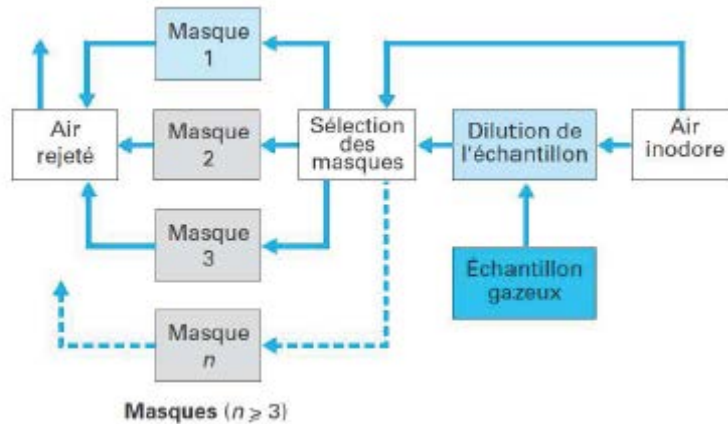


Figure I-9: Schéma du principe d'un olfactomètre dynamique.

I-13-Les différentes grandeurs mesurées en olfactométrie et les modes opératoires

La nuisance olfactive peut être caractérisée par trois mesures en olfactométrie : la concentration d'odeur (détection de l'odeur), l'intensité (quantification de l'odeur) et le caractère hédonique (appreciation de l'odeur) [16].

I -13-1-Mesure du facteur de dilution au seuil de perception

La concentration ou le niveau d'odeur a préalablement été défini au paragraphe II.3 et correspond au facteur de dilution (parfois noté K50 en France). La méthode de mesure est normalisée au niveau européen (norme EN 13725) . Celle-ci préconise d'exprimer la concentration d'odeur à partir de l'Unité Odeur Européenne (**ouE**) qui correspond à la quantité de composé qui, présente dans 1 m³ d'air, conduit au seuil de perception du jury. Ainsi, pour un hypothétique échantillon dont le facteur de dilution est égal à 2, la concentration d'odeur vaut 2 (ouE.m⁻³).

Deux modes de réponses sont préconisés : la méthode à « choix forcé » et la méthode « oui/non ». La concentration d'odeur est déduite des réponses du jury après un traitement statistique [13,16].

I -13-2-Mesure de l'intensité d'odeur

L'intensité d'une odeur est définie dans la norme NF X 43-103 comme « la grandeur de la sensation pour un stimulus supérieur à celui correspondant au seuil de perception » . Elle est fonction de la concentration du mélange odorant. La technique consiste à demander aux membres du jury de classer les échantillons, sur la base de leur intensité, en les comparant avec une gamme de référence odeur (niveaux supraliminaire) constituée généralement avec du butanol-1 ou de la pyridine.

En flairant directement l'échantillon, il est également possible de qualifier le caractère hédonique d'une odeur (plaisant, déplaisant) ainsi que son origine supposée (naturelle, urbaine, agricole ou industrielle).

Cette mesure reste néanmoins particulièrement subjective et peu utilisée [13, 14,17].

I -13-3-L'olfactométrie

Le terme 'olfactométrie' désigne à la fois la mesure des odeurs et la mesure des capacités olfactives d'un sujet. Dans les deux cas un capteur est mis en présence d'une odeur (le stimulus). Un stimulus connu permet de caractériser le capteur, un capteur étalonné permet d'objectiver le stimulus [6]. Les paramètres d'une odeur sont [10]:

- Quantitatif : son intensité, sa force.
- Qualitatif : sa description comme la référence à un objet odorant (ex. la rose) ou la constitution chimique du mélange odorant.
- Temporel : l'évolution dans le temps de son intensité et/ou de sa qualité (ex. note de tête d'un parfum).

Dans de nombreux domaines il s'avère nécessaire d'évaluer une odeur par exemple dans la parfumerie et les cosmétiques, le contrôle de qualité et les tests de préférence dans l'agroalimentaire, le traitement des nuisances olfactives, etc. Les méthodes utilisées pour mesurer les odeurs sont l'analyse physico-chimique et l'utilisation d'un 'jury de nez' entraîné. Dans le cas de l'évaluation de nuisance olfactive on utilise aussi les enquêtes auprès des populations concernées. Enfin des nez artificiels ou nez électroniques commencent à apparaître sur le marché.

La caractérisation des sources émettrices d'odeurs ; soit par [6,9] :

- l'analyse olfactométriques : résultats que ce type d'analyse peut fournir est le niveau d'odeur, le débit d'odeur, l'intensité de l'odeur ou encore l'aire de persistance de la nuisance. Les principes de mesures de niveau d'odeur sont décrits dans la norme AFNOR NFX 43-101.
- l'analyse physico-chimique : permet l'identification et la quantification des composés incriminés (analyse complexe car plusieurs centaines de composés dans les effluents).

La méthode physicochimique consiste à rechercher et quantifier les éléments chimiques présents dans l'atmosphère puis à se reporter à une table des propriétés olfactives de chaque corps pour caractériser l'odeur résultante. Cette méthode a des limites :

- la concentration des produits odorants est souvent si faible qu'ils ne sont pas détectables, même par les analyseurs les plus performants.

- les propriétés olfactives des mélanges sont différentes de celles des constituants pris séparément, et souvent varient avec la concentration.
- tous les composés chimiques ne sont pas odorants, mais ils peuvent influencer la sensation perçue avec le mélange.

I -13-4-Grandeurs olfactométriques (quelques définitions) :

- Seuil olfactif : pour chaque corps pur ou mélange odorant, on peut définir une concentration seuil pour laquelle l'effluent est ressenti comme odorant par 50 % des membres d'un jury constituant un échantillon de population. Dans le cas d'un corps pur, cette concentration est appelée par convention « seuil olfactif » .
- Niveau d'odeur : ce niveau est défini conventionnellement comme étant le facteur de dilution qu'il faut appliquer à un effluent pour qu'il ne soit pas ressenti comme odorant par 50 % des personnes d'un jury constituant un échantillon de population. On parle aussi de « Facteur de dilution au seuil de perception ».
- Débit d'odeur : le débit d'odeur est défini conventionnellement comme étant le produit du débit d'air rejeté, exprimé en Nm^3/h , par le facteur de dilution au seuil de perception. Combiné à un modèle de dispersion atmosphérique, cet indicateur permet de déterminer une aire de persistance de la nuisance en fonction des conditions météorologiques.
- Intensité d'odeur : l'intensité d'odeur ou intensité odorante caractérise la grandeur de la sensation olfactive. Sa mesure, réalisée par un jury entraîné, consiste à comparer l'intensité du mélange gazeux à l'intensité d'échantillons de référence.
- L'odorité: qui est la propriété que possède un corps ou un mélange d'avoir une odeur, c'est-à-dire d'être un stimulus olfactif.
- Le gène d'odeur: qui est le caractère attribué à une odeur soit pour sa qualité odorante, soit pour la fréquence avec laquelle elle est perçue. En dessous d'une certaine concentration appelée le seuil de perception, les odeurs ne sont plus perceptibles par 50% de la population [10,14].

I-14-Le jury de nez:

Pour un spécialiste des études de nuisances olfactives le terme 'olfactométrie' désigne la mesure des odeurs à l'aide d'un jury de nez. Par rapport à un instrument d'analyse l'être humain fait une évaluation directe de l'odeur.

Un jury est constitué de quatre à seize sujets sélectionnés sur leurs capacités olfactives, ils doivent être représentatifs de la moyenne de la population. Ces sujets sont entraînés en

fonction de la tâche qui leur sera confiée : s'ils doivent décrire des odeurs complexes leur entraînement consistera à détecter et nommer les constituants d'un mélange, s'ils doivent mesurer des intensités ils classeront par ordre concentration croissante des solutions. Si ces solutions sont distribuées dans une large gamme il est alors possible de déterminer les seuils de détection des sujets pour le produit considéré.

Ce jury, une fois entraîné, est placé dans les situations olfactives pour caractériser une source de nuisance, tester l'efficacité d'un procédé de désodorisation, évaluer la dilution d'une odeur dans l'environnement [7,10].

I-15-Les nez artificiels:

Un nez artificiel est constitué d'un ensemble de capteurs chacun faiblement sélectif à un composé chimique. Il existe des capteurs de type oxydes métalliques, de type polymères conducteurs et des biosenseurs à base de bicouches lipidiques. Ces capteurs sont associés à un dispositif de traitement des informations par réseau de neurone formel de type perceuteur multicouche.

Après une période d'entraînement, supervisée à l'aide d'une règle de rétropropagation de l'erreur, ces dispositifs sont capables d'identifier certains mélanges mais surtout d'évaluer les différences d'un mélange odorant avec un prototype. Les performances de ces nez électroniques sont très inférieures à celles d'un nez humain mais ils sont infatigables, c'est pourquoi ils commencent à être utilisés dans le contrôle de qualité principalement dans l'agro-alimentaire [6,8].

I-16-Les nuisances:

Les odeurs sont généralement dues à une multitude de molécules différentes, en concentration très faible, mélangées à l'air que nous respirons. La plupart des composés odorants sont détectés à des niveaux très faibles par rapport aux niveaux toxiques. A l'inverse, des gaz très toxiques comme le monoxyde de carbone n'ont aucune odeur.

Les nuisances sont liées à la perception des odeurs agréables ou désagréables, aimées ou détestées. L'odeur dépend de la dilution plus ou moins forte (d'où une intensité plus ou moins élevée) d'une ou d'un mélange de substances chimiques dans l'atmosphère [9].

Les substances appartiennent aux principaux composés suivants :

- composés azotés (amine, ammoniac...)
- acides gras volatils,
- aldéhydes et cétones,
- composés soufrés (hydrogène sulfuré, mercaptans, sulfures et disulfures...)

➤ mélange de ces composés.

Les substances sont issues de décomposition thermique ou anaérobie de composés chimiques, de produits animaux ou de déjections animales.

I-17-Conclusion:

Les odeurs et les nuisances olfactives sont des préoccupations environnementales dont l'importance est croissante, aussi bien du côté des industriels (milieu émetteur) qui cherchent à maîtriser ces nuisances que du côté de la population riveraine (milieu récepteur) qui exige le respect de son cadre de vie.

Leurs effets sont difficiles à caractériser de manière précise mais les nuisances olfactives doivent être prises en compte en matière de qualité de l'air car leurs conséquences sur la santé au sens large sont indéniables.

Les nuisances olfactives apparaissent comme le deuxième motif de plainte après le bruit leur association à une notion de toxicité est rarement justifiée sur le plan physiologique, les odeurs étant le plus souvent perçues à des concentrations très faibles, inférieures aux limites acceptables pour la santé.

Le seuil de perception olfactive peut varier couramment d'un facteur 10 à 100 entre des personnes différentes ou pour une même personne en fonction de nombreux facteurs (humidité relative, température, présence, d'autres composés dans l'air, fatigue...).

CHAPITRE II

Les outils et techniques utilisés

II-1-Introduction:

Les relations quantitatives structure-activité/propriété (QSAR/QSPR) sont de plus en plus utilisées, du fait de la croissance des moyens de calculs. Très récemment, la mise en place du nouveau règlement européen REACH, qui recommande leur utilisation pour limiter le recours à l'expérience, donne un nouvel essor au développement de tels modèles prédictifs.

Dans les dernières années, l'utilisation des méthodes QSAR n'a cessé de progresser. Elle est même devenue indispensable en chimie pharmaceutique et pour la conception de médicaments. Leur développement dans une gamme plus large d'applications leur ouvre d'ailleurs de grandes perspectives (ex : points d'ébullition, températures critiques, densité).

Il s'agit de présenter ici le principe des modèles QSPR ainsi que ceux des différents outils employés pour leur mise en place et leur évaluation : bases de données expérimentales, descripteurs, outils d'analyse de données [18].

II -2- Les relations structure-propriétés quantitatives "QSPR":

II -2-1-Définition de QSPR

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

- Choisir des descripteurs adaptés au problème structure-propriété,
- Exploiter les valeurs des descripteurs comme variables, afin de définir une relation qui les corrèle à la propriété en question, à l'aide de machines d'apprentissage. C'est la fouille de données.
- Établir des critères de performance et de validation qui aideront au choix des meilleurs modèles pour le problème posé et estimer des incertitudes de prédiction [19].

II-2-2- Principe de QSPR

Le principe des méthodes QSPR est, comme leur nom l'indique, de mettre en place une relation mathématique reliant de manière quantitative des propriétés moléculaires aussi bien électroniques que géométriques, appelées descripteurs, avec une observable macroscopique (activité biologique, toxicité, propriété physico-chimique, etc.), pour une série de composés chimiques similaires à l'aide de méthodes d'analyses de données. Aussi la forme générale de modèle est:

$$\text{Propriété} = f(\text{Descripteurs}) \quad (\text{II.1})$$

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

Une fois cette relation mise en place et validée sur un jeu validation, elle peut alors être employée pour la prédiction de la propriété de nouvelles molécules, pour lesquelles la valeur expérimentale n'est pas disponible, voire pour des molécules encore non synthétisées. De tels modèles peuvent également, dans certains cas, être utilisés pour mieux appréhender les phénomènes moléculaires mis en jeu dans la propriété d'intérêt[20].

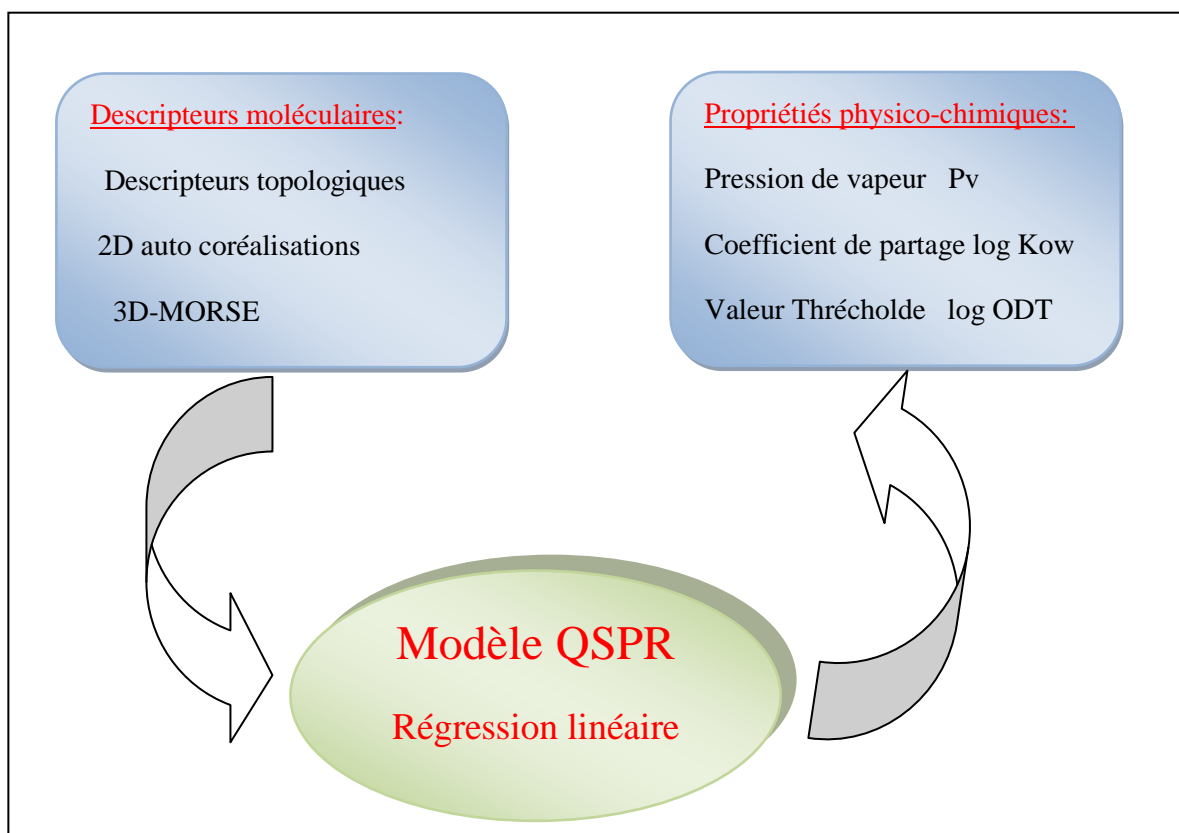


Figure II-1: principe de la méthode QSPR

II-2-3- Les méthodes mathématiques utilisés par le model QSPR

- Régression linéaire
- Régression linéaire multiple (MRL)
- Régression en composantes principales(PCR)
- Régression moindre carré (PLS)

II-2-4- Les types d'apprentissages d'un modèle QSPR [21]

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

- Non-supervisé: les activités et propriétés ne sont pas utilisées (seul les descripteurs).
Ce type est utilisé pour grouper des composés chimiques.
- Supervisé: établit une relation entre données entrées / input data (composés chimiques) et données produites output data (propriétés).

II-2-5-Techniques d'apprentissages utilisés en chimie

On a plusieurs techniques, le tableau suivant illustre les plus utilisées:

Tableau II-1 : différents types d'apprentissages

non-supervisé	supervisé
Réseau De Kohonen	Arbres de décision
analyse de composant principal	RL, RLM, SPL, PCR
Outils de Similarité	Machines de vecteur de soutien
Grouper	Algorithmes génétiques

II-3- Descripteurs moléculaires:**II-3-1 Définition d'un descripteur**

Le descripteur moléculaire est le résultat final d'une procédure logique et mathématique qui transforme l'information chimique chiffrée dans une représentation symbolique d'une molécule dans un nombre utile ou le résultat de quelque expérience standard.

Les descripteurs moléculaires sont les traits communs les plus considérables de structure moléculaire qui peut être utilisée pour développer la « Relation Structure – property ». Dans notre cas, la propriété est pression de vapeur ou coefficient Kow d'une molécule odorante. Beaucoup des descripteurs moléculaires ont été proposés dérivés de théories différentes et approches avec le but de prédire biologique et propriétés physico-chimique des molécules [22].

II-3-2 Bloc des descripteurs [23] :

Les descripteurs moléculaires calculés par **DRAGON** sont divisés en 20 blocs logiques. Il existe quatre types de descripteurs moléculaires 0D, 1D, 2D, 3D.

- Descripteurs moléculaires 0D, 1D : il y a 5 blocs, contiennent des nombres d'atomes, des groupements fonctionnels, des propriétés des molécules, et les charges.

- Descripteurs moléculaires 2D : il y a 9 bloc, possède des descripteurs simples et autres dérivées des algorithmes appliqués à une représentation topologique.
- Descripteurs moléculaires 3D: il y a 6 bloc, contiens des descripteurs moléculaires dérivés d'une représentation géométrique appelé aussi descripteurs géométriques.

Tableau II-2 : Les paramètres topologiques après le calcul par logiciel DRAGON

ZM1	premier index M1 de Zagreb
ZM1V	premier index de Zagreb par des degrés de sommet de valence
ZM2	second lieu index m2 de Zagreb
ZM2V	second lieu index de Zagreb par des degrés de sommet de valence s
Qindex	Index quadratique
SNar	Topologique simple de Narumi
HNar	Index topologique harmonique de Narum
GNar	Index topologique harmonique de Narumi

Tableau II-3: Les paramètres 2D auto corrélation. Après le calcul par logiciel DRAGON.

ATS1m	Auto corrélation de Broto-Moreau d'une structure topologique - le retard 1/a pesé masses atomique de b
ATS3m	Auto corrélation de Broto-Moreau d'une structure topologique - le retard 3/a pesé près
ATS4m	Auto corrélation de Broto-Moreau d'une structure topologique - le retard 4/a pesé près
ATS1v	Auto corrélation de Broto-Moreau d'une structure topologique - le retard 1/a pesé près
ATS2e	Auto corrélation de Broto-Moreau d'une structure topologique - le retard 2/a pesé près
ATS1p	Auto corrélation de Broto-Moreau d'une structure topologique - le retard 1/a pesé près
MATS1m	Auto corrélation de Moran - le retard 1/a pesé par les masses atomiques
MATS1v	Auto corrélation de Moran - le retard 1/a pesé par les volumes atomiques de van der Waals

II-3-3- Exemple: calcul ATSKw

Broto-Moreau auto corrélation L'ATSKw, w étant la propriété atomique pesait le graphique moléculaire et k le retard, est évalué en considérant séparément toutes les contributions de chaque longueur de chemin différente (retard) dans le graphique moléculaire, comme rassemblé en matrice topologique de distance. En d'autres termes, tout la auto corrélation spatial au lagk ATSKw est obtenu en additionnant tout le $\sum W_i$ des produits W_j de toutes les paires d'atomes i et j, pour lesquels la distance topologique égale le retard comme :

$$ATSKw = \sum_{i=1}^{nSK-1} \sum_{j>i} W_i \cdot W_j \cdot \delta_j \dots \dots \dots II - 2$$

- nSK est le nombre d'atomes de molécule d'non-Hydrogène
- d_{ij} est le delta de Kronecker ($d_{ij}=1$ si $d_{ij} = k$, zéro autrement, d_{ij} étant la distance topologique entre deux atomes considérés).

Les blocks des descripteurs moléculaires calculés par DRAGON sont représentés dans la figure suivant :

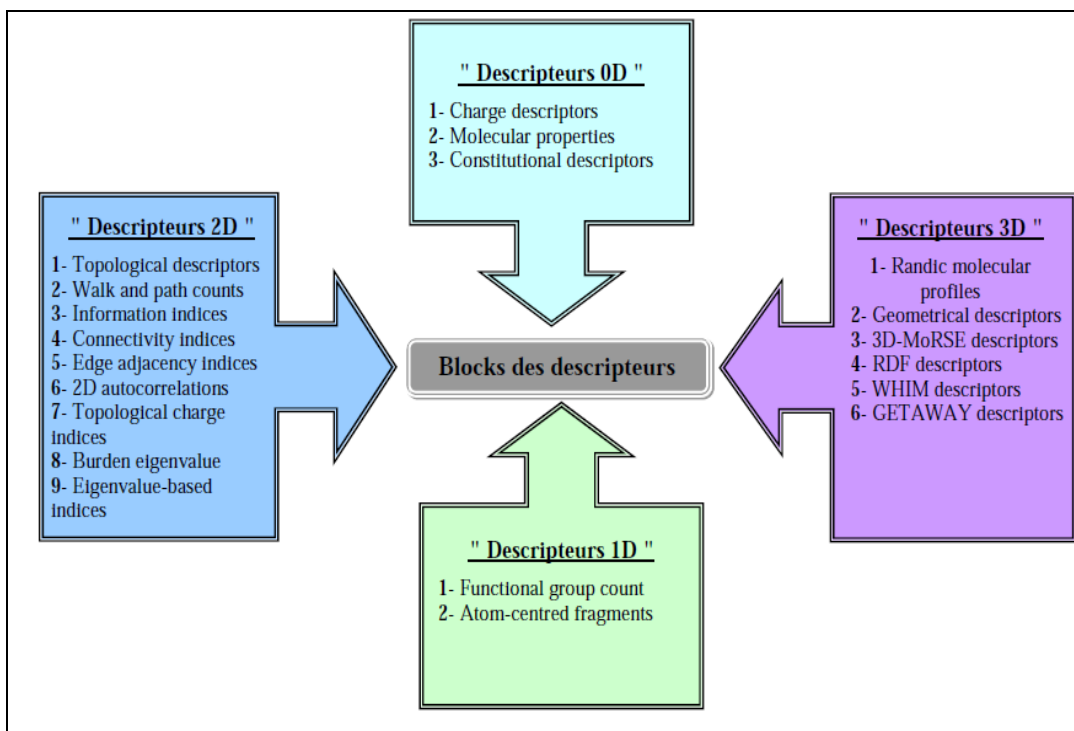


Figure II-2: Représentation des blocs des descripteurs moléculaires [23].

II-4- Importance des descripteurs:

Les descripteurs moléculaires jouent un rôle fondamental en chimie, sciences pharmaceutiques, la protection de l'environnement, recherche de la santé et contrôle de qualité, être obtenu quand les molécules sont transformées dans une représentation moléculaire qui autorise quelque traitement mathématique. Les descripteurs moléculaires sont très importants pour :

- Indiquons la description de la configuration de la molécule à étudier.
- Décrivons tous les paramètres descriptifs de la molécule.

Les descripteurs moléculaires sont utilisés pour, une connaissance de statistiques, chimio métriques, et les principes des approches QSAR/QSPR sont nécessaires en plus de la connaissance spécifique du problème [24].

II-4-1- Les descripteurs utilisés:

Les descripteurs utilisés dans cette étude sont calculés avec DRAGON, et sont divisés en deux groupes logiques qui représentée au-dessous sur un diagramme :

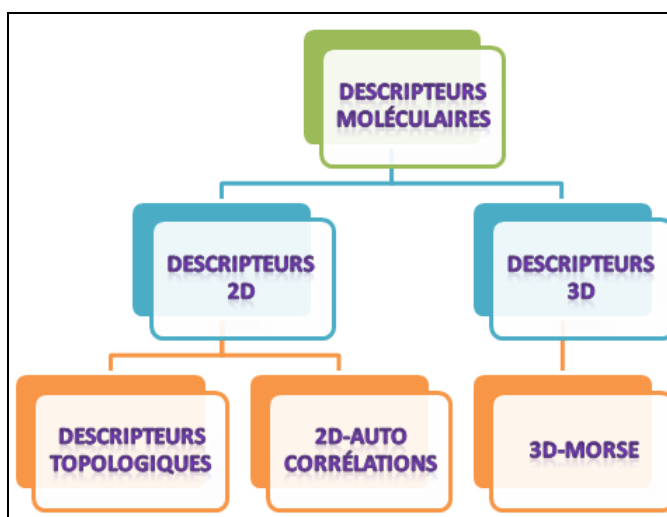


Figure II-3: Diagramme des descripteurs utilisés

A partir le diagramme précédent on choisi deux block des descripteurs moléculaires, descripteurs topologiques et descripteurs 3D-MoRSE.

II-5-Diagramme montre les étapes de prédiction :

Les premiers essais de modélisation QSPR développé par Wiener reliant point d'ébullition et structure moléculaire QSAR. Depuis, l'essor de nouvelles techniques de modélisation par apprentissage, linéaires d'abord, puis non linéaires, ont permis la mise en

place de nombreuses méthodes ; elles reposent pour la plupart sur la recherche d'une relation entre un ensemble de nombres réels, descripteurs de la molécule, et la propriété ou l'activité que l'on souhaite prédire.

Nous montrerons tout d'abord comment les molécules peuvent être représentées par des vecteurs de réels, et comment ces descripteurs sont sélectionnés. Nous introduirons ensuite les outils de modélisation les plus utilisés, c'est-à-dire la régression linéaire et la régression non linéaire à l'aide de logiciel XLSTAT, qui sont fondés sur le calcul de descripteurs.

La prédiction des propriétés s'effectue par plusieurs étapes intéressantes illustrant dans le diagramme suivant :

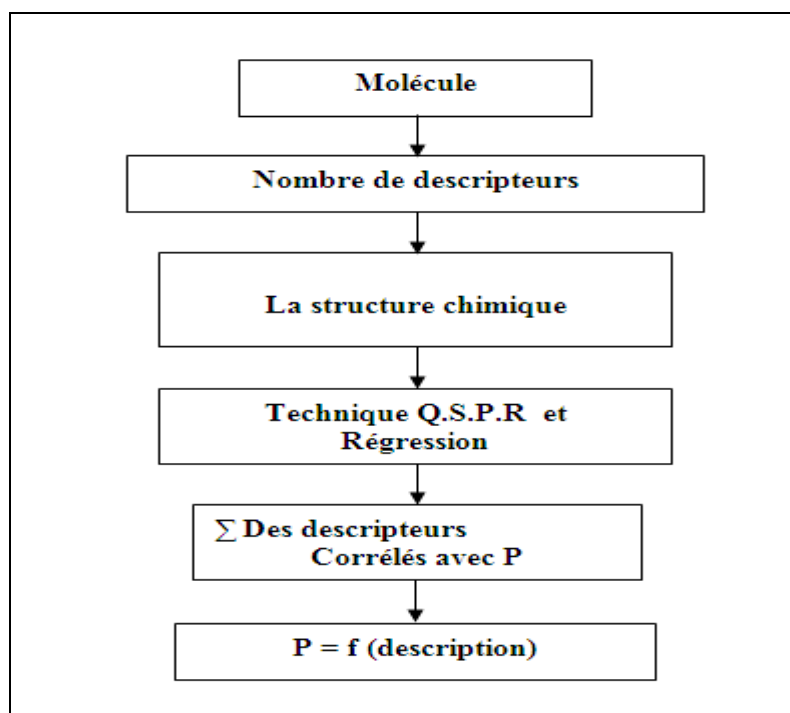


Figure II-4-Diagramme de prédiction par QSPR.

II-6- L'objectif de la prédiction :

L'objectif principal est d'établir un modèle de prédiction de propriété en se basant sur la structure moléculaire et plus précisément les descripteurs moléculaires. Ce modèle permet, en suite la classification des composés subi à la prédiction.

Le cycle de prédiction est :

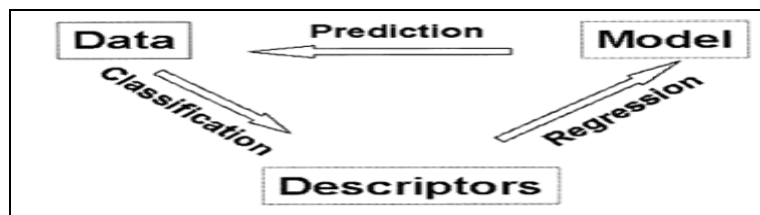


Figure II-5-montre le cycle d'optimisation du modèle.

II-7- Les étapes de travail: (Diagramme de travail)

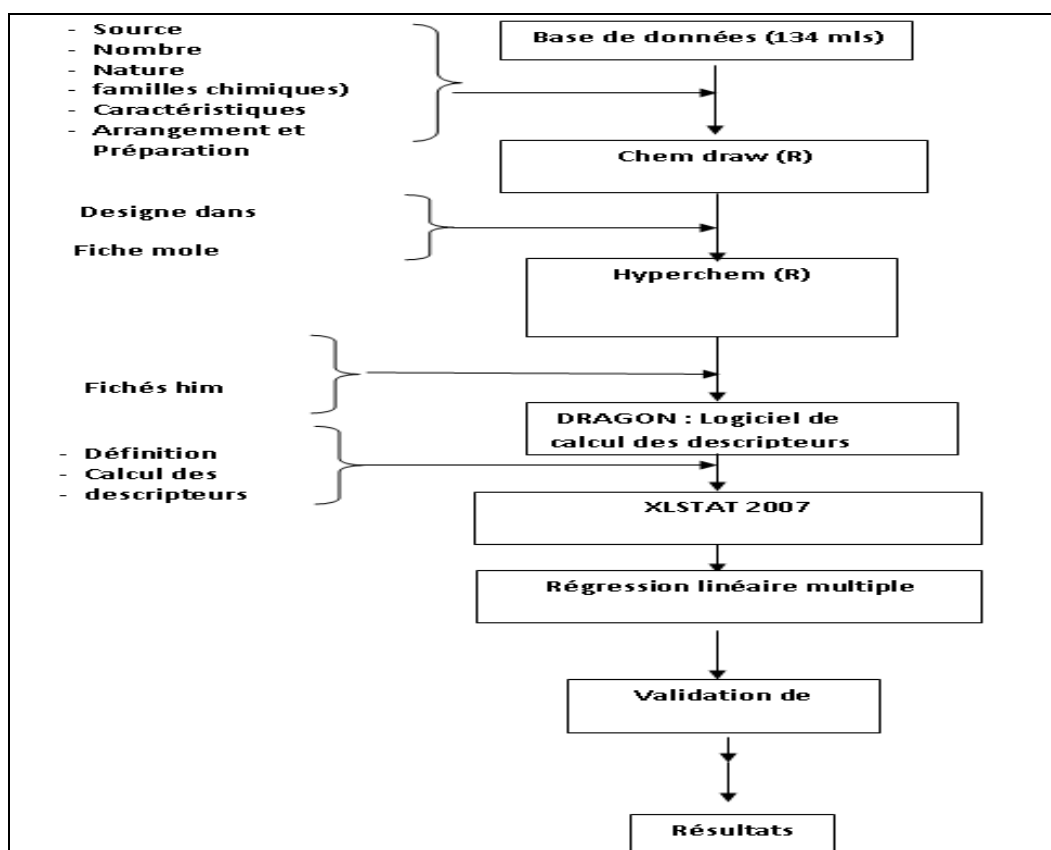


Figure II-6-Diagramme de notre travail.

Le diagramme précédant illustré les étapes de travail et les techniques quant utilisé dans la procédure de prédiction des propriétés étudiés (Pv, log K_{ow}, log ODT)

II-8- Base des données :

- La source : la base de données utilisée dans ce travail a été obtient à partir de travail qui porté le titre suivant :

- ✓ Prediction of odor detection thresholds using partition coefficients.

Oscar Rodriguez, Miguel A. Teixeira, Alírio E. Rodrigues.

- La nature des composées : ces composées sont de nature chimique.
- Les familles chimiques : dans cette base de données on trouve plusieurs familles chimiques comme (les l'aldéhydes ,les l'alcools , , les l'esters , les l'éthers ,les acides carboxyliques ,les acide amines , ... etc.)

II-8-1-Les caractéristiques et les propriétés de cette base de données sont:

- Physiques : tous les caractères physiques des composées organiques.
- Chimiques : toutes les caractères chimiques et biochimiques du composées organiques.
- Elle est présentée sous la numérotation CAS.
- Présenté les familles différents des composés odorants.

II-8-2-La raison du choix de cette base des données:

- Pour l'étude des aspects chimiques des différents types des composées odorants.
- Cette base de données traitée des propriétés jouent un rôle important dans l'olfaction.
- L'arrangement de cette base de données : après le littérateur.

Finalement nous avons trouvé cette base de données qui possède les caractéristiques et les propriétés suivant:

$$ODT \propto \frac{K_{AE}}{K_{OE}} = \frac{P_V}{K_{OE} C_{LE} RT}$$

Tell que :

$$K_{AE} = \frac{P_V}{C_{LE} RT}$$

II-8-3 -Préparation de base des données :

II-8-3-1-Logiciels « Hyperchem 7.5 »

HyperChem rassemble dans une même interface un ensemble d'outils dédiés à la modélisation moléculaire, qui est connu pour sa qualité, flexibilité, et facilité d'usage.

- Fonctionnalités:

HyperChem est le logiciel qui vous permet de faire réellement de la modélisation : il possède plus de méthodes de calculs (mécanique moléculaire, semi-empirique et ab- initio) pour que vous puissiez calculer plus de propriétés.

Hyperchem est utilisé dans cette étude pour construire et optimiser les molécules odoréfirants, chaque molécule est enregistrée comme un fichier nommé "Hin" après l'optimisation. Nous avons utilisés la méthode semi empirique MM⁺ pour l'optimisation. On á

134 molécules donc on obtient 134 fichiers Hin, en suite en va calculer les descripteurs moléculaires à partir de ce fichier par logiciel DRAGON pour chaque molécule [25].

II-8-3-2-Récupération et stabilisation les molécules de fichier Hin

II-8-3-2-1 Stabilisation structure des molécules (minimisation de l'énergie)

Pour stabilise forme de structure de chaque molécule ou minimisation de l'énergie on utilise le HyperChem, pouvez effectuer une minimisation de l'énergie (ou de la géométrie d'optimisation) d'une molécule en utilisant une variété de méthodes de calcul. Les deux mécanismes moléculaires et les méthodes semi-empiriques sont disponibles. Minimisation de l'énergie modifie la géométrie ou la forme d'une molécule d'abaisser l'énergie potentielle de la molécule et pour donner une conformation plus stable.

II-8-3-2-2 Mécanique moléculaire

Champs de force mécanique moléculaire utiliser les équations de la mécanique classique pour décrire les surfaces d'énergie potentielle et des propriétés physiques des molécules. Une molécule est décrite comme une collection d'atomes qui interagissent les uns avec les autres par de simples fonctions analytiques. Cette description est appelée un champ de force. Une composante d'un champ de force est l'énergie résultant de la compression et l'étirement d'un cautionnement [26].

HyperChem comprend quatre haut-mécanique moléculaire des champs de force: de nouvelles implémentations de techniques élaborées et publiées par des groupes de recherche respectés Mais nous dans ce travail on concerte à méthode MM⁺ de caractéristiques Field Force présente dans figure suivante :

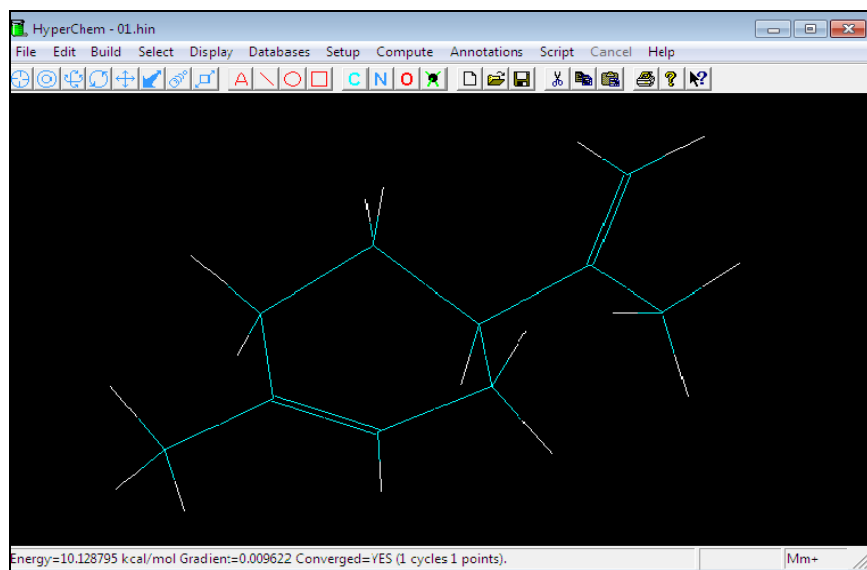


Figure II-7 : applique le Field Force à Composé : limonène ; Formule brute C₁₀H₁₆ ; CAS 7721-11-1

II-8-3-2-3 Récupération des fichiers HyperChem HIN :

Une fois que vous avez construit une structure en HyperChem, vous pouvez l'enregistrer pour une utilisation ultérieure. C'est une bonne idée puisque cela vous fait gagner du temps si vous voulez revoir votre structure à une date ultérieure. Pourquoi construire deux fois?! Vous pouvez le faire en allant dans Fichier et Enregistrer. Vous devez lui donner un nom comme Hin.

Le fichier peut être rappelé pour la visualisation et la manipulation à une date ultérieure. Enregistrez-le dans le dossier public dans le répertoire approprié.

Dans le calcul de descripteurs moléculaires, les structures chimiques des composés optimisés sont nécessaires. Les structures chimiques de tous les 134 composés dans notre jeu de données, ont été établies dans le logiciel HyperChem, et pré-optimisés en utilisant le champ avant MM⁺ mécanique

II-9- Calcule des descripteurs moléculaires :

Afin d'exploiter au maximum les informations contenues dans les structures moléculaires, celles-ci sont traduites en une série de grandeurs (en général scalaires) qui quantifient leurs caractéristiques physico-chimiques et structurelles.

Dans la prochaine étape pour tous les 134 composés, descripteurs moléculaires ont été calculés par le logiciel Dragon. Dragon peut calculer 1664 descripteurs moléculaires pour chaque structure dans notre jeu des données.

II-9-1 Le Logiciel « DRAGON »

DRAGON: c'est une application pour le calcul des descripteurs moléculaires. Ces descripteurs peuvent être utilisés pour évaluer l'influence de la structure moléculaire ou les relations propriétés-structure, aussi pour l'analyse de symétrie et la projection des bases de données des molécules [27].

II-9-2-Groupe des descripteurs moléculaires

En effet, les descripteurs moléculaires sont basés sur plusieurs théories différentes, tel que quantum chimie, théorie de l'information, chimie organique, théorie du graphique, et ainsi de suite, et est utilisé pour modéliser des plusieurs propriétés différentes de produits chimiques dans les champs du scientifique tel que toxicologie, chimie analytique, chimie physique,...etc [28].

Actuellement, il est possible de calculer 1664 descripteurs moléculaires ont inscrit dans Tableau suivant qui peuvent être classés en 20 groupes (blocs) logiques :

Tableau II-4 : les blocks des descripteurs calculés par logiciel DRAGON [28]

Descripteurs moléculaire	No	Descripteurs moléculaire	No
Descripteurs constitutionnelle	48	Autocorrélations 2D	96
Profils moléculaires Randic	41	Descripteurs ESCAPEDE	197
Descripteurs topologiques	119	Indices contiguïté Affilez	107
Descripteurs géométriques	74	Groupe fonctionnel chiffre	121
Marche et comptes de chem	47	Charge valeurs propres	64
Descripteurs RDF	150	Fragments d'atome centré	120
Connectivité indices	33	Indices de charge topologiqu	21
Descripteurs 3D-MORSE	160	Descripteurs de charge	14
Information indice	47	Indices valeurs propres à base	44
Descripteurs WHIM	99	Les propriétés moléculaire	28

En conséquence, à partir du 1664 calculée descripteurs moléculaires par le DRAGON. Les blocs des descripteurs moléculaires choisis et les résultats de calcul des descripteurs sont résumés dans le tableau suivant :

Tableau II-5 : Les descripteurs moléculaires après la réduction

Type descripteurs moléculaires		No après la réduction
2D	Descripteurs topologiques	44
	2D auto corrélation	30
3D	3D-MoRSE	86

II-10- Logiciel « XLSTAT»:

On choisit ce logiciel lorsqu'il contient plusieurs caractéristiques :

- XLSTAT c'est un logiciel très fiable et très stable.
- nous permet de réorganiser nos données et de les structurer afin de faciliter leur analyse.
- Avec XLSTAT l'analyse en composantes principales (ACP) et l'analyse des correspondances (AFC) deviennent faciles, de même que la régression linéaire (simple ou multiple),
- nous donne accès à un grand nombre de tests. Les résultats des tests comprennent une interprétation explicite.
- Il offre une large variété de tests statistiques pour que nous permette de tester des hypothèses.
- nous permet d'évaluer la qualité de nos données, et détermine les erreurs de calcul (écarts + résidus), ce que nous aide d'améliorer notre travail.
- Les résultats des essais incluent une première conclusion de niveau pour soulager notre interprétation.
- Il inclut quelques outils qui facilitent la visualisation des données.

II-10-1-Applications:

Marketing quantitatif, Analyse d'enquêtes, Analyse sensorielle, Analyse de portefeuille, Modélisation de risques, Analyse de données de survie, Recherche géologique, Ecologie, Expérimentations, Psychométrie, Chémométrie, Contrôle de qualité, Enseignement [29].

II-11-Techniques des régressions :

Depuis 1964, l'année de l'approche présente QSPR est née grâce à la contribution de groupe, et dans les dernières années est utilisé les descripteurs moléculaires. La plupart des modèles ont été construits QSPR en utilisant une technique de régression multi-variée.

L'analyse de régression établit une corrélation entre une variable dépendante représentant la propriété chimique et plusieurs variables indépendantes, qui sont, les descripteurs (prédicteurs). Cette corrélation est plus souvent exprimée sous la forme d'une régression linéaire ou non linéaire [30].

II-11-1-La régression Linéaire Simple (RLS)

La Régression Linéaire Simple (RLS) est un modèle particulier de régression dans lequel :

- Il n'y a qu'une seule variable explicative (numérique),
- Le modèle est linéaire dans la variable dans les paramètres.

Comme toutes les techniques prédictives, elle a deux objectifs :

- Construire un modèle dont les paramètres soient interprétables par le praticien en termes de propriétés de la population dont est extrait l'échantillon. On espère bien entendu que les paramètres du modèle seront de bons estimateurs des paramètres de la population sous-jacente.
- Utiliser le modèle pour faire des prédictions

La régression linéaire est sans aucun doute la méthode statistique la plus utilisée. On distingue habituellement la régression simple (une seule variable explicative) de la régression multiple (plusieurs variables explicatives) bien que le cadre conceptuel et les méthodes de calculs soient identiques [31].

II-11-2-Coefficients d'ajustement

Dans le tableau qui affichées les statistiques relatives à l'ajustement du modèle de régression on:

- R^2 : le coefficient de détermination du modèle. Ce coefficient, dont la valeur est comprise entre 0 et 1, n'est affiché que si la constante du modèle n'est pas fixée par l'utilisateur. Sa valeur est définie par:

$$R^2 = \frac{\sum_{i=1}^n w_i (y_i - \hat{y}_i)^2}{\sum_{i=1}^n w_i (y_i - \bar{y})^2} \quad (\text{II.3})$$

Avec :

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n w_i y_i \quad (\text{II.4})$$

Le R^2 s'interprète comme la proportion de la variabilité de la variable dépendante expliquée par le modèle. Plus le R^2 est proche de 1, meilleur est le modèle.

L'inconvénient du R^2 est qu'il ne prend pas en compte le nombre de variables utilisées pour ajuster le modèle.

- R^2 ajusté : le coefficient de détermination ajusté du modèle. Le R^2 ajusté peut être négatif si le R^2 est voisin de zéro. Ce coefficient n'est affiché que si la constante du modèle n'est pas fixée par l'utilisateur. Sa valeur est définie par :

$$\hat{R}^2 = 1 - (1 - R^2) \frac{W - 1}{W - p - 1} \quad (\text{II.5})$$

Le R^2 ajusté est une correction du R^2 qui permet de prendre en compte le nombre de variables utilisées dans le modèle.

- RMCE : la racine de la moyenne des carrés des erreurs (RMCE) est la racine carrée de la MCE.
- MAPE : l'erreur moyenne de pourcentage absolu (Mean Absolute Percentage Error) est calculée comme suit :

$$MAPE = \frac{100}{w} \sum_{i=1}^n w_i \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (\text{II.6})$$

II-11-3- Description de régression linéaire simple

La Régression Linéaire Simple traite la question suivante :

- Une grandeur y est mesurée
- Pour un certain nombre de valeurs d'une autre grandeur x .

Ces mesures se traduisent par le diagramme de dispersion suivant :

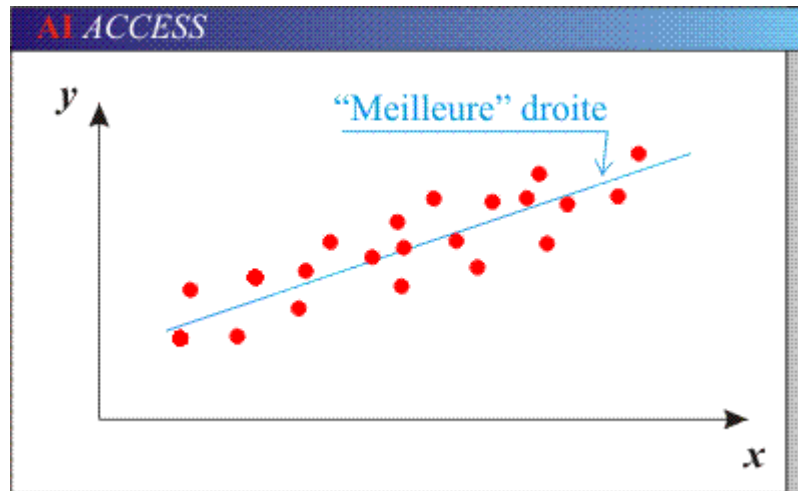


Figure II-8: Montre la meilleure droite de régression linéaire Simple.

Le Modèle Linéaire Simple (MLS) précise les conditions suivantes :

Il existe une droite d'équation : $y = b + a.x$ (II.7)

Telle que, pour chaque valeur x_i de x , la valeur y_i de Y_i est obtenue de la façon suivante :

- La hauteur de cette droite (donc y) est la valeur de la mesure si celle-ci n'était pas entachée d'erreur.
- Mais à cette valeur "parfaite" vient s'ajouter une **erreur de mesure**, réalisation d'une valeur x_i de moyenne nulle et de variance.

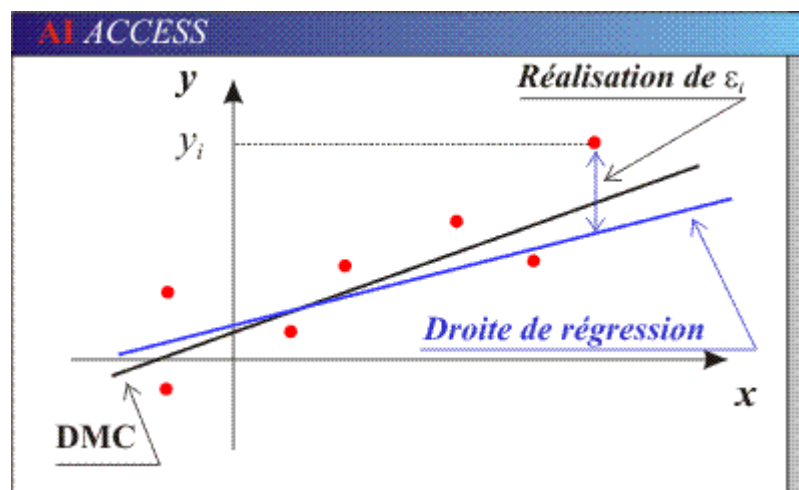


Figure II-9: Montre la réalisation de valeur $x_{i i}$ et la droite de régression

Cette droite (inconnue) s'appelle la **droite de régression** du mécanisme qui a engendré les données. Dans la littérature, la Droite des Moindres Carrés est souvent, appelée Droite de Régression. Nous réservons cette appellation à l'authentique droite de régression. Il est par contre justifié de baptiser la DMC "droite de régression empirique".

Les valeurs. Y_i sont donc définies par :

$$Y_i = \alpha + \beta \cdot x + \varepsilon_i \quad (\text{II.8})$$

Les ε_i étant de moyenne nulle, la moyenne de Y_i (pour un i donné) est égale à $\alpha + \beta \cdot x$.

Les valeurs, Y_i ont toutes leur moyenne sur la droite de régression (ce qui est d'ailleurs la définition de la régression).

Le MLS est donc défini par trois paramètres :

- β , la pente de la droite de régression,
- α , l'ordonnée à l'origine de la droite de régression,
- σ^2 , la variance des erreurs

L'objectif central de la RLS est de trouver de bons estimateurs de ces quantités ainsi que d'identifier les liens de dépendance (covariance) entre ces estimateurs [9].

II-11-4- Les étapes de régression

Nous avons fait les calculs avec XLSTAT software et les résultats pour ces descripteurs sont rapportés avec référence aux valeurs expérimentales. Avec la probabilité d'entré est :

$\alpha = 0.05$, on choisit des paramètres de validation (jeu aléatoire et 45% de la base de données).

Ouvrir une feuille d'Excel et lancer le logiciel XLSTAT par click sur l'icône de dernier, d'après la barre de menu choisir XLSTAT. Une liste va apparaitre, ce dernier contient les différentes taches de logiciel.

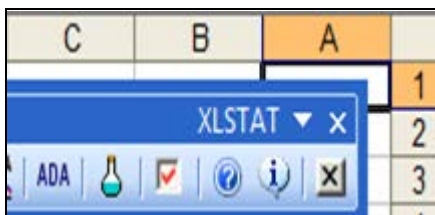


Figure II-10 : Schéma représente logiciel XLSTAT.

Partir la liste précédant, choisir " modélisation des données" ; puis " Régression linéaire " si on faire une régression linéaire ou non linéaire.

II-11-4-1 modèle linéaire

La régression linéaire est la méthode statistique la plus utilisé. On distingue deux types la régression simple et régression multiple bien que la cadre conceptuel et la méthode des calculs soient identiques. Le principe de régression linéaire est de modéliser une variable dépendante quantitative Y, à travers d'une combinaison linéaire de (p) variables explicatives quantitatives, X1, X2... Xp. le modèle déterministe (ne prenant pas en compte d'aléatoire) s'écrit pour une observation i.

La forme de l'équation est donnée qui les calculs de XLSTAT s'écrit comme suit :

$$P(x) = A + \sum_{i=0}^n w_i x_i \quad (\text{II.10})$$

Ou P : quel que soit propriété (Pv, log ODT, log K_{OW}) ;

A : constante ;

W : les poids (valeurs des coefficients) de les variables x, i=0, 1, 2,...n ;

x : les variables de la propriété (descripteurs moléculaires), i=0, 1, 2,...n.

Dans ce travaille on utilisé la régression linéaire pour estimer le coefficient de partage K_{OW}, la pression de vapeur et la seuil de détection d'odeur (log ODT) cette régression est faire par le choix de l'ongle "régression linéaire". Une fois cette dernière est sélectionnée, une boite de dialogue va apparaitre.

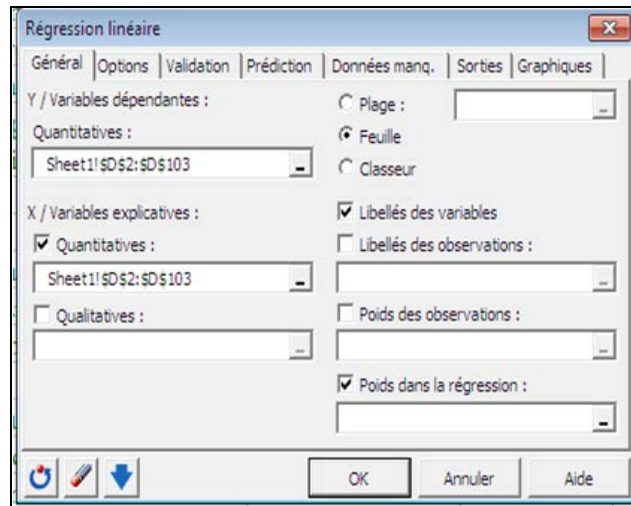


Figure II-11 : Boite dialogue de régression linéaire.

De la boîte précédente, sélectionner "variables(s) dépendante(s)" pour sélectionner les valeurs expérimentales des propriétés à modéliser (les Y_i), puis choisir "Variables quantitatives" pour sélectionner les variables indépendantes (les X_{ij}). Après cette étape, on passe aux autres onglets de boîte pour sélectionner les autres paramètres Puis on click sur le bouton OK. Une fois que vous avez cliqué sur le bouton " OK", les calculs commencent puis les résultats sont affichés.

II-11-5-Tableaux Résultants :

Le premier tableau de résultats fournit les coefficients d'ajustement du modèle. Le R^2 (coefficient de détermination) donne une idée du pourcentage de variabilité de la variable à modéliser, expliqué par les variables. Si ce coefficient proche de 1, la régression au modèle linéaire est mieux.

Tableau II-6: Coefficient d'ajustement

89.000	Observations
89.000	Somme des poids
15.000	DDL
0.973	R^2
0.844	R^2 ajusté
19461917.234	MCE
4411.566	RMCE
132919.923	MAPE
2.106	DW
74.000	Cp
1483.301	AIC
1667.460	SBC
0.290	PC
118187.914	Press RMCE

Le deuxième tableau fournit les détails sur le modèle, il est essentiel lorsque le modèle doit être pour faire des prévisions, simulations ou s'il doit être comparé à d'autres résultats.

Tableau II-7: Analyse de la variance

Pr > F	F	Moyenne des carrés	Somme des carrés	DDL	Source
< 0.0001	7.506	146089236.154	10664514239.231	73	Modèle
		19461917.234	291928758.506	15	Erreur
			10956442997.737	88	Total corrigé

II-11-6- Graphes résultants:

Le premier graphe permet de visualiser les données, la droite de régression et les deux intervalles de confiance (le plus proche de la courbe est l'intervalle autour de la moyenne de l'estimateur, le deuxième est l'intervalle autour de l'estimation ponctuelle aussi appelé intervalle de prédiction). Le troisième graphe indique que les résidus croissent en fonction du poids.

III-12- Conclusion:

Avant la pénétration dans la réalisation de notre travail, il est nécessaire de préparer les outils d'exécution qui sont:

- Présenté les étapes de modélisation et les techniques qui sont utilisés dans notre travail.
- l'arrangement de la base de données, qui contient 134 composés odorants de différentes familles chimiques.
- Les descripteurs moléculaires qui sont distribués sur 3 blocs et calculés à l'aide de Dragon. Après la réduction, le nombre de ces descripteurs deviens 160.
- Le logiciel XLSTAT pour la modélisation (régression étape par étape +validation).

CHAPITRE III

Résultats et discussion

III-1- Introduction :

Dans les chapitres précédents, on a parlé sur la conception des méthodes prédictives et leur impotences dans le calcul de propriétés physico-chimiques, on a détaillé la méthodologie de développement de la méthode « **QSPR** » par l'utilisation des outilles comme la régression.

Dans ce chapitre on va présenter quelques résultats de régression des données expérimentales obtenu par le logiciel XLSTAT. Effectués à chaque block des descripteurs utilisés dans ce travail. Les résultats sont présentés en deux formes:

- Résultats sous forme des tableaux.
- Résultats sous forme des figures.

III-2-Résultats de régression :

III-2-1-Coefficient de corrélations

Les coefficients de corrélations des blocks descripteurs, obtenus par la régression des propriétés (P_v , $\log K_{ow}$, $\log ODT$), donnent par le tableau suivant:

Tableau III-1 : Coefficient d'ajustement (corrélation) des propriétés étudiées.

Coefficient d'ajustement	Pv			log Kow			Log ODT		
	Block 2	Block 6	Block 14	Block 2	Block 6	Block 14	Block 2	Block 6	Block 14
R ²	0.187	0.488	0.973	0.529	0.498	0.919	0.578	0.003	0.854

- Le modèle quant utilisé à coures de l'étude :

$$P = b + \sum (a_i \cdot x_i) \dots \dots \dots III - 1$$

Tell que :

- P : le propriété (P_v , log ODT, log K_{ow}).
- b= constante.
- a_i = valeur de coefficient.
- x_i = variable (paramètre ou descripteur).

III-3- Block de descripteurs 02: Descripteurs topologiques :

III-3-1-Etude de propriété de log K_{ow}

Tableau III-2 : Coefficient d'ajustement

89.000	Observations
89.000	Somme des poids
45.000	DDL
0.529	R ²
0.079	R ² ajusté

Tableau III-3 : Analyse de variance.

Pr > F	F de Fisher	Carré moyen	mme des carr	ddl	Source
0.096	1.498	2.263	45.257	20	Modèle
		1.511	170.724	113	Résidus
			215.982	133	Total

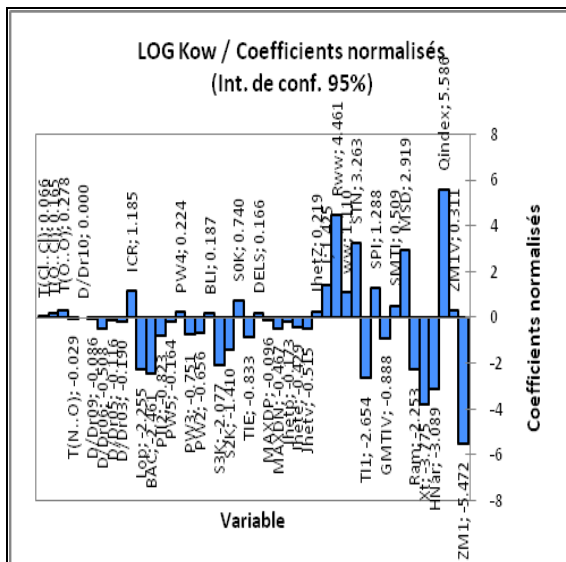


Figure III-1 : représentation de log K_{ow} en fonction des coefficients normalisés.

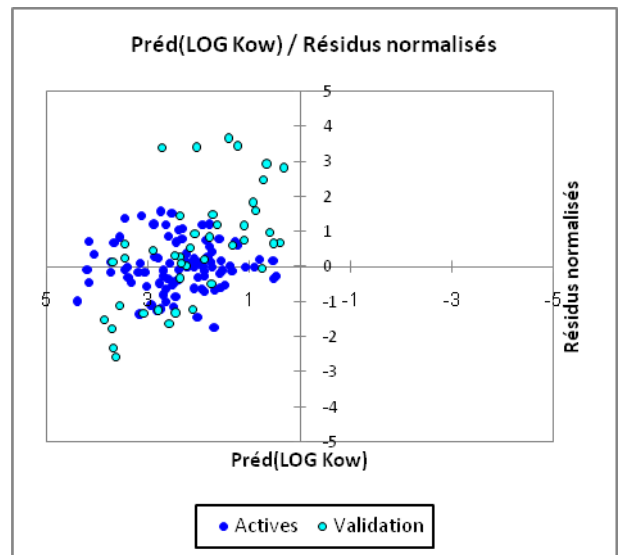


Figure III-2 : représentation de préd(log K_{ow}) en fonction de résidus normalisés.

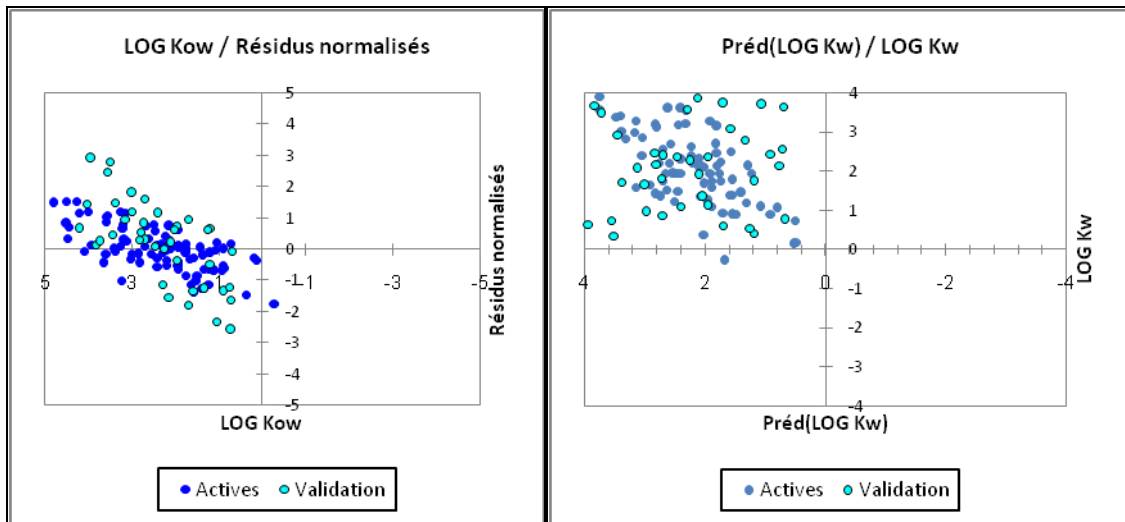


Figure III-3 : représentation de $\log K_{ow}$ en fonction du résidu normalisé. Figure III-4 : représentation de $\text{préd}(\log K_{ow})$ en fonction des $\log K_{ow}$.

III-3-1-1-Analyse des résultats

$$R^2=0.529 \text{ et } R^2 \text{ ajusté}=0.079$$

Donc pourcentage **52,9%** de variation de la variable à modélisé. Ce la nous indiquer que **52,9%** de la variabilité du **log K_{ow}** est expliquée par les descripteurs. Le reste de la variabilité est dû à des effets (autres variables explicatives) qui ne sont pas pris en compte ou bien les descripteurs sélectionnés ne sont pas totalement descriptives.

III-3-1-2-Analyse des résidus

$$\text{d'après le tableau: Le résidu} = 1 - R^2 = 1 - 0.529 = 0.471 = 100 * 0.471 = 47.1\%$$

Cela signifie, entre autres, que **52,9 %** des résidus doivent se trouver dans l'intervalle, toute valeur en dehors de cet intervalle est révélatrice d'une donnée suspecte. Afin de mettre en évidence les valeurs se trouvant hors de l'intervalle. Les paramètres permettent d'expliquer **52,9%** de la variabilité du **log K_{ow}** . Pour expliquer les variabilités restantes, d'autres sources de variabilité doivent donc être prises en compte.

Tableau III-4 : Paramètre de modèle

me supérieure 9	me inférieure 9	Pr > t	t de Student	Ecart-type	Valeur	Paramètre
28.327	3.775	0.011	2.590	6.196	16.051	Constante
-0.101	-0.834	0.013	-2.527	0.185	-0.467	ZM1
0.029	-0.032	0.906	-0.118	0.015	-0.002	ZM1V
1.025	-0.344	0.327	0.985	0.345	0.340	Qindex
-0.123	-18.843	0.047	-2.007	4.724	-9.483	HNar
-1.049	-18.562	0.029	-2.218	4.420	-9.805	Xt
1.017	-2.369	0.431	-0.791	0.855	-0.676	Ram
22.385	-0.480	0.060	1.898	5.771	10.952	MSD
0.010	-0.003	0.276	1.095	0.003	0.004	SMTI
0.001	-0.002	0.496	-0.683	0.001	-0.001	GMTIV
1.255	0.055	0.033	2.164	0.303	0.655	SPI
0.146	-0.157	0.942	-0.073	0.076	-0.006	TI1
13.303	0.328	0.040	2.081	3.275	6.815	STN
0.001	-0.002	0.556	-0.591	0.001	0.000	ww
1.589	-0.242	0.148	1.457	0.462	0.673	Rww
4.829	-1.595	0.321	0.997	1.621	1.617	J
0.420	-1.712	0.233	-1.200	0.538	-0.646	JhetZ
0.759	-13.815	0.079	-1.775	3.678	-6.528	Jhetv
3.556	-0.009	0.051	1.971	0.900	1.774	Jhete
11.629	-1.499	0.129	1.529	3.313	5.065	Jhetp
0.441	-0.981	0.454	-0.751	0.359	-0.270	MAXDN
9.822	-81.451	0.114	-1.683	21.278	-35.814	Mor16m
49.016	-80.638	0.609	-0.523	30.225	-15.811	Mor17m
60.982	-64.107	0.958	-0.054	29.161	-1.562	Mor18m
36.523	-24.967	0.693	0.403	14.335	5.778	Mor19m
9.859	-57.568	0.151	-1.518	15.719	-23.854	Mor20m
83.696	-93.629	0.906	-0.120	41.339	-4.967	Mor21m
66.182	-29.603	0.426	0.819	22.330	18.289	Mor22m

III-1-1-3-Discussion

L'étude de corrélation entre les descripteurs deux à deux ont conduit à classer les descripteurs entre deux extrêmes : Les descripteurs moins corrélés (T(O..O) , HNar) =0,08, (JhetZ) ,T(N..O)) =-0,011, (D/Dr09,T(O.. Cl)= - 0,015.Qui signifie qu' ils ne donnent pas la même information sur la propriété à estimer (**log K_{ow}**) (chaque descripteur à une information différente a celle de l'autre).C.à.d que l'information sur la distance topologique entre (N...O) est différente à celle de l'information sur (JhetZ) qui représente l'indice de type Balban de matrice du poids de distance Z, et que l'information donnée par D/Dr09 qui représente (Distance/l'indice de détour d'anneau d'ordre 9), et les informations données par les distances topologique T(O..O) est différente a celle de (HNar) qui représente l'indice topologique harmonique de Narumi .Ces descripteurs peuvent être importants pour le modèle.

Des descripteurs possédants une grande corrélation entre eux $(X_t, PW5) = -0.824$, $(X_t, PW3) = -0.841$, $(ICR, R_{ww}) = 0,756$. Cela signifie qu'ils vont nous donner des informations plus proches sur la propriété à estimer ($\log K_{ow}$). C.à.d que l'information sur l'indice de connectivité de la structure totale (X_t) est proche à celle de chemin/allure 3,5 ($PW3$ et $PW5$), et l'indice de d'information centrée radiale (ICR), est proche de l'indice réciproque de hyper-détour (R_{ww}). Ces descripteurs peuvent être éliminés du modèle.

Le calcul de la matrice de corrélation nous a permis d'éliminer **44** autres descripteurs puisqu'ils nous ramènent à la même information que les 45 restants $R > 0,95$. Les descripteurs ($T(O..O)$, X_t , R_{ww}) sont respectivement les plus corrélés avec $\log K_{ow}$. Ces variables nous donnent grandes informations sur $\log K_{ow}$, et sont parmi les paramètres du modèle. Les descripteurs ($GMTIV$, ww) sont les moins corrélés avec $\log K_{ow}$, ils nous donnent peut d'informations sur $\log K_{ow}$, et par conséquent ne seront pas présents sur l'équation du modèle.

L'interprétation du graphe ($\text{périd}(\log k_{ow}/\log k_{ow})$), nous montre qu'il y a **134** observations (molécules de la base de données), et sont à l'intérieur de l'intervalle de confiance, ce qui nous permet de dire que les structures bidirectionnelles de ces molécules sont régulières. Par contre les **44** observations qui ne sont pas corrélées des observations (molécules) restants de la base de données, ces molécules possèdent des structures chimiques bidimensionnelles irrégulières. On aussi à partir des paramètres de modèle que $T(I..I)$ et $T(Br..Br)$ et $T(O.Br)$ et $T(O.I)$ à sont nulle donc il reste 39 Observations qui présentent dans (**Annex 01**).

III-3-2-Etude de propriété de P_v

Tableau III-5 : Coefficient d'ajustement

89.000	Observations
89.000	Somme des poids
77.000	DDL
0.187	R^2
0.071	R^2 ajusté

Tableau III-6 : Analyse de variance.

Pr > F	F de Fisher	Carré moyen	Somme des carrés	ddl	Source
0.234	1.198	176990886.829	7787599020.492	44	Modèle
		147750834.154	13149824239.739	89	Résidus
			20937423260.231	133	Total

Tableau III-7 : Paramètre de modèle.

Borne supérieure 95 %	Borne inférieure 95 %	Pr > t	t de Student	Ecart-type	Valeur	Paramètre
742074.402	-367433.334	0.504	0.671	279194.673	187320.534	Constante
10278.393	-6759.076	0.682	0.410	4287.280	1759.659	ZM1
267.637	-923.855	0.277	-1.094	299.825	-328.109	ZM1V
22019.269	-35967.036	0.634	-0.478	14591.577	-6973.884	Qindex
254599.287	-257125.520	0.992	-0.010	128769.575	-1263.116	HNar
162.394	-177.282	0.931	-0.087	85.475	-7.444	SMTI
38.386	-27.113	0.733	0.342	16.482	5.637	GMTIV
7366.426	-5422.587	0.763	0.302	3218.206	971.920	DELS
1686.262	-986.253	0.604	0.520	672.507	350.004	S0K
8338.560	-1354.503	0.156	1.432	2439.146	3492.029	BAC
2725.053	-1490.829	0.562	0.582	1060.877	617.112	D/Dr06
1870.863	-4525.770	0.412	-0.825	1609.638	-1327.454	T(O..O)
2888.862	-3758.360	0.796	-0.260	1672.696	-434.749	T(O..CI)
1865.300	-4875.709	0.377	-0.887	1696.296	-1505.205	T(CI..CI)

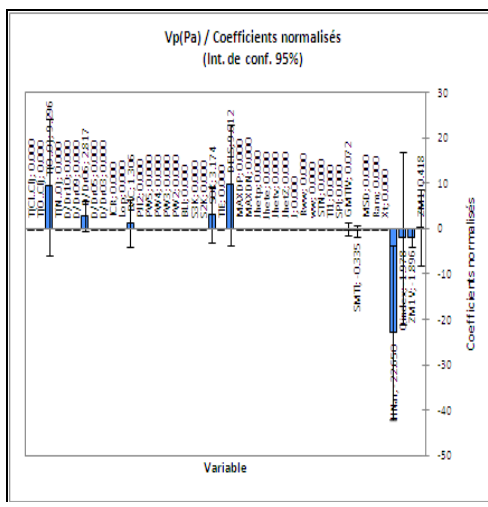


Figure III-5 : Variation P_v en fonction des coefficients normalisés.

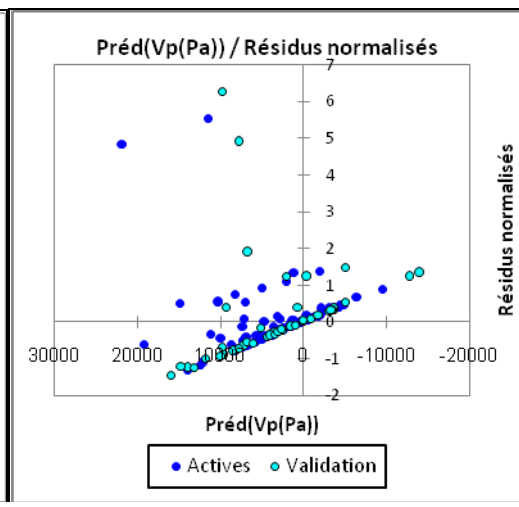


Figure III-6 : Variation $\text{pré}(P_v)$ en résidus normalisés.

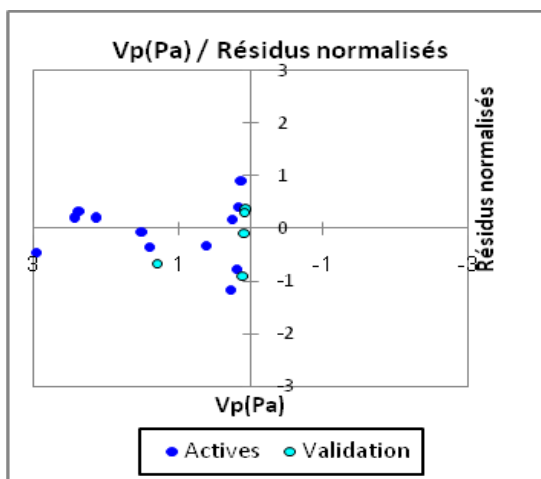


Figure III-7 : Variation P_v en résidus normalisés.

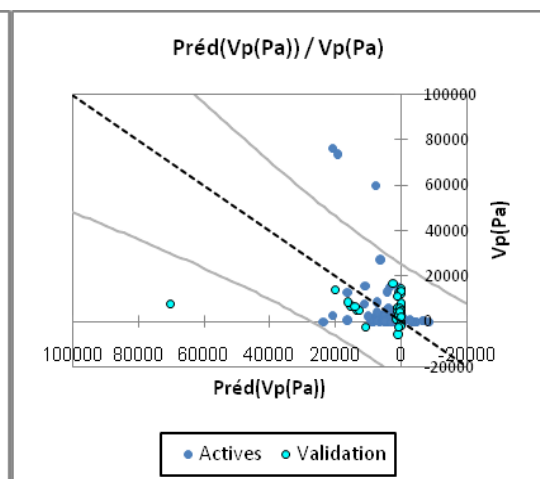


Figure III-8 : Variation $\text{pré}(P_v)$ en P_v .

III-3-2-1-Analyse des résultats

$$R^2=0.187 \text{ et } R^2 \text{ ajusté}=0.071$$

Donc pourcentage **18,7%** de variation de la variable à modéliser. Ce la nous indiquer que **18,7%** de la variabilité du P_v est expliquée par les descripteurs. Le reste de la variabilité est dû à des effets (autres variables explicatives) qui ne sont pas pris en compte. On a aussi que la précision est très faible pour cela donc le paramètre de modèle apparaitre uniquement 7 descripteur est les autre son éliminer care elle attient de propriétés Multi colinéarité

III-3-2-2-Analyse des résidus

$$\text{d'après le tableau : Le résidu} = 1 - R^2 = 1 - 0.187 = 0.813 = 100 * 0.813 = 81.3\%$$

Cela signifie que **18.7 %** des résidus doivent se trouver dans l'intervalle et les paramètres permettent d'expliquer **18,7%** de la variabilité du P_v . Pour expliquer les variabilités restantes, d'autres sources de variabilité doivent donc être prises en compte.

III-3-2-3-Discussion

Les descripteurs possédants une grande corrélation entre eux ($Lop, Qindex$) = 0.999, (Xt, Rww) = 0.999, ($T11, Hnar$) = 0.997,. Cela signifie qu'ils vont nous donner des informations plus proches sur la propriété à estimer (P_v). C.a.d que l'information sur Indice central d'élagage (Lop) est proche a celle de l'indice quadratique ($Qindex$), et l'indice de structure totale (Xt) proche de l'indice réciproque de hyper-détour (Rww) et l'indice topologique harmonique de Narumi plus proche de l'indice première de Mohar. Ces descripteurs peuvent être éliminés du modèle.

Le calcul de la matrice de corrélation nous a permis d'éliminer **7** autres descripteurs puisqu'ils nous ramènent à la même information que les 37 restants $R > 0,95$. Les descripteurs ($T(O..O)$, $ZM1$, $S0K$) sont respectivement les plus corrélés avec P_v . Ces variables nous donnent grandes informations sur P_v et sont parmi les paramètre du modèle. Les descripteurs ($GMTIV$, $SMTI$) sont les moins corrélés avec P_v , ils nous donnent peut d'informations sur P_v , et par conséquent ne seront pas présents sur l'équation du modèle .

Tableau III-10: Paramètre de modèle.

Borne supérieure (95%)	Borne inférieure (95%)	Pr > t	t	Ecart-type	Valeur	Source
238.047	-470.642	0.512	-0.661	175.932	-116.297	Constante
2.416	-1.210	0.506	0.670	0.900	0.603	ZM1
0.106	-0.108	0.982	-0.022	0.053	-0.001	ZM1V
2.363	-8.649	0.256	-1.150	2.734	-3.143	Qindex
42.175	-56.465	0.772	-0.292	24.487	-7.145	HNar
125.392	-55.917	0.444	0.772	45.010	34.738	Xt
7.562	-12.072	0.646	-0.463	4.874	-2.255	Ram
325.843	-144.208	0.440	0.778	116.690	90.817	MSD
-0.003	-0.068	0.032	-2.213	0.016	-0.035	SMTI
0.009	-0.002	0.239	1.194	0.003	0.003	GMTIV
2.630	-3.888	0.699	-0.389	1.618	-0.629	SPI
1.520	-2.097	0.749	-0.321	0.898	-0.289	TI1
14.710	-21.973	0.692	-0.399	9.106	-3.631	STN
0.013	-0.003	0.249	1.167	0.004	0.005	ww
4.040	-3.279	0.835	0.209	1.817	0.381	Rww
10.841	-31.401	0.332	-0.980	10.487	-10.280	J
1.730	-3.477	0.502	-0.676	1.293	-0.874	JhetZ
6.907	-36.349	0.177	-1.371	10.738	-14.721	Jhetv
7.298	-3.167	0.431	0.795	2.598	2.066	Jhete
29.467	-9.014	0.290	1.071	9.553	10.227	Jhetp
3.089	-4.029	0.792	-0.266	1.767	-0.470	MAXDN
0.975	-1.224	0.821	-0.228	0.546	-0.125	MAXDP
0.235	-2.014	0.118	-1.593	0.558	-0.889	DELS
1.522	0.176	0.015	2.541	0.334	0.849	TIE
0.406	-0.027	0.085	1.764	0.107	0.189	SOK
2.461	-1.776	0.746	0.325	1.052	0.342	S2K
0.468	-1.977	0.220	-1.243	0.607	-0.755	S3K
5.706	-13.593	0.415	-0.823	4.791	-3.944	BLI
452.862	-85.409	0.176	1.375	133.626	183.727	PW2
108.116	-36.246	0.321	1.003	35.838	35.935	PW3
223.056	-33.228	0.143	1.492	63.622	94.914	PW4
100.785	-29.512	0.276	1.102	32.346	35.637	PW5
9.183	-2.077	0.210	1.271	2.795	3.553	PJ12
0.946	-0.778	0.846	0.196	0.428	0.084	BAC
4.837	-10.421	0.465	-0.737	3.788	-2.792	Lop
1.500	-12.862	0.118	-1.593	3.565	-5.681	ICR
3.446	-2.397	0.720	0.361	1.451	0.524	D/Dr03
1.198	-0.815	0.704	0.383	0.500	0.191	D/Dr05
1.087	-0.638	0.603	0.523	0.428	0.224	D/Dr06
0.837	-0.544	0.671	0.427	0.343	0.146	D/Dr09
0.699	-0.332	0.477	0.717	0.256	0.184	D/Dr10
1.417	-1.192	0.863	0.174	0.648	0.113	T(N..O)
0.487	-0.972	0.506	-0.670	0.362	-0.243	T(O..CI)
2.750	0.473	0.007	2.851	0.565	1.612	T(CI..CI)

III-3-3-1-Analyse des résultats

$$R^2=0.578 \text{ et } R^2 \text{ ajusté}=0.175$$

Donc pourcentage **57,8%** de variation de la variable à modéliser. Ce la nous indiquer que **57,8%** de la variabilité du **Log(ODT)** est expliquée par les descripteurs. Le reste de la variabilité est dû à des effets (autres variables explicatives) qui ne sont pas pris en compte ou bien les descripteurs sélectionnés ne sont pas totalement descriptives.

III-3-3-2-Analyse des résidus

$$\text{D'après le tableau: Le résidu} = 1-R^2 = 1-0.578=0.422= 100 *0.422 =42.2\%$$

Cela signifie, entre autres, que **57,8 %** des résidus doivent se trouver dans l'intervalle, toute valeur en dehors de cet intervalle est révélatrice d'une donnée suspecte. Afin de mettre en évidence les valeurs se trouvant hors de l'intervalle .Les paramètres permettent d'expliquer

57,8% de la variabilité du **log ODT**. Pour expliquer les variabilités restantes, d'autres sources de variabilité doivent donc être prises en compte.

III-3-3-3-Discussion

L'étude de corrélation entre les descripteurs deux à deux ont conduit à classer les descripteurs entre deux extrêmes : Les descripteurs moins corrélés ($T(N..O)$, $HNar$) = 0,063, ($JhetZ$, $T(N..O)$) = 0,03, ($D/Dr03$, $T(O.. Cl)$) = - 0,022. Qui signifie qu'ils ne donnent pas la même information sur la propriété à estimer (**log ODT**) (chaque descripteur à une information différente à celle de l'autre). C.à.d que l'information sur la distance topologique entre ($N...O$) est différente à celle de l'information sur ($JhetZ$) qui représente l'indice de type Balban de matrice du poids de distance Z , et que l'information donnée par $D/Dr03$ qui représente (Distance/l'indice de détour d'anneau d'ordre 3), et les informations données par les distances topologique $T(N..O)$ est différente à celle de ($HNar$) qui représente l'indice topologique harmonique de Narumi. Ces descripteurs peuvent être importants pour le modèle.

Des descripteurs possédants une grande corrélation entre eux (Xt , $PW3$) = - 0.814, (Xt , $PW4$) = - 0.849, (ICR , LOP) = 0,756,. Cela signifie qu'ils vont nous donner des informations plus proches sur la propriété à estimer (**log ODT**). C.à.d que l'information sur l'indice de connectivité de la structure totale (Xt) est proche à celle de chemin/allure 3,4 ($PW3$ et $PW4$), et l'indice de d'information centrique radial (ICR), est proche de l'indice central d'élagage (Lop).. Ces descripteurs peuvent être éliminés du modèle.

Le calcul de la matrice de corrélation nous a permis d'éliminer **43** autres descripteurs puisqu'ils nous ramènent à la même information que les 2 restants $R > 0,95$. Les descripteurs ($T(Cl..Cl)$, $PW5$, Xt) sont respectivement les plus corrélés avec **log ODT**. Ces variables nous donnent grandes informations sur **log ODT**, et sont parmi les paramètres du modèle. Les descripteurs (STN , $SMTI$) sont les moins corrélés avec **log ODT**, ils nous donnent peut d'informations sur **log ODT**, et par conséquent ne seront pas présents sur l'équation du modèle.

L'interprétation du graphe (**périd(log ODT/ log ODT)**), nous montre qu'il y a **134** observations (molécules de la base de données), et sont à l'intérieur de l'intervalle de confiance, ce qui nous permet de dire que les structures bidirectionnelles de ces molécules sont régulières. Par contre les **4** observations qui ne sont pas corrélés des observations

(molécules) restants de la base de données, ces molécules possèdent des structures chimiques bidimensionnelles irrégulières qui présenter dans le (Annex 03).

III-4-Block de descripteurs 14: Descripteurs 3D-MorSE:

III-4-1-Etude de propriété de log ODT

Tableau III-11 : Coefficient d'ajustement.

89.000	Observations
89.000	Somme des poids
14.000	DDL
0.854	R ²
0.082	R ² ajusté

Tableau III-12 : Analyse de variance

Pr > F	F	Moyenne des carrés	Somme des carrés	DDL	Source
0.442	1.106	2.651	196.209	74	Modèle
		2.397	33.559	14	Erreur
			229.769	88	Total corrigé

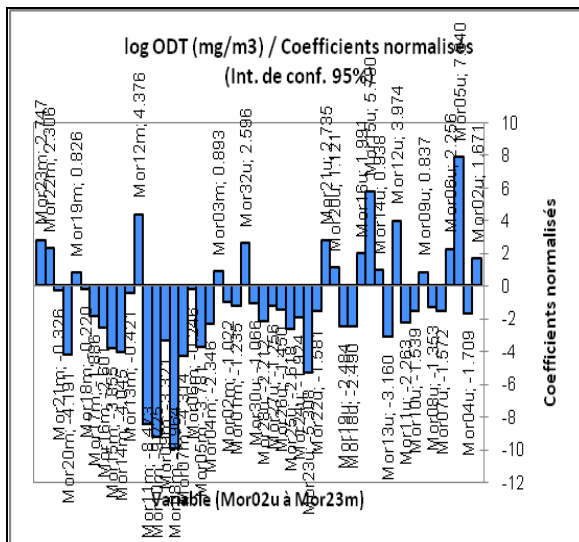


Figure III-13 : log ODT en fonction de Coefficient normalisés.

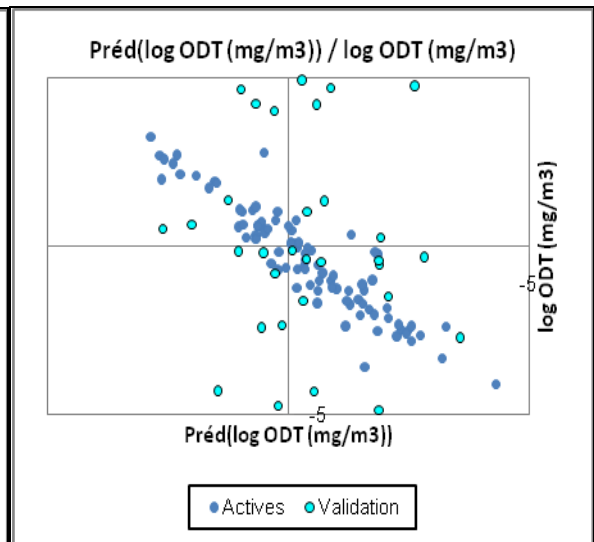


Figure III-14 : préd(log ODT en fonction de log ODT

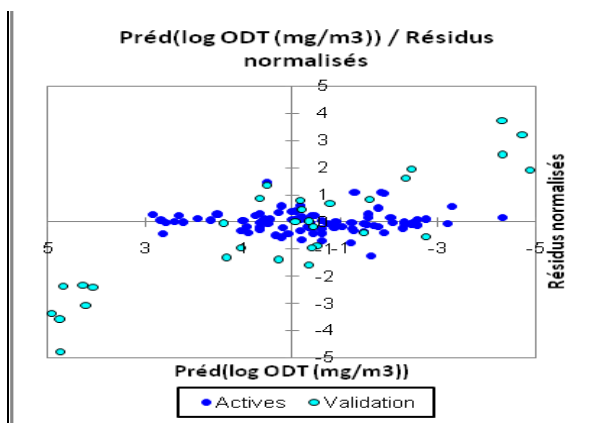


Figure III-15 : log ODT en fonction de résidus normalisés.

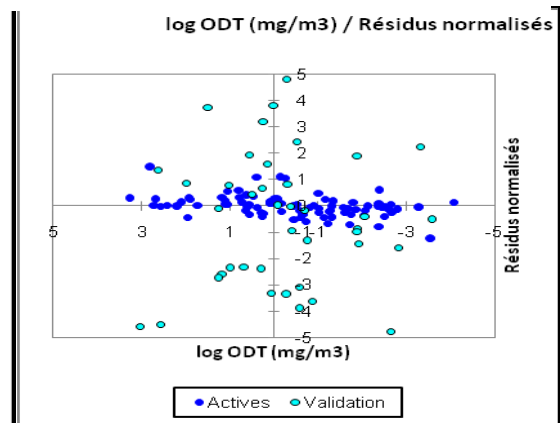


Figure III-16 : préd(log ODT en fonction de résidus normalisés

Tableau III-13 : Paramètre de modèle

Moyenne	Maximum	Minimum	ns données ma	ns données ma	Observations	Variable
-0.559	3.260	-3.569	45	0	45	log ODT (mg/
91.708	206.131	11.610	45	0	45	ISIZ
24.393	44.363	7.610	45	0	45	IAC
1.207	1.764	0.881	45	0	45	AAC
2.232	3.414	0.918	45	0	45	IDE
4.716	6.506	1.500	45	0	45	IDM
2.284	3.507	0.811	45	0	45	IDDE
95.362	310.670	2.755	45	0	45	IDET
1.266	1.727	0.619	45	0	45	IVDE
18.305	49.119	4.174	45	0	45	Uindex
0.709	2.042	0.431	45	0	45	Vindex
1.884	3.916	0.000	45	0	45	Yindex
0.293	0.655	0.196	45	0	45	SIC0
3.055	4.237	0.800	45	0	45	CIC0
2.213	3.174	1.157	45	0	45	IC1
45.564	94.219	7.610	45	0	45	TIC1
0.529	0.802	0.268	45	0	45	SIC1
2.048	3.829	0.594	45	0	45	CIC1
2.897	4.022	1.000	45	0	45	IC2
0.681	0.877	0.431	45	0	45	SIC2
1.364	3.005	0.547	45	0	45	CIC2
-19.002	4.271	-999.000	45	0	45	IC3
46.639	140.955	-999.000	45	0	45	TIC3
-21.456	0.969	-999.000	45	0	45	SIC3
-18.969	4.393	-999.000	45	0	45	IC4
-21.455	1.000	-999.000	45	0	45	SIC4
-21.221	2.656	-999.000	45	0	45	CIC4
-65.898	1.000	-999.000	45	0	45	SIC5
-65.607	3.299	-999.000	45	0	45	CIC5

III-4-1-1-Analyse des résultats

$$R^2=0.854 \text{ et } R^2 \text{ ajusté}=0.082$$

Donc pourcentage **85,4** %de variation de la variable à modéliser. Ce la nous indiquer que **85,4%** de la variabilité du **log ODT** est expliquée par les descripteurs. Le reste de la variabilité est dû à des effets (autres variables explicatives) qui ne sont pas pris en compte ou bien les descripteurs sélectionnés ne sont pas totalement descriptives.

III-4-1-2-Analyse des résidus

D'après le tableau: Le résidu = $1-R^2 = 1-0.854=0.146= 100 *0.146 =14.6\%$

Cela signifie, entre autres, que **85,4 %** des résidus doivent se trouver dans l'intervalle, toute valeur en dehors de cet intervalle est révélatrice d'une donnée suspecte. Afin de mettre en évidence les valeurs se trouvant hors de l'intervalle. Les paramètres permettent d'expliquer **85,4%** de la variabilité du **log ODT**. Pour expliquer les variabilités restantes, d'autres sources de variabilité doivent donc être prises en compte.

III-4-1-3-Discussion

L'étude de corrélation entre les descripteurs deux à deux ont conduit à classer les descripteurs entre deux extrêmes : Les descripteurs moins corrélés (Mor30v ,Mor5m) =0,00054,(Mor08v,Mor15v) =0,008 (Mor2u,Mor24m)= 0,002. Qui signifie qu' ils ne donnent pas la même information sur la propriété à estimer (**log ODT**) (chaque descripteur à une information différente a celle de l'autre).C.à.d que l'information sur 3D-MoRSE - signal 30/pesés par les volumes atomiques de van der Waals est différente à celle de l'information sur (Mor15m) qui représente 3D-MoRSE -signal 15/pesés par masse atomique, et que l'information donnée par Mor08v qui représente 3D-MoRSE - signal 8/pesés par les volumes atomiques de van der Waals, est différente a celle des informations données par 3D-MoRSE - signal 15/pesés par les volumes atomiques de van der Waals, et que l'information donnée par Mor2u qui represente 3D-MoRSE - signal 2/ non pesés est différente a celle des informations données par 3D-MoRSE - signal 2/ pesés de masse atomique. Ces descripteurs peuvent être importants pour le modèle.

Des descripteurs possédants une grande corrélation entre eux (Mor08u,Mor11u)=-0,662, (Mor12m,Mor08u)=0,444,. Cela signifie qu'ils vont nous donner des informations plus proches sur la propriété à estimer (**log ODT**).C.a.d que l'information sur 3D-MoRSE - signal 8/ non pesés est proche a celle de 3D-MoRSE - signal 11/ non pesés, et 3D-MoRSE - signal 12/ pesés de masse atomique est proche 3D-MoRSE - signal 8/ non pesés). Ces descripteurs peuvent être éliminés du modèle.

Le calcul de la matrice de corrélation nous a permis d'éliminer **75** autres descripteurs puisqu'ils nous ramènent à la même information que les 9 restants $R > 0,95$. Les descripteurs (Mor18v,Mor09m,Mor15u) sont respectivement les plus corrélés avec **log ODT**. Ces

variables nous donnent grandes informations sur **log ODT** sont parmi les paramètres du modèle. Les descripteurs (Mor01v,Mor01m) sont les moins corrélés avec **log ODT**, ils nous donnent peut d'informations sur **logODT**, et par conséquent ne seront pas présents sur l'équation du modèle.

III-4-2-Etude de propriété de **log K_{ow}**

Tableau III-14 : Coefficient d'ajustement.

Tableau III-15 : Analyse de variance

89.000 Observations	Pr > F	F	Moyenne des carrés	Somme des carrés	DDL	Source
89.000 Somme des poids	0.297	1.175	1.521	65.402	43	Modèle
13.000 DDL			1.295	58.265	45	Erreur
0.919 R ²				123.667	88	Total corrigé
0.448 R ² ajusté						

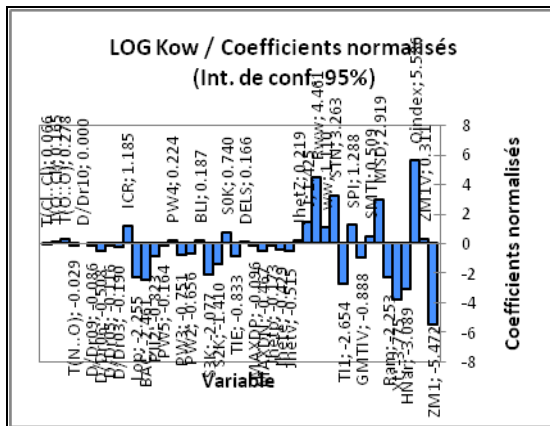


Figure III-17 : log K_{ow} en fonction de coefficients normalisés.

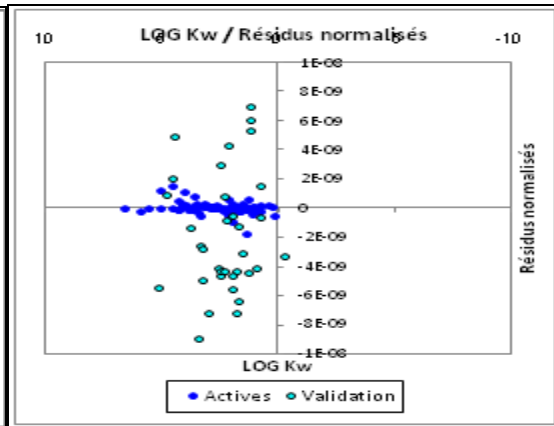


Figure III-18 : log K_{ow} en fonction de résidus normalisés.

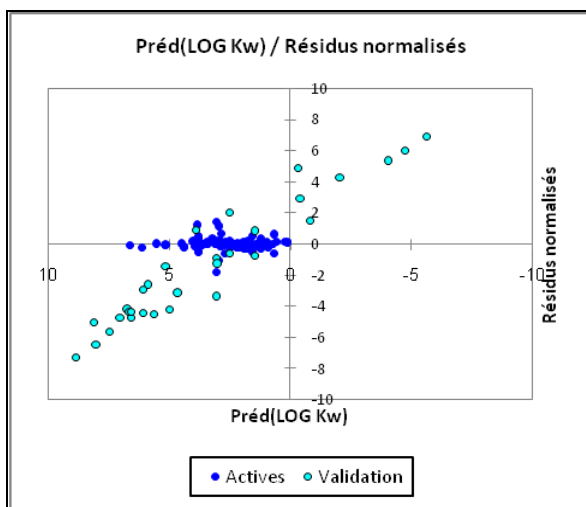


Figure III-19 : préd(log K_{ow}) en fonction des résidus normalisés.

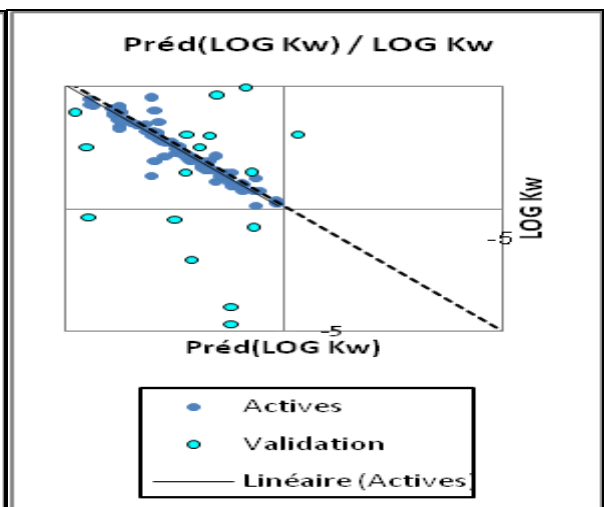


Figure III-20 : prés(log K_{ow}) en fonction de log K_{ow}.

Tableau III-16 : Paramètre de modèle.

e supérieure	Se inférieure	g	Pr > t	t de Student	Ecart-type	Valeur	Paramètre
1.028	-9.656	0.111	-1.624	2.656	-4.314	Constante	
0.892	-0.167	0.175	1.379	0.263	0.363	Mor02u	
1.102	-1.721	0.662	-0.441	0.702	-0.309	Mor04u	
1.974	-5.675	0.335	-0.973	1.901	-1.850	Mor05u	
3.097	-2.008	0.670	0.429	1.269	0.544	Mor06u	
2.433	-3.672	0.685	-0.408	1.517	-0.619	Mor07u	
2.594	-7.822	0.318	-1.010	2.589	-2.614	Mor08u	
6.155	-5.223	0.870	0.165	2.828	0.466	Mor09u	
6.395	-2.581	0.397	0.855	2.231	1.907	Mor10u	
1.384	-7.999	0.163	-1.418	2.332	-3.307	Mor11u	
3.374	-8.101	0.411	-0.829	2.852	-2.364	Mor12u	
10.308	-2.199	0.198	1.304	3.109	4.055	Mor13u	
8.772	-4.682	0.544	0.612	3.344	2.045	Mor14u	
15.863	-2.929	0.173	1.385	4.671	6.467	Mor15u	
2.660	-12.386	0.200	-1.300	3.740	-4.863	Mor16u	
11.826	-6.812	0.770	0.294	5.129	1.507	Mor18u	
0.983	-15.012	0.084	-1.764	3.975	-7.014	Mor19u	
1.315	-15.627	0.096	-1.700	4.211	-7.156	Mor20u	
6.722	-17.705	0.370	-0.905	6.071	-5.491	Mor21u	
0.666	-30.029	0.060	-1.924	7.629	-14.681	Mor22u	
19.692	-6.856	0.336	0.973	6.598	6.418	Mor23u	
28.396	-4.142	0.140	1.500	8.087	12.127	Mor24u	
28.698	-0.912	0.065	1.888	7.359	13.893	Mor25u	
6.869	-13.445	0.518	-0.651	5.049	-3.288	Mor26u	
11.772	-18.177	0.669	-0.430	7.443	-3.203	Mor27u	
14.471	-16.219	0.909	-0.115	7.628	-0.674	Mor28u	
19.436	-14.850	0.789	0.269	8.522	2.293	Mor30u	
11.648	-6.920	0.611	0.512	4.615	2.364	Mor32u	
0.125	-0.264	0.478	-0.716	0.097	-0.069	Mor01m	
3.010	-0.209	0.087	1.750	0.800	1.400	Mor02m	
5.746	-4.951	0.882	0.149	2.659	0.397	Mor03m	
1.778	-13.398	0.130	-1.540	3.772	-5.810	Mor04m	
5.947	-14.917	0.391	-0.865	5.186	-4.485	Mor05m	
11.771	-4.006	0.327	0.990	3.921	3.883	Mor06m	
13.236	-1.122	0.096	1.697	3.667	6.057	Mor07m	
32.609	-1.743	0.077	1.808	8.538	15.433	Mor08m	
35.041	-0.446	0.056	1.961	8.820	17.297	Mor09m	
28.056	-5.482	0.182	1.354	8.336	11.287	Mor10m	
36.134	-0.534	0.057	1.953	9.114	17.800	Mor11m	
38.353	2.755	0.025	2.323	8.648	20.554	Mor12m	
34.943	-2.091	0.081	1.785	9.204	16.426	Mor13m	
12.103	-13.768	0.898	-0.129	6.430	-0.833	Mor14m	
0.203	-30.885	0.053	-1.985	7.727	-15.341	Mor15m	
2.292	-58.289	0.069	-1.860	15.057	-27.999	Mor16m	
4.461	-34.760	0.127	-1.554	9.753	-15.160	Mor17m	
21.539	-17.510	0.836	0.208	9.795	2.015	Mor18m	
23.273	-7.717	0.318	1.010	7.702	7.778	Mor19m	
42.608	-4.311	0.107	1.642	11.661	19.149	Mor20m	
57.990	-7.020	0.121	1.577	16.158	25.485	Mor21m	
41.426	-21.113	0.517	0.653	15.544	10.157	Mor22m	
34.894	-16.256	0.467	0.733	12.713	9.319	Mor23m	
56.964	-17.144	0.285	1.081	18.419	19.910	Mor24m	
72.041	-12.548	0.164	1.415	21.024	29.747	Mor25m	
23.860	-18.145	0.786	0.274	10.440	2.857	Mor26m	
18.391	-48.541	0.369	-0.906	16.635	-15.075	Mor27m	
-10.701	-89.005	0.014	-2.552	19.462	-49.853	Mor28m	
5.629	-49.731	0.116	-1.603	13.759	-22.051	Mor29m	
18.140	-63.690	0.269	-1.119	20.313	-22.725	Mor30m	
5.127	-77.354	0.085	-1.762	20.500	-36.113	Mor31m	
12.099	-50.329	0.224	-1.232	15.516	-19.115	Mor32m	
0.111	-0.180	0.636	-0.477	0.072	-0.034	Mor01v	
-0.499	-4.417	0.015	-2.524	0.974	-2.458	Mor02v	
3.011	-4.152	0.750	-0.320	1.780	-0.670	Mor03v	
4.991	-12.329	0.398	-0.852	4.305	-3.669	Mor04v	
17.462	-14.968	0.878	0.155	8.060	1.247	Mor05v	
11.042	-12.821	0.881	-0.150	5.931	-0.890	Mor06v	
22.046	-24.153	0.927	-0.092	11.482	-1.053	Mor08v	
28.675	-13.497	0.473	0.724	10.481	7.589	Mor09v	
15.097	-21.892	0.713	-0.370	9.193	-3.398	Mor10v	
12.063	-40.116	0.285	-1.082	12.968	-14.027	Mor11v	
2.268	-59.975	0.068	-1.865	15.470	-28.954	Mor12v	
2.897	-55.950	0.076	-1.814	14.626	-26.626	Mor13v	
45.265	-15.877	0.339	0.967	15.196	14.694	Mor15v	
89.546	-7.401	0.095	1.705	24.095	41.072	Mor16v	
16.223	-52.952	0.291	-1.068	17.193	-18.364	Mor18v	
27.664	-31.239	0.903	-0.122	14.640	-1.788	Mor20v	
33.211	-61.407	0.552	-0.600	23.517	-14.098	Mor21v	
70.634	-6.532	0.101	1.671	19.179	32.051	Mor22v	
35.305	-57.746	0.630	-0.485	23.127	-11.221	Mor23v	
32.335	-83.371	0.379	-0.887	28.758	-25.518	Mor24v	
21.579	-100.393	0.200	-1.300	30.315	-39.407	Mor25v	
82.941	-24.084	0.274	1.106	26.600	29.429	Mor27v	
145.900	-12.378	0.096	1.697	39.338	66.761	Mor28v	
52.508	-84.850	0.638	-0.474	34.139	-16.171	Mor30v	
62.976	-18.971	0.286	1.080	20.367	22.002	Mor31v	
5.199	-11.880	0.435	-0.787	4.245	-3.340	Mor17e	
8.376	-13.685	0.635	-0.477	5.458	-2.604	Mor29e	

III-4-2-1-Analyse des résultats

$$R^2=0.529 \text{ et } R^2 \text{ ajusté}=0.448$$

Donc pourcentage **91,9** %de variation de la variable à modéliser. Ce la nous indiquer que **91,9%** de la variabilité du **logK_{ow}** est expliquée par les descripteurs. Le reste de la variabilité est dû à des effets (autres variables explicatives) qui ne sont pas pris en compte ou bien les descripteurs sélectionnés ne sont pas totalement descriptives.

III-4-2-2-Analyse des résidus;

D'après le tableau: Le résidu = $1-R^2 = 1-0.919=0.081= 100 *0.081 =8.1\%$

Cela signifie, entre autres, que **91,9 %** des résidus doivent se trouver dans l'intervalle, toute valeur en dehors de cet intervalle est révélatrice d'une donnée suspecte. Afin de mettre en évidence les valeurs se trouvant hors de l'intervalle. Les paramètres permettent d'expliquer **91.9%** de la variabilité du $\log K_{ow}$. Pour expliquer les variabilités restantes, d'autres sources de variabilité doivent donc être prises en compte.

III-4-2-3-Discussion

L'étude de corrélation entre les descripteurs deux à deux ont conduit à classer les descripteurs entre deux extrêmes : Les descripteurs moins corrélés (Mor28v, Mor05u)=-0,001, (Mor27v, Mor25m)=-0,00022 (Mor3m, Mor19m)=-0,005. Qui signifie qu'ils ne donnent pas la même information sur la propriété à estimer ($\log K_{ow}$) (chaque descripteur à une information différente à celle de l'autre). C.à.d que l'information sur 3D-MoRSE - signal 28/pesés par les volumes atomiques de van der Waals est différente à celle de l'information sur (Mor05u) qui représente 3D-MoRSE - signal 05/non pesés, et que l'information donnée par Mor27v qui représente 3D-MoRSE - signal 27/pesés par les volumes atomiques de van der Waals, est différente à celle des informations données par 3D-MoRSE - signal 25/pesés par les masses atomiques de van der Waals, et que l'information donnée par Mor3m qui représente 3D-MoRSE - signal 3/ pesés de masse atomique est différente à celle des informations données par 3D-MoRSE - signal 19/ pesés de masse atomique. Ces descripteurs peuvent être importants pour le modèle.

Des descripteurs possédants une grande corrélation entre eux (Mor04u, Mor02u)=-0,581, (Mor19m, Mor02u)=0,749. Cela signifie qu'ils vont nous donner des informations plus proches sur la propriété à estimer ($\log K_{ow}$). C.à.d que l'information sur 3D-MoRSE - signal 4/ non pesés est proche à celle de 3D-MoRSE - signal 2/ non pesés, et 3D-MoRSE - signal 19/ pesés de masse atomique est proche 3D-MoRSE - signal 2/ non pesés). Ces descripteurs peuvent être éliminés du modèle.

Le calcul de la matrice de corrélation nous a permis d'éliminer **75** autres descripteurs puisqu'ils nous ramènent à la même information que les 9 restants $R > 0,95$. Les descripteurs (Mor08v, Mor28m, Mor24u) sont respectivement les plus corrélés avec $\log K_{ow}$. Ces variables nous donnent grandes informations sur $\log K_{ow}$ sont parmi les paramètres du modèle. Le descripteurs (Mor01v) est moins corrélé avec $\log K_{ow}$, il nous donnent peut d'informations sur $\log K_{ow}$, et par conséquent ne seront pas présents sur l'équation du modèle.

III-4-3-Etude de propriété de $\log P_v$

Tableau III-17 : Coefficient d'ajustement.

89.000	Observations
89.000	Somme des poids
15.000	DDL
0.973	R ²
0.844	R ² ajusté

Tableau III-18 : Analyse de variance

Pr>F	F	Moyenne des carrés	Somme des carrés	DDL	Source
<0.0001	7.506	146089236.154	10664514239.231	73	Modèle
		19461917.234	291928758.506	15	Erreur
			10956442997.737	88	Total corrigé

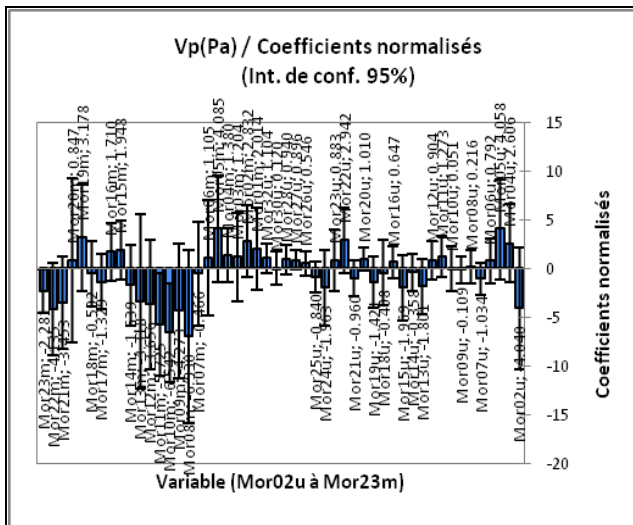


Figure III-21 : V_p en fonction de coefficients normalisés.

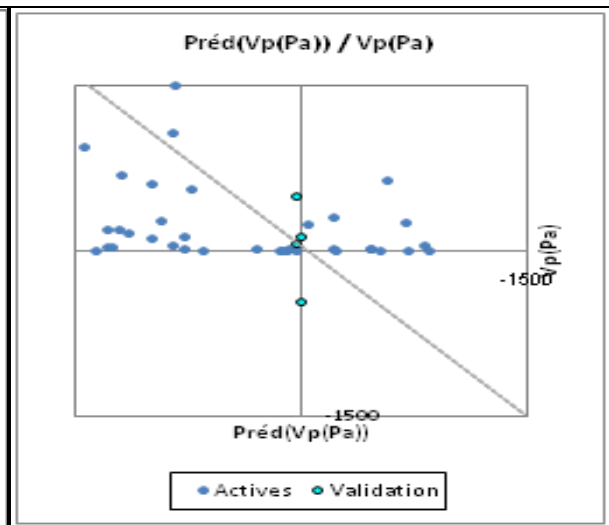


Figure III-22 : $\text{préd}(V_p)$ en fonction de V_p

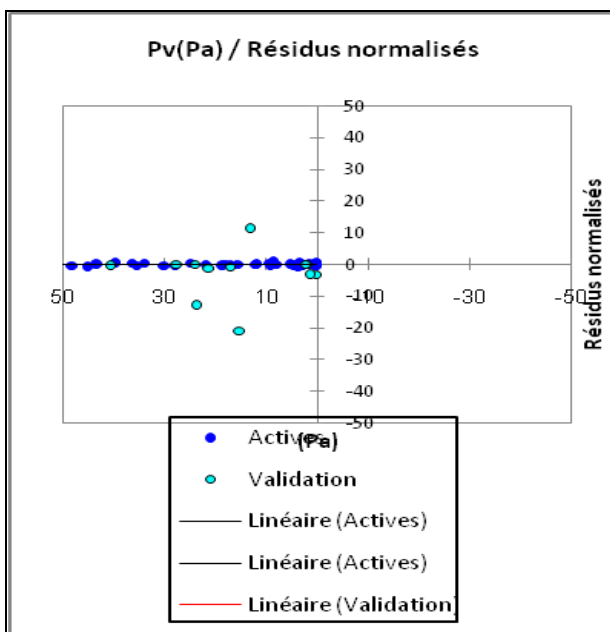


Figure III-23 : V_p en fonction de résidus normalisés.

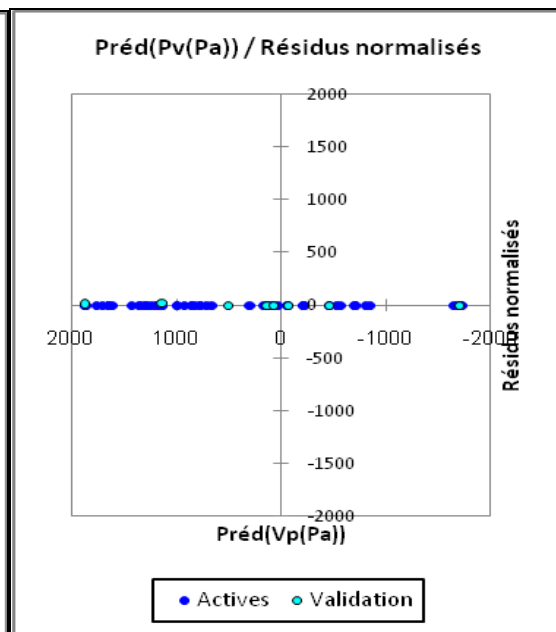


Figure III-24 : $\text{préd}(V_p)$ en fonction de résidus normalisés

Tableau III-19 : Paramètre de modèle

e supérieure	Se inférieure	Pr > t	t de Student	Ecart-type	Valeur	Paramètre
35556.574	-43496.380	0.841	-0.202	19647.900	-3969.903	Constante
5471.806	-2359.319	0.428	0.800	1946.356	1556.243	Mor02u
3449.868	-17438.647	0.184	-1.347	5191.652	-6994.390	Mor04u
26421.757	-30175.593	0.894	-0.133	14066.762	-1876.918	Mor05u
21444.305	-16327.573	0.786	0.273	9387.860	2558.366	Mor06u
29347.480	-15821.959	0.550	0.602	11226.458	6762.761	Mor07u
12553.736	-64509.904	0.181	-1.356	19153.474	-25978.084	Mor08u
25491.537	-58695.539	0.432	-0.793	20923.940	-16602.001	Mor09u
24574.699	-41838.205	0.603	-0.523	16506.329	-8631.753	Mor10u
27272.861	-42149.331	0.668	-0.431	17254.261	-7438.235	Mor11u
25114.845	-59786.956	0.415	-0.822	21101.579	-17336.056	Mor12u
52925.737	-39609.955	0.773	0.289	22998.914	6657.891	Mor13u
57917.753	-41619.800	0.743	0.329	24739.163	8148.976	Mor14u
128834.063	-10206.733	0.093	1.716	34557.339	59313.665	Mor15u
86584.722	-24738.195	0.269	1.118	27668.309	30923.263	Mor16u
32431.069	-120262.906	0.253	-1.157	37950.713	-43915.918	Mor18u
7523.262	-110819.294	0.086	-1.756	29412.977	-51648.016	Mor19u
37087.338	-88255.025	0.416	-0.821	31152.716	-25583.843	Mor20u
145040.971	-35684.913	0.230	1.217	44917.792	54678.029	Mor21u
101276.121	-125831.244	0.829	-0.218	56445.492	-12277.562	Mor22u
67833.928	-128587.791	0.537	-0.622	48818.851	-30376.931	Mor23u
106146.995	-134594.356	0.813	-0.238	59834.096	-14223.680	Mor24u
132967.612	-86105.865	0.669	0.430	54448.741	23430.874	Mor25u
74595.039	-75703.449	0.988	-0.015	37355.337	-554.205	Mor26u
177788.828	-43792.070	0.230	1.217	55071.938	66998.379	Mor27u
94515.486	-132546.395	0.738	-0.337	56434.187	-19015.454	Mor28u
143769.962	-109904.899	0.789	0.269	63048.604	16932.532	Mor30u
109231.040	-28149.674	0.241	1.187	34144.740	40540.683	Mor32u
0.664	-0.382	0.571	0.581	0.242	0.141	Mor01m
5.613	-1.403	0.217	1.296	1.624	2.105	Mor02m
23.558	-1.000	0.069	1.984	5.684	11.279	Mor03m
12.987	-10.701	0.838	0.209	5.482	1.143	Mor04m
40.028	-8.520	0.184	1.402	11.236	15.754	Mor05m
33.912	1.041	0.039	2.297	7.608	17.476	Mor06m
15.168	-8.152	0.527	0.650	5.397	3.508	Mor07m
19.318	-20.117	0.966	-0.044	9.127	-0.399	Mor08m
29.357	-31.096	0.951	-0.062	13.992	-0.869	Mor09m
36.578	-25.230	0.698	0.397	14.305	5.674	Mor10m
47.001	-9.267	0.171	1.449	13.023	18.867	Mor11m
77.213	1.604	0.042	2.252	17.499	39.409	Mor12m
98.532	9.080	0.022	2.599	20.703	53.806	Mor13m
7.990	-34.618	0.200	-1.350	9.861	-13.314	Mor14m
39.547	-43.350	0.923	-0.099	19.186	-1.902	Mor15m
5.451	-80.721	0.082	-1.887	19.944	-37.635	Mor16m
16.505	-64.451	0.223	-1.279	18.737	-23.973	Mor17m
94.332	-48.262	0.497	0.698	33.002	23.035	Mor18m
62.893	-5.033	0.089	1.840	15.721	28.930	Mor19m
33.083	-18.582	0.555	0.606	11.957	7.251	Mor20m
57.947	-27.686	0.459	0.763	19.819	15.131	Mor21m
64.156	-15.427	0.209	1.323	18.419	24.364	Mor22m
55.885	-46.993	0.855	0.187	23.810	4.446	Mor23m
13.579	-67.510	0.174	-1.437	18.768	-26.965	Mor24m
37.140	-31.656	0.866	0.172	15.922	2.742	Mor25m
73.225	-38.994	0.521	0.659	25.972	17.116	Mor26m
127.387	-18.071	0.128	1.624	33.665	54.658	Mor27m
55.969	-28.692	0.499	0.696	19.594	13.639	Mor28m
96.744	-36.216	0.343	0.983	30.773	30.264	Mor29m
51.392	-58.971	0.884	-0.148	25.543	-3.789	Mor30m
6.157	-100.179	0.078	-1.910	24.611	-47.011	Mor31m
85.970	-78.965	0.928	0.092	38.173	3.503	Mor32m
0.162	-0.595	0.239	-1.235	0.175	-0.216	Mor01v
1.941	-7.639	0.221	-1.285	2.217	-2.849	Mor02v
5.972	-15.268	0.362	-0.945	4.916	-4.648	Mor03v
20.212	-17.922	0.899	0.130	8.826	1.145	Mor04v
11.595	-50.867	0.197	-1.358	14.456	-19.636	Mor05v
14.332	-25.912	0.545	-0.622	9.314	-5.790	Mor06v
82.196	-24.511	0.264	1.168	24.697	28.842	Mor08v
104.493	-27.527	0.230	1.259	30.555	38.483	Mor09v
36.794	-30.164	0.834	0.214	15.497	3.315	Mor10v
62.125	-28.038	0.429	0.817	20.867	17.043	Mor11v
62.777	-56.291	0.908	0.118	27.557	3.243	Mor12v
23.534	-131.628	0.156	-1.505	35.911	-54.047	Mor13v
114.102	-72.894	0.642	0.476	43.279	20.604	Mor15v
1.941	-7.639	0.221	-1.285	0.000	0.000	Mor16v
121.802	-129.972	0.945	-0.070	58.271	-4.085	Mor18v
0.162	-0.595	0.239	-1.235	0.000	0.000	Mor20v
87.016	-94.966	0.926	-0.094	42.118	-3.975	Mor21v
82.196	-24.511	0.264	1.168	0.000	0.000	Mor22v
104.493	-27.527	0.230	1.259	0.000	0.000	Mor23v
65.837	-86.507	0.774	-0.293	35.259	-10.335	Mor24v
15.168	-8.152	0.527	0.650	5.397	0.000	Mor25v
19.318	-20.117	0.966	-0.044	9.127	0.000	Mor27v
29.357	-31.096	0.951	-0.062	13.992	0.000	Mor28v
36.578	-25.230	0.698	0.397	14.305	0.000	Mor30v
47.001	-9.267	0.171	1.449	13.023	0.000	Mor31v
77.213	1.604	0.042	2.252	17.499	0.000	Mor17e
98.532	9.080	0.022	2.599	20.703	0.000	Mor29e

III-4-3-1-Analyse des résultats

$$R^2=0.973 \text{ et } R^2 \text{ ajusté}=0.844$$

Donc pourcentage **97,3** % de variation de la variable à modéliser. Ce nous l'indiquer sélectionnés que 97,3% de la variabilité du P_V est expliquée par les descripteurs. Le reste de la variabilité est dû à des effets (autres variables explicatives) qui ne sont pas pris en compte ou bien les descripteurs ne sont pas totalement descriptifs.

III-4-3-2-Analyse des résidus

$$\text{D'après le tableau: Le résidu} = 1 - R^2 = 1 - 0.973 = 0.027 = 100 * 0.027 = 2.7\%$$

Cela signifie, entre autres, que **97,3** % des résidus doivent se trouver dans l'intervalle, toute valeur en dehors de cet intervalle est révélatrice d'une donnée suspecte. Afin de mettre en évidence les valeurs se trouvant hors de l'intervalle. Les paramètres permettent d'expliquer **97.3%** de la variabilité du P_V . Pour expliquer les variabilités restantes, d'autres sources de variabilité doivent donc être prises en compte.

III-4-3-3-Discussion

L'étude de corrélation entre les descripteurs deux à deux ont conduit à classer les descripteurs entre deux extrêmes : Les descripteurs moins corrélés (Mor11v, Mor25u) = -0,00046, (Mor11v, Mor11m) = -0,009 (Mor17e, Mor16m) = -0,005 Qui signifie qu'ils ne donnent pas la même information sur la propriété à estimer ($\log K_{ow}$) (chaque descripteur à une information différente à celle de l'autre). C.à.d que l'information sur 3D-MoRSE - signal 11/pesés par les volumes atomiques de van der Waals est différente à celle de l'information sur (Mor25u) qui représente 3D-MoRSE - signal 25/non pesés, et que l'information donnée par Mor11v qui représente 3D-MoRSE - signal 11/pesés par les volumes atomiques de van der Waals, est différente à celle des informations données par 3D-MoRSE - signal 11/pesés par masse atomiques de van der Waals, et que l'information donnée par Mor17e qui représente 3D-MoRSE - signal 17/pesés par Sanderson atomique électronégativités est différente à celle des informations données par 3D-MoRSE - signal 16/ pesés de masse atomique. Ces descripteurs peuvent être importants pour le modèle.

Des descripteurs possédants une grande corrélation entre eux ($\text{Mor10u, Mor04u} = 0,263, (\text{Mor06m, Mor18u}) = -0,112$). Cela signifie qu'ils vont nous donner des informations plus proches sur la propriété à estimer (\mathbf{P}_V). C.à.d que l'information sur 3D-MoRSE - signal 10/ non pesés est proche à celle de 3D-MoRSE - signal 4/ non pesés, et 3D-MoRSE - signal 06/ pesés de masse atomique est proche 3D-MoRSE - signal 18/ non pesés). Ces descripteurs peuvent être éliminés du modèle.

Le calcul de la matrice de corrélation nous a permis d'éliminer **75** autres descripteurs puisqu'ils nous ramènent à la même information que les 9 restants $R > 0,95$. Les descripteurs ($\text{Mor12v, Mor22m, Mor16u}$) sont respectivement les plus corrélés avec \mathbf{P}_V . Ces variables nous donnent grandes informations sur \mathbf{P}_V sont parmi le paramètre du modèle. Le descripteurs (Mor01v, Mor01m) est moins corrélé avec \mathbf{P}_V , il nous donnent peut d'informations sur \mathbf{P}_V , et par conséquent ne seront pas présents sur l'équation du modèle.

CONCLUSION GENERAL

CONCLUSION GENERAL:

D'après les analyses statistiques des valeurs de (propriétés)_{préd} et (propriétés)_{exp}, (résidus normalisés) et l'interprétation des graphes, on voit que les valeurs de résidus varient, et les valeurs qui sont hors de l'intervalle de confiance présentent les points aberrants qui indiquent les molécules non corrèles.

Ces molécules possèdent des structures chimiques bidimensionnelles irrégulières. Cette irrégularité engendre par l'existence des halogènes comme (Br,Cl), des liaisons d'hydrogène, van der waals, et l'aromaticité, ou bien des groupements comme (NH₂) qui sont capables de donner un doublet électronique.

La qualité du modèle dépend de nombre des points aberrants, et de nombre des variables (paramètres). La valeur de R^2 nous indique le degré de corrélation des propriétés avec les paramètres structuraux des différents blocs est très petite car la base de données est très étroite et ne permet pas d'identifier un bon modèle qui correspondent la prédiction des propriétés.

Si on compare les résultats des blocks entre-deux on peut dire que :

- ✓ Les paramètres du block 2 donnent une représentation suffisante pour la structure moléculaire des composées de la base des données utilisée.
- ✓ Les paramètres du block 14 donnent une représentation suffisante des liaisons **d'hydrogène, van der waals, et l'aromaticité.**
- ✓ D'après ces résultats, on observe que le degré d'estimation des propriétés change suivant la qualité de descripteurs de chaque bloc.
- ✓ D'après cette étude et ces résultats on peut identifier et classier les composées odorants suivants leur propriétés qui dépende des valeurs de (log_{ow}), log ODT, et P_v.

Enfin on recommande les points suivant:

- ✓ Collection des plusieurs molécules pour élargisse la base de donner pour donner des mieux résultats.
- ✓ Fait la corrélation avec des autres outilles qui modaliser les propriétés par petits erreurs.

BIBLIOGRAPHIE

BIBLIOGRAPHIE

- [1] **Enrico Riboni., Myriam Robert.** (2000), *Traitement des odeurs par ozonisation dans les stations d'épuration des eaux usées.*
- [2] **Martin G., Laffort P.** (1991), *Odeurs et désodorisations dans l'environnement*, Lavoisier, Tec&Doc, Paris.
- [3] <http://www.silapedagogie.com/odorat.htm/>
- [4] <http://www.filsantejeunes.com/L-odorat/>
- [5] **Institut national de recherche et de sécurité (INRS).** *Comparaison des seuils olfactifs de substances chimiques avec des indicateurs de sécurité utilisés en milieu professionnel.* Cahiers de notes documentaires no. 156, 3e trimestre. Paris; 1994.
- [6] **Anne-Marie GOURONNEC** Ingénieur à l'IRSN « *Analyses olfactométriques ou mesure des odeurs par analyse sensorielle* » (Institut de radioprotection et de sûreté nucléaire) © Techniques de l'Ingénieur P 446 1 – 21.
- [7] **J.-P. Dumont** « *Arômes et saveurs des aliments* ».
- [8] **Karen J. Rossiter**, “*Structure - Odor Relationships*”, Chem. Rev. 1996, 96, 3201-3240
- [9] **Altervino** – 2006 « *Cours de dégustation* ».
- [10] **Division Environnement Amiens** « *NOTE sur les odeurs* » 15 octobre 2004
MH/MD – 2004-1019
- [11] Zald, D. H. and Pardo, J. V. (2000), *Functional neuroimaging of the olfactory system in humans* Int J Psychophysiol 36, 165-81.
- [12] Tegoni, M., Pelosi, P., Vincent, F., Spinelli, S., Campanacci, V., Grolli, S., Ramoni, R. and Cambillau, C. (2000), *Mammalian odorant binding proteins* Biochim Biophys Acta 1482, 229-40
- [13] **Popescu M., Blanchard J.-M., Jean C.** (1998), *Analyse et traitement physicochimique des rejets atmosphériques industriels - Emissions, fumées, odeurs et poussières*, Lavoisier, Tec&Doc, Paris.
- [14] **Rognon C., Pourtier L.** (2000), *Les odeurs dans l'environnement*, Les techniques de l'ingénieur Traité environnement (G 2900): 1-11.

Bibliographie

- [15] **Gostelow P., Parsons S.A., Stuetz R.M.** (2001), *Odour measurements for sewage treatment works*, Water Research **35** (3): 579-597.
- [16] **Perrin M.-L., Thal M.-F., Zettwoog P.** (1991), *Olfactométrie dans l'industrie- Mesures des odeurs à l'émission et dans l'environnement*, Les techniques de l'ingénieur Traité analyse/mesures (P 445): 1-8
- [17] **Gouronnec A.-M.** (2004), *Analyse olfactométrique ou mesure des odeurs par analyse sensorielle*, Les techniques de l'ingénieur Traité techniques d'analyse (p 446): 1-21.
- [18] **A.R. Katritzky., D.C. Fara., R.O. Petrukhin**, (2002), Top. Med. Chemm 1333-1356.
- [19] **Hansch C.** (1969), a quantitative approach to biochemical structure activity relationships. Accounts of chemical research, 2, 232-239.
- [20] **R. Leardi** , (2001), *Chemometr.* 15, 559-569.
- [21] **Talete srl**, Dragon Profesional Software, Via V.Pisani, 13-20124 Milano (ITALY), 2005.
- [22] Dragon_ Aide blocs des descripteurs.
- [23] **Saadi Khaled.** (2009).Contribution á l'étude de la Relation structure chimique-odeur Utilisation de la technique Random Forest (Application à la famille des pyrazines), Mémoire de Magister. UNIVERSITÉ KASDI Merbah Ourgla.p30.
- [24] **AI ACCESS**, 91940, Les Ulis, Franonce.
- [25] **SOUYEI Belgacem.** (2007).contribution à la prédiction de coefficient de partage Octanol/eau par la technique QSPR. Mémoire de Magister. UNIVERSITÉ KASDI Merbah de Ourgla. P 45 à 57.
- [26] **Thomas-Danguin T.** (1997).Intensité olfactive des composés purs et de mélanges:application au masquage des odeurs, Université Claude Bernard,Lyon,p224.
- [27] **Martin G., Laffort P.** (1991).Odeurs et désodorisations dans l'environnement, Lavoisier, Tec&Doc, Paris.
- [28] **Le Cloirec P.** (2002). *Introduction au traitement de l'air*, Les techniques de l'ingénieur Traité environnement (G 1700): 1-8.
- [29] **Afnor.** (1996). *Qualité de l'air-Mesurage de l'odeur d'un effluent gazeux. Méthodes supraliminales*, NF X 43-103.
- [30] **Moletta R.** (2002), *Gestion des problèmes environnementaux dans les industries agroalimentaires*, Lavoisier, Tec&Doc, Paris.

Bibliographie

[31] **GHAOUES SOUHEILA.** (2011). Evaluation de la qualité physico-chimique et organoleptique de cinq marques de laits reconstitués partiellement écrémés commercialisés dans l'est Algérien 2010/2011, Mémoire de Magister Université Hadj Lakhdar Batna. p50.

dedicaces

J'ai le grand honneur de dédier ce modeste travail A celui qui a fait de moi un homme, mon père.

A l'être le plus cher de ma vie, ma mère.

A mes frères (mini, moussa, sidelli, taha, dalile, dodo, qador, habchi, Abd Allah, jma, nadire, dahma, joura, ali, yassin, nour eldine, belkassame, momni, abd alssamed, saih, mohammed.)

Et mes sœurs (nana, halouma, bassoussa, sarito)

A toute ma famille et surtout mon grande mère.

A tous mes amis, à tous ceux qui sèment le bonheur dans mon chemin.

A tous les enseignants et toutes les enseignantes science de matière.

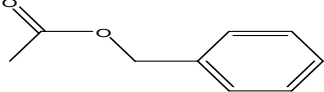
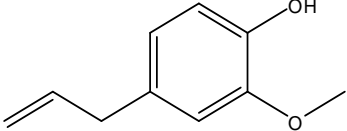
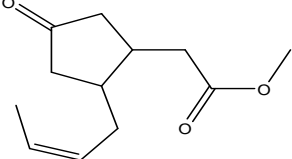
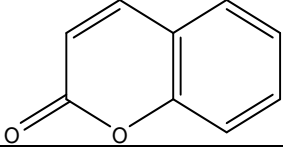
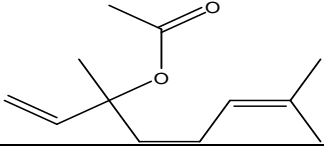
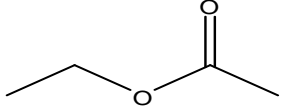
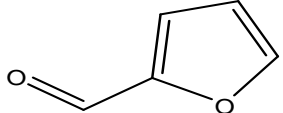
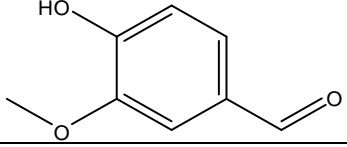
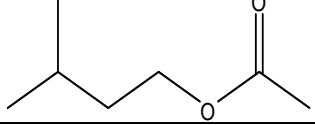
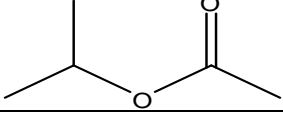
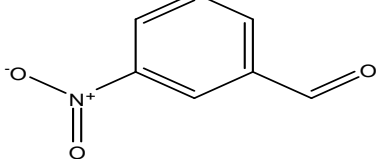
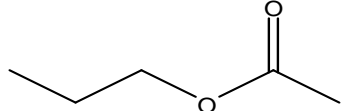
A tous la promotion master chimie appliqué

A mon cher pays « l'ALGERIE »,

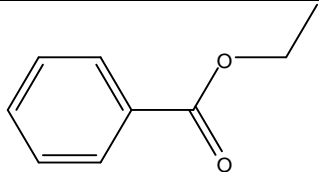
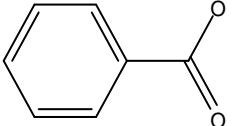
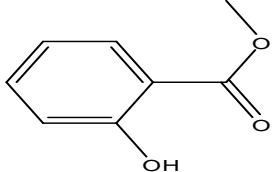
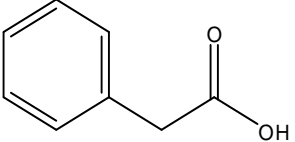
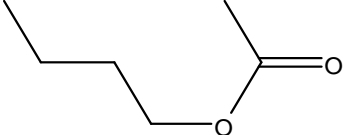
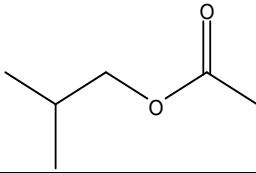
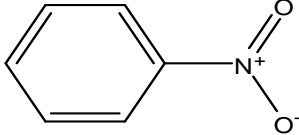
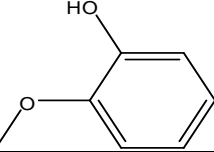
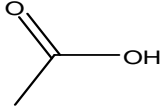
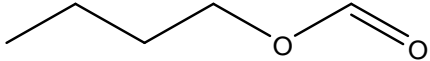
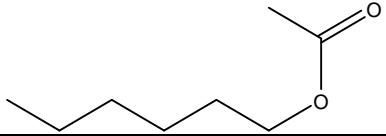
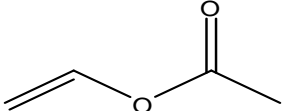
Regadi Dahmane.

ANNEXES

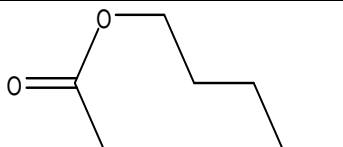
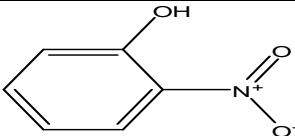
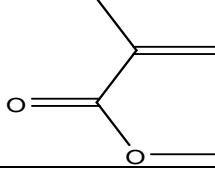
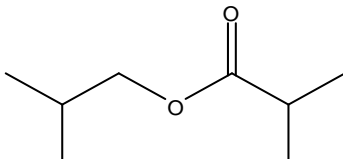
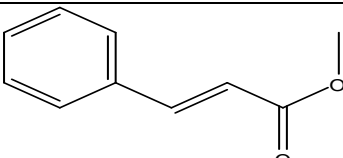
LES ANNEXES

No	Molécules	CAS	Formule	Structure
5	Benzyl Acetate	140-11-4	$C_9H_{10}O_2$	
7	Eugenol	97-53-0	$C_{10}H_{12}O_2$	
8	Methyldihydro jasmonate	24851-98-7	$C_{13}H_{22}O_3$	
9	Coumarin	91-64-5	$C_9H_6O_2$	
16	Linalyl Acetate	115-95-7	$C_{12}H_{20}O_2$	
17	Ethyl Acetate	141-78-6	$C_4H_8O_2$	
18	Furfural	98-01-1	$C_5H_4O_2$	
22	Vanillin	121-33-5	$C_8H_8O_3$	
23	Isoamyl Acetate	123-92-2	$C_7H_{14}O_2$	
24	Isopropyl Acetate	108-21-4	$C_5H_{10}O_2$	
27	3-Nitrobenzaldehyde	99-61-6	$C_7H_5NO_3$	
34	Propyl Acetate	109-60-4	$C_5H_{10}O_2$	

LES ANNEXES

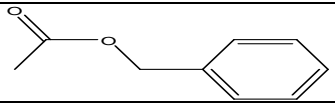
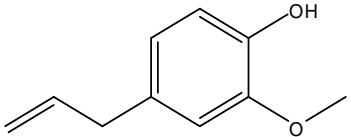
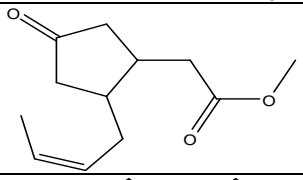
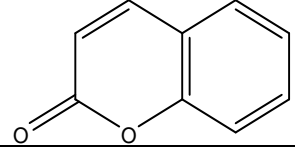
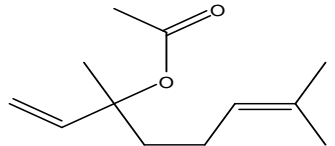
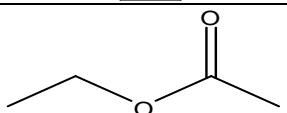

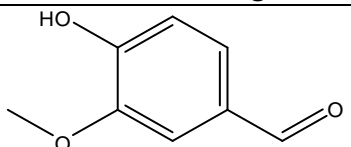
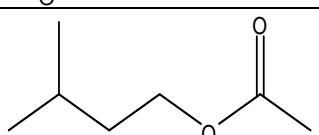
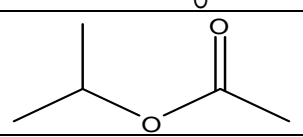
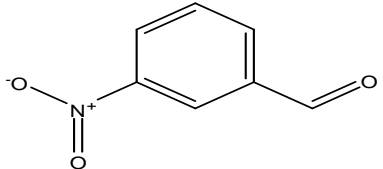
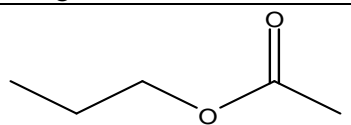
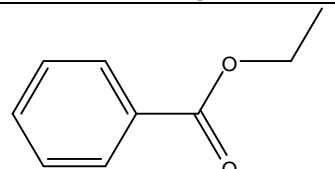
46	Ethyl Benzoate	93-89-0	$C_9H_{10}O_2$	
56	Methyl Benzoate	93-58-3	$C_8H_8O_2$	
57	Methyl Salicylate	119-36-8	$C_8H_8O_3$	
58	Phenylacetic Acid	103-82-2	$C_8H_8O_2$	
61	n-Butyl Acetate	123-86-4	$C_6H_{12}O_2$	
62	Isobutyl Acetate	110-19-0	$C_6H_{12}O_2$	
77	Nitrobenzene	98-95-3	$C_6H_5NO_2$	
81	Anethole	104-46-1	$C_{10}H_{12}O$	
83	Acetic Acid	64-19-7	$C_2H_4O_2$	
84	Butyl Formate	592-84-7	$C_5H_{10}O_2$	
89	Hexyl Acetate	142-92-7	$C_8H_{16}O_2$	
96	Vinyl Acetate	108-05-4	$C_4H_6O_2$	

LES ANNEXES

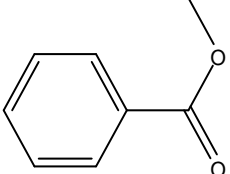
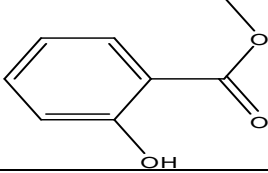
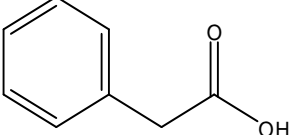
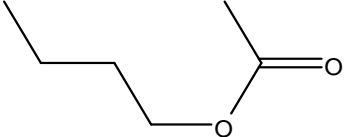
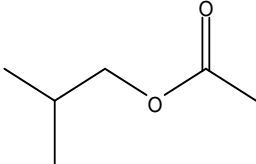
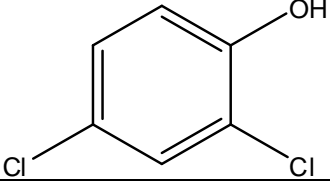
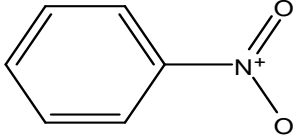
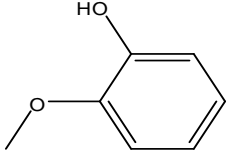
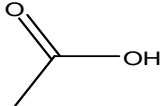
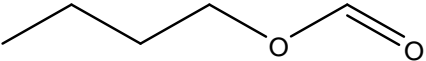
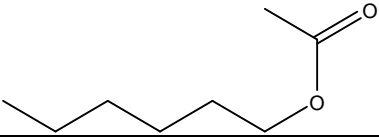
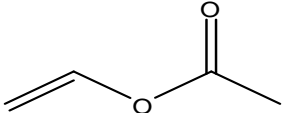
104	Pentyl Acetate	628-63-7	$C_7H_{14}O_2$	
107	2-Nitrophenol	88-75-5	$C_6H_5NO_3$	
109	Methyl	80-62-6	$C_5H_8O_2$	
121	Dodecanoic acid	143-07-7	$C_{12}H_{24}O_2$	
127	4-Methyl-2-pentanol	108-11-2	$C_6H_{14}O$	

Annex 01 : identifier les composés non corrèles dans ce block à partir de résidu et LogK_{ow}.

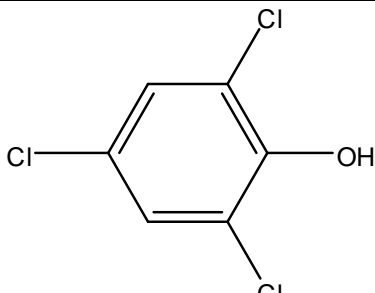
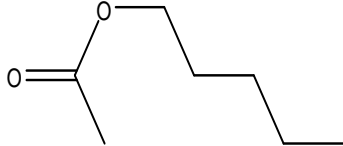
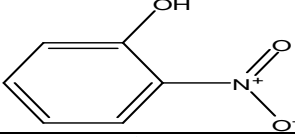
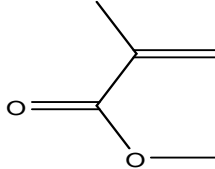
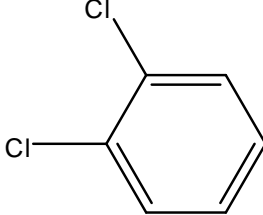
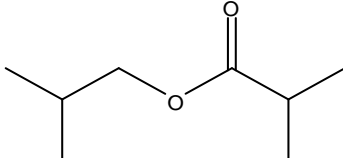
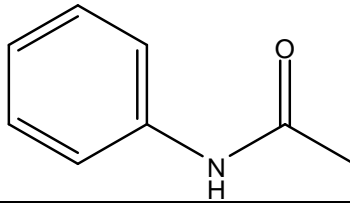
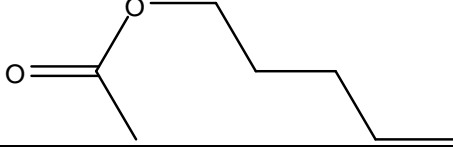
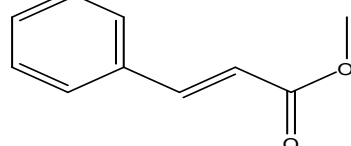
LES ANNEXES

No	Molécules	CAS	Formule	Structure
5	Benzyl Acetate	140-11-4	$C_9H_{10}O_2$	
7	Eugenol	97-53-0	$C_{10}H_{12}O_2$	
8	Methyl dihydrojasmonate	24851-98-7	$C_{13}H_{22}O_3$	
9	Coumarin	91-64-5	$C_9H_6O_2$	
16	Linalyl Acetate	115-95-7	$C_{12}H_{20}O_2$	
17	Ethyl Acetate	141-78-6	$C_4H_8O_2$	
18	Furfural	98-01-1	$C_5H_4O_2$	
22	Vanillin	121-33-5	$C_8H_8O_3$	
23	Isoamyl Acetate	123-92-2	$C_7H_{14}O_2$	
24	Isopropyl Acetate	108-21-4	$C_5H_{10}O_2$	
27	3-Nitrobenzaldehyde	99-61-6	$C_7H_5NO_3$	
34	Propyl Acetate	109-60-4	$C_5H_{10}O_2$	
46	Ethyl Benzoate	93-89-0	$C_9H_{10}O_2$	

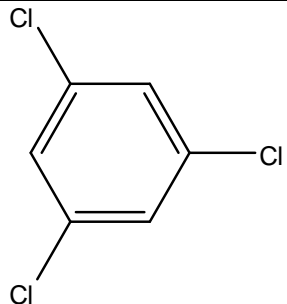
LES ANNEXES

56	Methyl Benzoate	93-58-3	$C_8H_8O_2$	
57	Methyl Salicylate	119-36-8	$C_8H_8O_3$	
58	Phenylacetic Acid	103-82-2	$C_8H_8O_2$	
61	n-Butyl Acetate	123-86-4	$C_6H_{12}O_2$	
62	Isobutyl Acetate	110-19-0	$C_6H_{12}O_2$	
69	2,4-Dichlorophenol	120-83-2	120-83-2	
77	Nitrobenzene	98-95-3	$C_6H_5NO_2$	
81	Anethole	104-46-1	$C_{10}H_{12}O$	
83	Acetic Acid	64-19-7	$C_2H_4O_2$	
84	Butyl Formate	592-84-7	$C_5H_{10}O_2$	
89	Hexyl Acetate	142-92-7	$C_8H_{16}O_2$	
96	Vinyl Acetate	108-05-4	$C_4H_6O_2$	

LES ANNEXES

99	2,4,6-Trichlorophenol	88-06-2	$C_6H_3Cl_3O$	
104	Pentyl Acetate	628-63-7	$C_6H_{12}O_2$	
107	2-Nitrophenol	88-75-5	$C_6H_5NO_3$	
109	Methyl	80-62-6	$C_5H_8O_2$	
115	1,2-Dichlorobenzene	88-06-2	$C_6H_4Cl_2$	
121	Dodecanoic acid	143-07-7	$C_{12}H_{24}O_2$	
122	Acetanilide	103-84-4	C_8H_9NO	
124	Pentyl Acetate	628-63-7	$C_7H_{14}O_2$	
128	Isobutyl Isobutyrate	97-85-8	$C_8H_{16}O_2$	

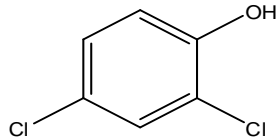
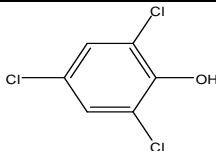
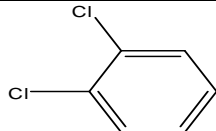
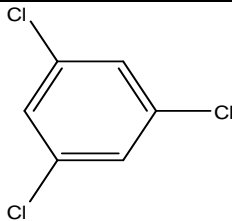
LES ANNEXES

133	1,3,5- Trichlorobenzene	108-70-3	$C_6H_3Cl_3$	
-----	----------------------------	----------	--------------	---

ANNEX 02 : identifier les composés non corrèles dans ce block à partir de résidu et **Pv(exp)**

.

LES ANNEXES

No	Composé	CAS	Formule	structure
69	2,4-Dichlorophenol	120-83-2	$C_6H_3Cl_2O$	
99	2,4,6-Trichlorophenol	88-06-2	$C_6H_2Cl_3O$	
114	1,2-Dichlorobenzène	95-50-1	$C_6H_4Cl_2$	
133	1,3,5-Trichlorobenzène	108-70-3	$C_6H_3Cl_3$	

Annex 03 : identifier les composés non corrèles dans ce block à partir de résidu **log ODT**

	Molécules	CAS	Formule	M (g/mol)	log ODT(mg/m ³)	Pv (Pa)	log Kow
1	Limonene	01-11-7721	C ₁₀ H ₁₆	136.23	-0.165	2.05E+02	4.45
2	Linalool	78-70-6	C ₁₀ H ₁₈ O	154.25	-1.901	1.21E+01	3.28
3	Nonanal	124-19-6	C ₉ H ₁₈ O	142.24	-1.716	7.09E+01	3.56
4	p-Cymene	99-87-6	C ₁₀ H ₁₄	134.22	-0.255	1.95E+02	4.02
5	Benzyl Acetate	140-11-4	C ₉ H ₁₀ O ₂	150.17	-0.575	2.19E+01	1.93
6	p-Cresol	106-44-5	C ₇ H ₈ O	108.14	-2.496	2.81E+01	1.94
7	Eugenol	97-53-0	C ₁₀ H ₁₂ O ₂	164.2	-2.119	1.39E+00	2.2
8	Methylidihydro jasmonate	24851-98-7	C ₁₃ H ₂₂ O ₃	226.31	-2.643	9.47E-02	2.5
9	Coumarin	91-64-5	C ₉ H ₆ O ₂	146.14	-2.504	1.73E-01	1.39
10	Aniline	62-53-3	C ₆ H ₇ N	92.13	0.594	9.77E+01	0.94
11	Benzaldehyde	100-52-7	C ₇ H ₆ O	106.13	-0.081	1.30E+02	1.64
12	Benzene	71-43-2	C ₆ H ₆	78.11	1.406	1.35E+04	2.2
13	Butanal	123-72-8	C ₄ H ₈ O	72.11	-1.014	1.28E+04	0.91
14	1-Butanol	71-36-3	C ₄ H ₁₀ O	74.12	0.58	1.14E+03	0.88
15	o-Xylene	95-47-6	C ₈ H ₁₀	106.17	0.382	7.99E+02	3.14
16	Linalyl Acetate	115-95-7	C ₁₂ H ₂₀ O ₂	196.29	-0.577	1.55E+01	3.83
17	Ethyl Acetate	141-78-6	C ₄ H ₈ O ₂	88.11	1.263	1.49E+04	0.71
18	Furfural	98-01-1	C ₅ H ₄ O ₂	96.09	-0.677	2.97E+02	0.73
19	1-Heptanol	111-70-6	C ₇ H ₁₆ O	116.2	-0.099	4.33E+01	2.47
20	2-Heptanone	110-43-0	C ₇ H ₁₄ O	114.19	-0.363	6.31E+02	1.97
21	Hexanal	66-25-1	C ₆ H ₁₂ O	100.16	-1.135	1.45E+03	1.97
22	Vanillin	121-33-5	C ₈ H ₈ O ₃	152.15	-4.078	5.12E-02	1.1
23	Isoamyl Acetate	123-92-2	C ₇ H ₁₄ O ₂	130.19	-0.522	7.57E+02	2.12
24	Isopropyl Acetate	108-21-4	C ₈ H ₁₆ O ₂	130.19	1.156	8.09E+03	1.06
25	Isoamyl Alcohol	123-51-3	C ₈ H ₁₈ O	88.15	-0.477	5.55E+02	1.22
26	Methacrylonitrile	126-98-7	C ₅ H ₈ N	67.09	1.048	7.53E+03	0.74
27	3-Nitrobenzaldehyde	99-61-6	C ₇ H ₅ NO ₃	151.12	0.477	1.29E+00	1.75
28	Nonane	111-84-2	C ₉ H ₂₀	128.26	1.933	6.17E+02	5.54

Photo de Base de données à partir de molécule 01 jusqu'a molécule 28

29	Pentane	109-66-0	C ₅ H ₁₂	72.15	3.028	7.03E+04	3.41
30	1-Pentanol	71-41-0	C ₆ H ₁₂ O	88.15	0.543	3.75E+02	1.41
31	2-Pentanone	107-87-9	C ₆ H ₁₀ O	86.13	1.056	5.15E+03	0.91
32	3-Pentanone	96-22-0	C ₆ H ₁₀ O	86.13	1.066	4.77E+03	0.91
33	Phenol	108-95-2	C ₆ H ₆ O	94.11	-0.731	8.19E+01	1.48
34	Propyl Acetate	109-60-4	C ₈ H ₁₆ O ₂	102.13	0.661	4.69E+03	1.24
35	Tetrachloromethane	56-23-5	CCl ₄	153.82	2.702	1.51E+04	2.86
36	Toluene	108-88-3	C ₇ H ₈	92.14	0.99	3.69E+03	2.68
37	Trichloromethane	67-66-3	CHCl ₃	119.38	2.173	2.67E+04	1.76
38	1,1,1-Trichloroethane	71-55-6	C ₂ H ₃ Cl ₃	133.41	2.56	1.63E+04	2.11
39	2-Chlorophenol	95-57-8	C ₆ H ₅ ClO	128.56	-1.874	1.17E+02	2.04
40	Acetophenone	98-86-2	C ₈ H ₈ O	120.15	-1.315	3.99E+01	1.67
41	Chlorobenzene	108-90-7	C ₆ H ₅ Cl	112.56	0.605	1.49E+03	2.81
42	Cyclohexane	110-82-7	C ₆ H ₁₂	84.16	2.111	1.25E+04	3.39
43	Cyclohexanone	108-94-1	C ₆ H ₁₀ O	98.15	0.368	3.99E+02	0.76
44	Dichloromethane	75-09-2	CH ₂ Cl ₂	84.93	2.687	5.97E+04	1.19
45	p-Xylene	106-42-3	C ₈ H ₁₀	106.17	0.258	1.06E+03	3.14
46	Ethyl Benzoate	93-89-0	C ₉ H ₁₀ O ₂	150.18	-0.897	2.40E+01	2.73
47	1-Octanol	111-87-5	C ₈ H ₁₈ O	130.23	-0.81	1.52E+01	3
48	Nonanol	143-08-8	C ₈ H ₂₀ O	144.26	-0.74	5.43E+00	3.53
49	Dodecanol	112-53-8	C ₁₂ H ₂₆ O	186.34	-1.813	2.79E-01	5.13
50	p-Anisaldehyde	123-11-5	C ₈ H ₈ O ₂	136.15	-2.646	3.32E+00	1.7
51	Camphor	76-22-2	C ₁₀ H ₁₈ O	152.24	-1.219	3.00E+01	2.13
52	Carvone	6485-40-1	C ₁₀ H ₁₄ O	150.22	-1.03	8.75E+00	2.27
53	Citronellol	106-22-9	C ₁₀ H ₂₀ O	156.27	-2.075	2.44E+00	3.38
54	Diphenylmethane	101-81-5	C ₁₃ H ₁₂	168.24	-1.565	2.13E+00	4.21
55	Menthol	89-78-1	C ₁₀ H ₂₀ O	156.27	-1.303	4.31E+00	3.2
56	Methyl Benzoate	93-58-3	C ₈ H ₈ O ₂	136.15	-1.885	4.53E+01	2.2
57	Methyl Salicylate	119-36-8	C ₈ H ₈ O ₃	152.15	-1.283	9.33E+00	2.23

Photo de Base de données à partir de molécule 29 jusqu'a molécule 57

Numéro	Nom de la molécule	Code CAS	Formule chimique	Poids moléculaire	LogP	LogS	LogP + LogS
59	Phenylethylalcohol	60-12-8	C ₈ H ₁₀ O	122.17	-1.696	1.85E+01	1.38
60	Thymol	89-83-8	C ₁₀ H ₁₄ O	150.22	-1.751	2.59E-01	1.19
61	n-Butyl Acetate	123-86-4	C ₈ H ₁₂ O ₂	116.16	0.302	1.53E+03	1.77
62	Isobutyl Acetate	110-19-0	C ₈ H ₁₂ O ₂	116.16	0.558	2.40E+03	1.59
63	Methyl Ethyl Ketone	78-93-3	C ₄ H ₈ O	72.11	1.476	1.53E+04	0.37
64	Methyl Isobutyl Ketone	108-10-1	C ₈ H ₁₂ O	100.16	0.298	2.43E+03	1.25
65	Propylene Oxide	75-56-9	C ₃ H ₆ O	58.08	1.993	7.63E+04	0.13
66	Heptane	142-82-5	C ₇ H ₁₆	100.21	2.809	6.03E+03	4.48
67	n-Octane	111-65-9	C ₈ H ₁₈	114.23	2.621	1.89E+03	5.01
68	1-Octyne	629-05-0	C ₈ H ₁₄	110.2	2.785	1.92E+03	3.66
69	2,4-Dichlorophenol	120-83-2	C ₆ H ₄ Cl ₂ O	163	-3.569	1.81E+01	3
70	N,N-Dimethylaniline	121-69-7	C ₈ H ₁₁ N	121.18	-1.018	6.16E+01	2.33
71	2,4-Dimethylpentane	108-08-7	C ₇ H ₁₆	100.21	3.26	1.26E+04	4.11
72	Cumene	98-82-8	C ₉ H ₁₂	120.2	-0.527	5.97E+02	3.56
73	m-Cresol	108-39-4	C ₇ H ₈ O	108.14	-2.69	2.76E+01	1.94
74	o-Cresol	95-48-7	C ₇ H ₈ O	108.14	-2.451	5.05E+01	1.94
75	2-Hexanone	591-78-6	C ₆ H ₁₂ O	100.16	0.085	1.77E+03	1.44
76	1-Naphthol	90-15-3	C ₁₀ H ₈ O	144.17	-2.387	1.85E-01	2.71
77	Nitrobenzene	98-95-3	C ₆ H ₅ NO ₂	123.11	-0.675	3.65E+01	1.95
78	Indole	120-72-9	C ₈ H ₇ N	117.15	-3.563	3.97E+00	2.14
79	Octanal	124-13-0	C ₈ H ₁₆ O	128.22	-1.596	2.76E+02	3.03
80	Guaiacol	90-05-1	C ₇ H ₈ O ₂	124.14	-2.053	2.39E+01	1.19
81	Anethole	104-46-1	C ₁₀ H ₁₂ O	148.21	-0.615	9.16E+00	3.17
82	Cinnamaldehyde	104-55-2	C ₉ H ₈ O	132.16	-2.355	3.53E+00	2.12
83	Acetic Acid	64-19-7	C ₂ H ₄ O ₂	60.05	0.271	1.85E+03	-0.29
84	Butyl Formate	592-84-7	C ₈ H ₁₆ O ₂	102.13	1.9	3.55E+03	1.36
85	Cyclohexanol	108-93-0	C ₆ H ₁₂ O	100.16	-0.407	1.17E+02	1.34
86	Cyclooctane	292-64-8	C ₈ H ₁₆	112.22	0.556	6.08E+02	4.51
87	Ethylbenzene	100-41-4	C ₈ H ₁₀	106.17	0.069	1.23E+03	3.21

Photo de Base de données à partir de molécule 59 jusqu'a molécule 87

Numéro	Nom de la molécule	Code CAS	Formule chimique	Poids moléculaire	LogP	LogS	LogP + LogS
88	Hexane	110-54-3	C ₆ H ₁₄	86.18	2.56	2.01E+04	3.94
89	Hexyl Acetate	142-92-7	C ₈ H ₁₆ O ₂	144.22	-0.581	1.85E+02	2.83
90	Anisole	100-66-3	C ₇ H ₈ O	108.14	-0.038	5.65E+02	2.13
91	Isoprene	78-79-5	C ₅ H ₈	68.12	0.74	7.32E+04	2.41
92	2-Methylnaphthalene	91-57-6	C ₁₁ H ₁₀	142.2	-2.398	8.07E+00	3.91
93	Isobutyl Alcohol	78-83-1	C ₄ H ₁₀ O	74.12	0.657	2.19E+03	0.69
94	Acrylonitrile	107-13-1	C ₃ H ₃ N	53.06	1.102	1.29E+04	0.19
95	Styrene	100-42-5	C ₈ H ₈	104.15	-0.292	8.28E+02	2.7
96	Vinyl Acetate	108-05-4	C ₄ H ₆ O ₂	86.09	-0.084	1.57E+04	0.73
97	Undecane	1120-21-4	C ₁₁ H ₂₄	156.31	1.713	7.52E+01	6.6
98	Mesitylene	108-67-8	C ₉ H ₁₂	120.2	0.092	3.09E+02	3.6
99	1,2,4-Trimethylbenzene	95-63-6	C ₉ H ₁₂	120.2	-0.431	2.56E+02	3.6
100	Iodoform	75-47-8	CHI ₃	393.73	-0.701	4.52E+00	3.51
101	2,4,6-Trichlorophenol	88-06-2	C ₆ H ₃ Cl ₃ O	197.45	-2.787	2.36E+00	3.58
102	Diphenyl Ether	101-84-8	C ₁₂ H ₁₀ O	170.21	-2.026	2.97E+00	4.21
103	Pentylbenzene	538-68-1	C ₁₁ H ₁₆	148.25	0.778	4.81E+01	4.8
104	Pentyl Acetate	628-63-7	C ₇ H ₁₄ O ₂	130.19	-0.026	5.24E+02	2.3
105	1-Octene	111-66-0	C ₈ H ₁₆	112.22	1.036	2.39E+03	4.5
106	2-Octanol	123-96-6	C ₈ H ₁₈ O	130.23	-1.331	4.08E+01	2.82
107	2-Nitrophenol	88-75-5	C ₆ H ₅ NO ₃	139.11	-1.921	1.32E+01	1.71
108	Naphthalene	91-20-3	C ₁₀ H ₈	128.18	-0.748	2.12E+01	3.45
109	Methyl Methacrylate	80-62-6	C ₅ H ₈ O ₂	100.12	-0.331	4.92E+03	1.35
110	(-)-Menthone	14073-97-3	C ₁₀ H ₁₈ O	154.25	0.046	3.41E+01	2.63
111	1-Hexen-3-ol	4798-44-1	C ₆ H ₁₂ O	100.16	-1.155	4.80E+02	1.58
112	1-Hexanol	111-27-3	C ₆ H ₁₄ O	102.18	-0.187	1.26E+02	1.94
113	Galaxolide	1222-05-5	C ₁₃ H ₂₆ O	258.41	-3.311	7.27E-02	5.9
114	2-Ethyl-1-Butanol	97-95-0	C ₈ H ₁₈ O	102.18	-0.009	2.41E+02	1.75
115	1,2-Dichlorobenzene	95-50-1	C ₆ H ₄ Cl ₂	147	-0.149	1.61E+02	3.28
116	1-Chlorobutane	109-69-3	C ₄ H ₉ Cl	92.57	1.96	1.37E+04	2.56

Photo de Base de données à partir de molécule 88 jusqu'a molécule 116

117	Butyl Ether	142-96-1	C ₈ H ₁₈ O	130.23	-0.3	9.47E+02	3.11	
118	Bromobenzene	108-86-1	C ₆ H ₅ Br	157.02	0.614	5.49E+02	2.99	
119	Borneol	507-70-0	C ₁₀ H ₁₈ O	154.25	-2.604	5.31E+00	2.71	
120	Propylbenzene	103-65-1	C ₉ H ₁₂	120.1916	1.158	4.12E+02	3.74	
121	Dodecanoic acid	143-07-7	C ₁₂ H ₂₄ O ₂	200.32	0.222	8.81E-02	5.03	
122	Acetanilide	103-84-4	C ₉ H ₉ NH(COC	154.25	2.431	1.16E-01	1.08	
123	Butylbenzene	68411-44-9	C ₁₀ H ₁₄	134.22	-1.237	1.40E+02	4.27	
124	2,6-Dichloroanisole	1984-65-2	C ₇ H ₆ Cl ₂ O	177.03	-2.398	2.51E+01	3.24	
125	2,3-Dimethylphenol	526-75-0	C ₈ H ₁₀ O	122.17	-3.301	1.24E+01	2.4	
126	2,4-Dimethylphenol	105-67-9	C ₈ H ₁₀ O	122.17	-1.9	1.72E+01	2.4	
127	2,5-Dimethylphenol	95-87-4	C ₈ H ₁₀ O	122.17	-0.879	1.71E+01	2.4	
128	Isobutyl Isobutyrate	97-85-8	C ₈ H ₁₆ O ₂	144.22	1.255	6.80E+02	2.47	
129	4-Methyl-2-pentanol	108-11-2	C ₈ H ₁₄ O	102.18	0.146	4.91E+02	1.57	
130	1-Penten-3-ol	616-25-1	C ₅ H ₁₀ O	86.13	0.227	1.49E+03	1.05	
131	Methyl Cinnamate	103-26-4	C ₁₀ H ₁₀ O ₂	162.19	-0.569	1.49E+00	2.18	
132	Thiophene	110-02-1	C ₄ H ₄ S	84.14	0.085	1.09E+04	1.95	
133	Tribromomethane	75-25-2	CHBr ₃	252.75	0.5	6.89E+02	2.29	
134	1,3,5-Trichlorobenzene	108-70-3	C ₆ H ₃ Cl ₃	181.45	0.792	3.56E+01	4.04	

Photo de Base de données à partir de molécule 117 jusqu'a molécule 134

Résumé

Dans ce travail, nous avons étudié le degré de corrélation entre le coefficient de partage octanol/eau ($\log K_{ow}$), pression de vapeur (P_v) et le seuil de détection d'odeur ($\log ODT$) avec deux types de descripteurs moléculaires (**3D**, **2D**). Pour cette raison on a utilisé la technique de **Q.S.P.R.** on applique la régression linéaire par **XLSTAT** comme outil pour faciliter la prédiction avec 2 blocs de descripteurs qui sont (**3D-MoRSE**, **topologiques**).

La base de données qui on utilise contient **134** molécules odorant. Après les calculs et l'analyse des résultats, nous avons trouvé que la corrélation est remarquable dans le bloc **3D-MoRS**, tel que ($R^2=0.973$), et la corrélation suffisant avec bloc de descripteur topologique ($R^2= 0.584$).

Les mots clés: descripteurs, odeur, descripteurs moléculaires, QSPR, prédiction.

Abstract:

In this work, we have study the degree of correlation between of partition the coefficient octanol/ water ($\log K_{ow}$) and vapor pressure (P_v) and the threshold of detection of odor ($\log ODT$) with two types of the molecular descriptors (**3D**, **2D**). For this reason we used the méthode of **Q.S.P.R.** and applied the linear regression by **XLSTAT** as tool for facility the prediction with 2 blocks of descriptors which are (**3D-MoRSE**, **topological**).

The data base which we use contains **134** molecules odorous. After calculations and the analysis of the results, we have find that the correlation is remarkable in the block **3D-MoRSE**, such as $R^2 = 0.973$, and the correlation sufficient with topological block of descriptor ($R^2 = 0.584$).

Key words: descriptors, odor, molecular correlation, QSPR, prediction.



Remerciements

En premier lieu, je tiens à remercier notre Dieu, qui m'a donné la force pour accomplir ce travail.

J'adresse mes vifs remerciements à mon promoteur consultant, Dr. KORICHI Mourad pour m'avoir dirigé tout au long de ce travail.

Je remercie également Ms: TEKHA Mebarqa, pour ses explications, remarques judicieuses et conseils qui m'ont été précieux pour la réalisation de ce travail.

Je présente mes chaleureux remerciements aux membres de jury qui ont consulté mon travail et aussi aux enseignants pour leur aides et orientation durant mes études.

Je remercie également tous les membres qui m'ont beaucoup aidé à réaliser ce travail dans de bonnes conditions.

Mes derniers remerciements et ce ne sont pas les moindres, vont à tous ceux qui ont participé de près ou de loin pour l'aboutissement de ce travail.



Liste des figures

Titre	page
Figure I-1: représentation le nez (HARLE, 2009).	03
Figure I-2: présentation de polarité des composés odoriférants.	03
Figure I-3: Énantiomères de carvone.	04
Figure I-4: Formule semi développée de l'acétate d'isoamyle.	04
Figure I-5 : Explication de loi de Steven.	05
Figure I-6: Représentation schématique des projections du système olfactif dans le système nerveux central.	07
Figure I-7 : La voie principale de transduction chimio-électrique olfactive.	08
Figure I-8: Représentation schématique d'un récepteur couplé aux protéines G.	08
Figure I-9: Schéma du principe d'un olfactomètre dynamique.	10
Figure II-1: principe de la méthode QSPR.	16
Figure II-2: Représentation des blocs des descripteurs moléculaires.	19
Figure II-3: Diagramme des descripteurs utilisés.	20
Figure II-4-Diagramme de prédiction par QSPR.	21
Figure II-5-montre le cycle d'optimisation du modèle.	22
Figure II-6-Diagramme de notre travail.	22
Figure II-7 : applique le Field Force à Composé : limonène ; Formule brute $C_{10}H_{16}$; CAS 7721-11-124	22
Figure II-8: Montre la meilleure droite de régression linéaire Simple.	30
Figure II-9: Montre la réalisation de valeur $x_{i i}$ et la droite de régression.	30
Figure II-10 : Schéma représente logiciel XLSTAT.	32
Figure II-11 : Boite dialogue de régression linéaire.	33
Figure III-1-représentation de $\log K_{ow}$ en fonction des coefficients normalisés (b2).	36
Figure III-2-représentation de $\text{préd}(\log K_{ow})$ en fonction des résidus normalisés (b2).	36
Figure III-3-représentation de $\log K_{ow}$ en fonction des résidus normalisés (b2).	37
Figure III-4-représentation de $\text{préd}(\log K_{ow})$ en fonction de $K_{ow}(b2)$.	37
Figure III-5-représentation de $\log P_v$ en fonction des coefficients normalisés (b2).	40
Figure III-6-représentation de $\text{préd}(\log P_v)$ en fonction des résidus normalisés (b2).	40
Figure III-7-représentation de $\log P_v$ en fonction des résidus normalisés (b2).	40
Figure III-8-représentation de $\text{préd}(\log P_v)$ en fonction de $P_v(b2)$.	40

Figure III-9-représentation de préd(log ODT) en fonction des résidus normalisés (b2).	42
Figure III-10-représentation de log ODT en fonction des résidus normalisés (b2).	42
Figure III-11-représentation de log ODT en fonction des coefficients normalisés (b2).	42
Figure III-12-représentation de préd(log ODT) en fonction de log ODT (b2).	42
Figure III-13-représentation de log ODT en fonction des coefficients normalisés (b14).	45
Figure III-14-représentation de préd(log ODT) en fonction de log ODT (b14).	45
Figure III-15-représentation de log ODT en fonction des résidus normalisés (b14).	45
Figure III-16-représentation de préd(logODT) en fonction des résidus normalisés (b14).	45
Figure III-17-représentation de log K_{ow} en fonction des coefficients normalisés (b14).	48
Figure III-18-représentation de log K_{ow} en fonction des résidus normalisés (b14).	48
Figure III-19-représentation de préd(log K_{ow}) en fonction des résidus normalisés (b14).	48
Figure III-20-représentation de préd(log K_{ow}) en fonction de K_{ow} (b14).	48
Figure III-21-représentation de log P_v en fonction des coefficients normalisés (b14).	51
Figure III-22-représentation de préd(log P_v) en fonction de P_v (b14).	51
Figure III-23-représentation de log P_v en fonction des résidus normalisés (b14).	51
Figure III-24-représentation de préd(log P_v) en fonction des résidus normalisés (b14).	51

Liste des tableaux

Titre	page
Tableau II-1 : différents types d'apprentissages.	17
Tableau II-2 : Les paramètres topologiques après le calcul par logiciel DRAGON.	18
Tableau II-3: Les paramètres 2D auto corrélation. Après le calcul par logiciel DRAGON.	18
Tableau II-4 : les blocks des descripteurs calculés par logiciel DRAGON.	26
Tableau II-5 : Les descripteurs moléculaires après la réduction.	26
Tableau II-6: Coefficient d'ajustement	33
Tableau II-7: Analyse de la variance.	34
Tableau III-1 : Coefficient d'ajustement (corrélation) des propriétés étudiées.	35
Tableau III-2 : Coefficient d'ajustement ($\log K_{ow}$, block2).	36
Tableau III-3 : Analyse de variance ($\log K_{ow}$, block2).	36
Tableau III-4 : Paramètre de modèle ($\log K_{ow}$, block2).	38
Tableau III-5 : Coefficient d'ajustement (P_V , block2).	39
Tableau III-6 : Analyse de variance (P_V , block2).	39
Tableau III-7 : Paramètre de modèle (P_V , block2).	40
Tableau III-8 : Coefficient d'ajustement ($\log ODT$, block2).	42
Tableau III-9 : Analyse de variance ($\log ODT$, block2).	42
Tableau III-10 : Paramètre de modèle ($\log ODT$, block2).	43
Tableau III-11 : Coefficient d'ajustement ($\log ODT$, block14).	45
Tableau III-12 : Analyse de variance ($\log ODT$, block14).	45
Tableau III-13 : Paramètre de modèle ($\log ODT$, block14).	46
Tableau III-14 : Coefficient d'ajustement ($\log K_{ow}$, block14).	48
Tableau III-15 : Analyse de variance ($\log K_{ow}$, block14).	48
Tableau III-16 : Paramètre de modèle ($\log K_{ow}$, block14).	49
Tableau III-17 : Coefficient d'ajustement (P_V , block14).	51
Tableau III-18 : Analyse de variance (P_V , block14).	51
Tableau III-19 : Paramètre de modèle (P_V , block14).	52

Nomenclature

Symboles	Définitions
K_{AE}	Coefficient de partage air- eau
K_{OE}	Coefficient de partage octanol - eau
$C_{I,E}$	Solubilité dans l'eau
P_V	Pression de vapeur
ODT	Seuil détection d'odeur
CAS	Services Abstracts Chimiques
QSPR	Relation structure-propriété quantitative
DDL	Le nombre de degrés de liberté pour le modèle retenu (correspondant à la partie erreurs)
MCE	La moyenne des carrés des erreurs
RMCE	La racine de la moyenne des carrés des erreurs (RMCE) est la racine carrée de la MCE.
t	La tolérance
F	Coefficient de Fisher
AAD	La déviation absolue moyenne
R^2	Coefficient de détermination
AAE	Erreur moyenne absolue
RMSE	La racine carrée de l'erreur moyenne
AAEP	Erreur moyenne absolue prédire
MAPE	Pourcentage d'Erreur Absolue Moyenne
DW	Le coefficient de Durbin-Watson
C_p	Le coefficient C_p de Mallows
AIC	Le critère d'information d' Akaike (<i>Akaike's Information Criterion</i>).
SBC	Le critère Bayésien de Schwarz (<i>Schwarz's Bayesian Criterion</i>)
PC	Le critère de prédiction d' Amemiya (<i>Amemiya's Prediction Criterion</i>)
SCE	La somme des carrés des erreurs (ou résidus) du modèle
W_i	Le poids (valeur ou coefficient) des variables
OUE	Unité Odeur Européenne

SOMMAIRE

Dédicace

Remerciements

Liste des figures

Liste des tableaux

Nomenclature

Sommaire

Introduction général 01

CHAPITRE I: GENERALITES SUR L'ODEUR

I-1- Introduction	02
I-2- Définition d'une odeur	02
I-3- l'olfaction	02
I-4- Mécanismes de l'olfaction	03
I -5- Les propriétés physico-chimiques des composés odoriférants:	03
I -5-1-Odeurs et propriétés physiques	04
I -6- La perception des odeurs	05
I -6-1- Seuil de détection olfactif et seuil de reconnaissance olfactif	05
I -6-2-Sensibilité olfactive et seuils de perception	06
I-6-3- Solubilité des odeurs	06
I-6-4-Volatilité des odeurs	07
I -7-Structures moléculaires odoriférantes	07
I -8-Les sources d'odeurs	07
I -10-La transduction du message olfactif	08
I -9-Système olfactif	08
I -11-Récepteur d'odeur	08
I -12-Méthodes des mesures d'odeur	09
I -12-1-L'analyse physico-chimique	09
I -12-1-L'analyse olfactométrique	09
I-13-Les différentes grandeurs mesurées en olfactométrie et les modes opératoires	10
I -13-1-Mesure du facteur de dilution au seuil de perception	10
I -13-2-Mesure de l'intensité d'odeur	10

I -13-3-L'olfactométrie	11
I-14-Le jury de nez	11
I -13-4-Grandeurs olfactométriques (quelques définitions)	12
I-15-Les nez artificiels	13
I-16-Les nuisances	13
I-17-Conclusion	14

CHAPITRE II: LES OUTILS ET LES TECHNIQUES UTILISE

II-1-Introduction	15
II -2-Quantitative Structure-Property Relationships	15
II -2-1-Définition de QSPR	15
II-2-2-Principe de QSPR	15
II-2-3-Les méthodes mathématiques utilisés par le model QSPR	16
II-2-4-Les types d'apprentissages d'un modèle QSPR	17
II-2-5-Techniques d'apprentissages utilisés on chimie	17
II-3-Descripteurs moléculaires	17
II-3-1-Définition d'un descripteur	17
II-3-2-Bloc des descripteurs	17
II-3-3-Exemple: calcul ATSKw	19
II-4-Importance des descripteurs	20
II-4-1- Les descripteurs utilisés	20
II-5-Diagramme montre les étapes de prédiction	20
II-6-L'objectif de la prédiction	21
II-7-Les étapes de travail: (Diagramme de travail)	22
II-8-Base des données	22
II-8-1-Les caractéristiques et les propriétés de cette base de données sont	23
II-8-2-La raison du choix de cette base des données	23
II-8-3 -Préparation de base des données	23
II-8-3-1-Logiciels « Hyperchem 7.5 »	23
II-8-3-2-Récupération et stabilisation les molécules de fichier Hin	24
II-8-3-2-1 Stabilisation structure des molécules (minimisation de l'énergie)	24

II-8-3-2-2 Mécanique moléculaire	24
II-8-3-2-3 Récupération des fichiers HyperChem HIN	25
II-9-Calcul des descripteurs moléculaires	25
II-9-1-Le Logiciel « DRAGON»	25
II-9-2-Groupe des descripteurs moléculaires	25
II-10- Logiciel « XLSTAT»	27
II-10-1-Applications	27
II-11-Techniques des régressions	27
II-11-1-La régression Linéaire Simple (RLS)	28
II-11-2-Coefficients d'ajustement	28
II-11-3- Description de régression linéaire simple	29
II-11-4- Les étapes de régression	31
II-11-4-1 modèle linéaire	32
II-11-5-Tableaux Résultants	33
II-11-6- Graphes résultants	34
II-12- Conclusion	34
CHPITRE III: RESULTATS ET DISCUSSION	
III-1-Introduction	35
III-2-Résultats de régression	35
III-2-1-Coefficient de corrélations	35
III-3- Block de descripteurs 02: Descripteurs topologiques	36
III-3-1-Etude de propriété de $\log K_{OW}$	36
III-3-1-2-Analyse des résidus	36
III-3-1-3-Discussion	38
III-3-2-Etude de propriété de P_v	39
III-3-2-1-Analyse des résultats	41
III-3-2-2-Analyse des résidus	41
III-3-2-3-Discussion	41
III-3-3-Etude de propriété de $\log (ODT)$	42
III-3-3-1-Analyse des résultats	43

III-3-3-2-Analyse des résidus	43
III-3-3-3-Discussion	44
III-4-Block de descripteurs 14: Descripteurs 3D-MoRSE	45
III-4-1-Etude de propriété de log ODT	45
III-4-1-1-Analyse des résultats	46
III-4-1-2-Analyse des résidus	46
III-4-1-3-Discussion	47
III-4-2-Etude de propriété de log K_{ow}	48
III-4-2-1-Analyse des résultats	49
III-4-2-2-Analyse des résidus	49
III-4-2-3-Discussion	50
III-4-3-Etude de propriété de log P_v	51
III-4-3-1-Analyse des résultats	53
III-4-3-2-Analyse des résidus	53
III-4-3-3-Discussion	53
CONCLUSION GENERAL	54
BIBLIOGRAPHIE	55
ANNEXES	
Résumé	

INTRODUCTION GENERAL

INTRODUCTION GENERAL

INTRODUCTION GENERAL

La modélisation moléculaire implique l'utilisation de méthodes de calcul théoriques (Mécanique moléculaire, dynamique moléculaire) permettant d'évaluer les propriétés physico-chimiques de la molécule odoriférante qui est étudiée.

L'importance de la modélisation moléculaire dans tous les domaines, nous conduit à utiliser leurs méthodes pour expliquer un phénomène très intéressant biologiquement, ou chimiquement dans notre travail c'est la relation structure - propriétés de quelques molécules odorantes.

Les différentes techniques de développement moléculaire peuvent aider considérablement les chercheurs pour trouver des relations entre la structure chimique et propriété de composés organiques.

Les relations structure-propriétés quantitatives "QSPR" ont été utilisées dans cette étude pour relier les paramètres structuraux d'une molécule (les descripteurs moléculaires, géométriques, topologique, etc.) à ses propriétés physico-chimiques.

On utilise la régression linéaire par XLSTAT comme outil pour faciliter les procédures de prédiction.

Notre travail est divisé en trois chapitres :

Dans le premier chapitre on a un inventaire non exhaustif des connaissances en matière d'odeurs, de composés odorants et d'olfaction sera tout d'abord présenté, afin de bien comprendre le cheminement et le traitement de l'information olfactive.

Dans le deuxième chapitre nous présenterons les outils et les techniques qui sont utilisés dans la prédiction de modèle.

Troisième chapitre : nous avons présenté les résultats suivis par discussion et enfin conclusion générale qui comporte les recommandations pour améliorer les résultats.