

## RELATION STRUCTURE/SOLUBILITE AQUEUSE DE HAP PRIORITAIRES

Reçu le 01/09/2008 – Accepté le 23/06/2009

### Résumé

Une relation structure/solubilité aqueuse des HAP retenus comme prioritaires par l'agence environnementale américaine (US-EPA) a été recherchée en favorisant l'approche hybride algorithme génétique/régression multilinéaire, les paramètres structuraux étant calculés avec le logiciel DRAGON. Parmi la centaine de modèles à 2 régresseurs obtenus nous avons sélectionné celui qui présente les meilleures valeurs du paramètre de prédiction ( $Q^2$ ) et du coefficient de détermination ( $R^2$ ):

$\log S = 23,624 - 8,6316 \text{ VEA1} - 11,145 \text{ Kp}$  ;

$n=15$  ;  $ES=0,464$  ;  $Q^2(\%)=98,38$  ;  $R^2(\%)=98,90$  ;  $F=544,19$  ;  $P=0,000$

La première variable explicative (VEA1) est associée aux propriétés locales d'une molécule, alors que la seconde reflète sa forme globale.

**Mots clés:** structure/propriété, hydrocarbures aromatiques polycycliques, descripteurs moléculaires

### Abstract

A structure/aqueous solubility relationship was searched for PAHs considered as important by the US Environmental Protection Agency, while promoting the hybrid genetic algorithm/multilinear regression approach, the structural parameters being calculated with the software DRAGON. Among about a hundred of 2 regressor models gotten, we selected the one that present best values of the prediction parameter ( $Q^2$ ) and of the determination coefficient ( $R^2$ ).

$\log S = 23.624 - 8.6316 \text{ VEA1} - 11.145 \text{ Kp}$  ;

$n=15$  ;  $SD=0.464$  ;  $Q^2(\%)=98.38$  ;  $R^2(\%)=98.90$  ;  $F=544.19$  ;  $P=0.000$

The first explanatory variable (VEA1) is related to the local properties of a molecule, whereas the second reflects its global shape.

**Keywords:** structure/property ; polycyclic aromatic hydrocarbons; molecular descriptors.

I.TOUHAMI  
D.MESSADI

<sup>1</sup>Laboratoire de Sécurité  
Environnementale et Alimentaire  
(LASEA)  
Université BADJI Mokhtar BP12 –  
Annaba – 23000  
ALGERIE

### ملخص

العلاقة بنية / د و بانية مانية بالنسبة للهيدروكربونات العطرية المتعددة الحلقة المصنفة بالأولوية حسب الوكالة الأمريكية للبيئة (US.EPA) التي تم تحقيقها بتفضيل برنامج هجين وراثي / التراجع الخطي المتعدد. المقاييس البنيوية تم حسابها بالبرنامج DRAGON.

من بين مئة نموذج لمتغيرين تم اختيار أفضل مقياس دو أحسن قيم المتنبأ  $Q^2$  وأحسن معامل التحديد :  $R^2$  .

$\log S = 23.624 - 8.6316 \text{ VEA1} - 11.145 \text{ Kp}$  ;

$n=15$  ;  $SD=0.464$  ;  $Q^2(\%)=98.38$  ;  $R^2(\%)=98.90$  ;  $F=544.19$  ;  $P=0.000$

المتغير التوضيحي الأول (Kp) مرتبط بالخواص المحلية للجزئي. أما الثاني (VEA1) يعكس شكله الشامل.

**الكلمات المفتاحية:** خواص , هيدروكربونات العطرية المتعددة الحلقة , مؤشرات بنيوية معتمدة على القيم الخاصة الهوليسنتكية الموزونة.

I – INTRODUCTION

Les hydrocarbures aromatiques polycycliques (HAP) sont des molécules organiques issues de la combustion incomplète des matières carbonées suite à des processus naturels, mineurs, et des processus anthropiques, majoritaires. Les HAP sont constitués d’atomes de carbone et d’hydrogène formant au moins deux anneaux aromatiques condensés. Ils sont libérés dans tous les compartiments de l’environnement.

Le nombre de HAP identifiés à ce jour est de l’ordre de 130 [1]. Parmi ceux-ci, une liste restreinte est généralement considérée pour les études environnementales. Ce sont les 16 polluants retenus comme prioritaires par l’agence environnementale américaine (US-EPA) [2]. Elle contient les HAP les plus étudiés et posant des problèmes environnementaux majeurs du fait de leur toxicité.

Le transport et le devenir des HAP dans l’environnement dépendent, en partie, de leurs propriétés physicochimiques, comme la solubilité dans l’eau, et de leurs répartitions relatives entre les différentes phases environnementales.

La solubilité dans l’eau de HAP a pu être reliée à la température de fusion, en association avec le volume moléculaire et l’aire de la surface de van der Waals qui le délimite [3]. Des modèles à 2 régresseurs ont ainsi été établis. Ils nécessitent, cependant, la connaissance de la température de fusion, un paramètre expérimental aussi difficile à évaluer théoriquement que la solubilité elle-même.

Dans ce travail nous avons relié la solubilité dans l’eau (logS) des HAP prioritaires, à des descripteurs moléculaires calculés avec le logiciel DRAGON [4], en favorisant l’approche hybride combinant algorithme génétique, pour le choix de sous-ensembles de descripteurs significatifs, et régression linéaire multiple.

La qualité de l’ajustement et la robustesse du modèle ont été vérifiées par validation interne.

II. PARTIE THEORIQUE

II-1 La régression linéaire multiple

Supposons qu’on ait mesuré sur n individus (k+1) variables représentées par des vecteurs de  $R^n$ :  $\tilde{y}$ ,  $\tilde{x}_1$ ,  $\tilde{x}_2, \dots, \tilde{x}_k$ ;  $\tilde{y}$  est la variable dépendante ou à expliquer (logS), les  $\tilde{x}_j$  les variables explicatives ou prédictes (descripteurs moléculaires). On cherche alors à reconstruire  $\tilde{y}$  au moyen des  $\tilde{x}_j$  par une formule linéaire.

On pose :

$$\tilde{y} = \beta_0 \tilde{1} + \sum_{(j)} \tilde{x}_{(j)} \beta_{(j)} + \varepsilon_j \tag{1}$$

$\tilde{y}$  est un vecteur de dimension n contenant la solubilité des HAP étudiés,  $\tilde{1}$  est un vecteur unité, c'est-à-dire une matrice colonne formée d’éléments égaux à 1,  $\tilde{X}_{(j)}$  indique la matrice (n x j), et  $\varepsilon_j$  correspond aux erreurs (résidus) qui doivent suivre une loi Normale, posséder une moyenne nulle, et une matrice de dispersion  $I \sigma^2$  [5]. Les estimateurs  $\{\beta\}$  sont calculés en utilisant la technique des moindres carrés ordinaires.

II-2 Paramètres d’évaluation de la qualité de l’ajustement

Deux paramètres sont couramment utilisés :

- Le coefficient de détermination multiple :

$$R^2 = 1 - \frac{SCE}{SCT} = 1 - \frac{\sum_1^n (y_i - \hat{y}_i)^2}{\sum_1^n (y_i - \bar{y})^2} \tag{2}$$

où  $\hat{y}_i$  est la valeur estimée de logS, et  $\bar{y}$  la valeur moyenne des valeurs observées.

- La racine de l’erreur quadratique moyenne de prédiction :

$$\sigma_n = \sqrt{\frac{1}{n} \sum_1^n (y_i - \hat{y}_{(i)})^2} \tag{3}$$

II-3 Robustesse du modèle

La stabilité du modèle a été explorée en utilisant la validation croisée par omission d’une observation (LOO pour : cross-validation by leave-one-out) [6]. Elle consiste à recalculer le modèle sur (n-1) HAP, l’équation obtenue servant alors à estimer la solubilité du composé éliminé, notée  $\hat{y}_{(i)}$ . On répète le procédé pour chacun des n HAP.

La somme des carrés des erreurs de prédiction, désignée par l’acronyme PRESS (pour : Predictive Residual Sum of Squares)

$$PRESS = \sum_1^n (y_i - \hat{y}_{(i)})^2 \tag{4}$$

est une mesure de la dispersion de ces estimations. On l’utilise pour définir le coefficient de prédiction :

$$Q_{LOO}^2 = \frac{SCT - PRESS}{SCT} \tag{5}$$

Contrairement à  $R^2$ , qui augmente avec le nombre de paramètres p de la régression,  $Q_{LOO}^2$  affiche une courbe avec maximum, ou avec palier, obtenu pour un certain nombre de paramètres, puis décroît par la suite de façon

monotone. Ce fait confère une grande importance au coefficient  $Q_{LOO}^2$ . Une valeur de  $Q_{LOO}^2 > 0,5$  est, généralement, considérée comme satisfaisante, et une valeur supérieure à 0,9 est excellente [7].

#### II-4 Détection des observations aberrantes

Elle a été basée sur la non-satisfaction à trois au moins (pour n'en privilégier aucun) des six tests statistiques couramment utilisés pour la détection de telles observations en analyse de régression [5,6]:

1/ Les résidus ordinaires  $e_i$ , différences entre les valeurs observées ( $y_i$ ) et estimées par le modèle ( $\hat{y}_i$ ). 2/ Les résidus studentisés internes  $r_i$ , obtenus en divisant  $e_i$  par  $s\sqrt{1-h_{ii}}$ . 3/ Les résidus studentisés externes, pour lesquels l'estimation  $S_{(i)}^2$  de  $\sigma^2$  est calculée pour (n-1) observations, la  $i^{ème}$  étant exclue ( $t_i = e_i/S_{(i)}\sqrt{1-h_{ii}}$ ). 4/ Les leviers  $h_{ii}$ , éléments diagonaux de la matrice ( $\tilde{H}$ ) de passage du vecteur  $y$  au vecteur  $\hat{y}$ , permettent de juger de l'influence d'une observation  $i$  dans la détermination de l'équation de régression (lorsque  $h_{ii} > 3p/n$ ). 5/ la statistique représentée par DFITS permet de mesurer l'influence d'une observation  $i$  sur la valeur ajustée. Belsley, Kuh et Welsch [8] considèrent qu'une observation pour laquelle  $DFITS > 2\sqrt{p/n}$  est inhabituelle. 6/ La distance de Cook  $D_i$ , permet d'étudier l'influence d'une observation  $i$  sur les coefficients de régression estimés par les moindres carrés. Cook [9] et Weisberg [10] suggèrent de comparer  $D_i$  au paramètre de Fisher  $F(.5, p, n-p)$ . Comme  $F(.5, p, n-p) \approx 1$ , on considère que les observations pour lesquelles  $D_i > 1$  sont influentes.

#### II-5 Test de randomisation [6]

Ce test permet de mettre en évidence des corrélations dues au hasard. Il consiste à générer un vecteur solubilité par permutation aléatoire des composantes du vecteur réel. On calcule alors sur le vecteur obtenu (considéré comme vecteur expérimental réel) un modèle, selon la méthode habituelle. Ce procédé est répété plusieurs fois (100 dans notre cas).

### III. COLLECTE DES DONNEES ET METHODOLOGIE

Les solubilités des HAP prioritaires (exceptée celle du Fluorène) rapportées dans [11], ont été utilisées pour l'analyse des relations structure/solubilité aqueuse.

Les structures des 15 molécules ont été obtenues à l'aide du logiciel Hyperchem Pro 6.03 [12], et les géométries finales à l'aide de la méthode semi-empirique AM1 du même logiciel. Les structures moléculaires ont été optimisées à l'aide de l'algorithme Polak-Ribiere. Les

géométries obtenues ont été transférées dans le logiciel DRAGON pour le calcul de plus de 1600 descripteurs appartenant à 20 classes différentes.

La sélection des variables explicatives, a été réalisée par algorithme génétique, dans la version MOBYDIGS [13], en maximisant  $Q_{LOO}^2$ . Les principaux paramètres ont été fixés comme suit : taille de la population du modèle  $P=100$  ; nombre maximal de variables,  $L= 2$  ; probabilité de croisement,  $pc > 0,9$  ; probabilité de mutation,  $pM < 0,1$ .

### IV. RESULTATS ET DISCUSSION

La solubilité dans l'eau des HAP a été reliée à des paramètres structuraux (Tableau I) qu'on peut rattacher aux indices basés sur les valeurs propres (VEA1) et aux indices WHIM (Kp) [14].

**Tableau 1** : Variables dépendante et explicatives des HAP prioritaires.

n°	Nom	logS	VEA1	Kp
1	Naphtalène	- 8.3061	3.048	0.526
2	Acenaphtylène	-10.5651	3.307	0.500
3	Acenaphtène	-10.7154	3.307	0.481
-	Fluorène	-	3.363	0.623
4	Phénanthrène	-12.0910	3.363	0.609
5	Anthracène	-15.1938	3.542	0.692
6	Fluoranthène	-13.7957	3.685	0.500
7	Pyrène	-14.2417	3.773	0.500
8	Benzo[a]anthracène	-17.5044	3.947	0.692
9	Chrysène	-18.6577	3.933	0.701
10	Benzo[b]fluoranthène	-16.8053	4.041	0.573
11	Benzo[k]fluoranthène	-17.9889	4.024	0.664
12	Benzo[a]pyrène	-19.1630	4.134	0.621
13	Dibenzo[ah]anthracène	-22.4381	4.338	0.757
14	Benzo[ghi]perylène	-19.7911	4.346	0.500
15	Indeno[1,2,3,c,d]pyrène	-20.0709	4.294	0.565

Le premier descripteur est associé aux propriétés locales d'une molécule, alors que le second est un paramètre de forme globale, l'association de ces deux caractéristiques pouvant expliquer les différences des solubilités aqueuses des HAP considérés.

L'indice VEA1, reflète les différences intramoléculaires. Il est égal à la somme des coefficients  $I_{iA}$  du vecteur propre associé à la valeur propre la plus négative de la matrice d'adjacence.

$$VEA1 = \sum_{i=1}^A I_{iA} \quad (6)$$

## RELATION STRUCTURE/SOLUBILITE AQUEUSE DE HAP PRIORITAIRES

Les  $li_A$  constituent des descripteurs des sommets du graphe, au nombre de A, utilisés pour caractériser l'invariance des propriétés locales dans une molécule [14].

Les descripteurs moléculaires à invariant holistique pondéré (WHIM) constituent une classe de descripteurs moléculaires géométriques. Ils permettent de saisir dans le détail les informations relatives à la taille, la forme, la symétrie et la distribution des atomes d'une molécule par rapport à des cadres de référence fixes. Le calcul des descripteurs WHIM repose sur l'analyse en composantes principales de la matrice de covariance des coordonnées atomiques pondérées, dont les éléments sont définis par :

$$S_{jk} = \frac{\sum_{i=1}^n w_i (q_{ij} - \bar{q}_j)(q_{ik} - \bar{q}_k)}{\sum_{i=1}^n w_i} \quad (7)$$

où n représente le nombre d'atomes de la molécule,  $w_i$  le poids du ième atome,  $q_{ij}$  la jème coordonnée cartésienne de l'atome i ( $j = 1, 2, 3$ ), alors que  $\bar{q}_j$  est la moyenne de cette jème coordonnée.

Six modes de pondération sont proposés, et selon le mode adopté on obtient différentes matrices de covariance et différents axes principaux pour la molécule.

On distingue les descripteurs WHIM dirigés, calculés individuellement selon les directions des composantes principales, et les descripteurs WHIM non dirigés, ou globaux, calculés pour la molécule entière à partir des combinaisons des premiers [14,15].

L'indice de forme globale K est un descripteur WHIM non dirigé, plus général que le facteur acentrique qu'il remplace avantageusement [16]. Il est défini par la relation :

$$K = \frac{\left| \frac{\lambda_m}{\sum_m \lambda_m} - \frac{1}{3} \right|}{\frac{4}{3}} \quad m = 1, 2, 3 ; 0 \leq K \leq 1 \quad (8)$$

$\lambda_1, \lambda_2, \lambda_3$  étant les valeurs propres de la matrice de covariance des coordonnées atomiques pondérées.

Le modèle retenu a pour équation :

$$\log S = 23,624 (\pm 1,208) - 8,6316 (\pm 0,3234) \text{ VEA1} - 11,145 (\pm 1,469) \text{ Kp} \quad (9)$$

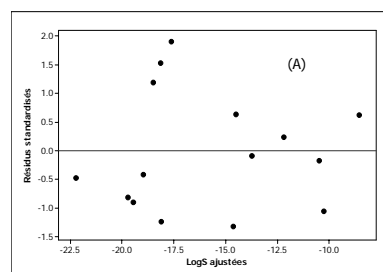
$n = 15$  ;  $\sigma_n = 0,506$  ;  $R^2(\%) = 98,9$  ;  $Q2(\%) = 98,38$  ;  $F = 544,19$  ;  $p = 0,000$

Les valeurs des paramètres statistiques montrent que les 2 descripteurs VEA1 et Kp permettent de corréler les solubilités aqueuses des 15 HAP. En effet, la valeur du coefficient de détermination signifie que 98,9% de la

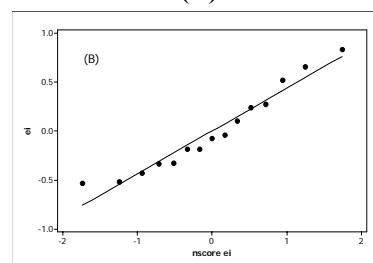
variabilité de  $\log S$ , peut être expliquée par ces 2 descripteurs. La grande valeur du paramètre de Fisher ( $F = 544,19$  ;  $p = 0,000$ ) indique un modèle très hautement significatif.

La figure 1-A reproduit le graphe des résidus standardisés  $d_i (= e_i / \sqrt{\text{CME}}$  ; CME étant le carré moyen des écarts) en fonction des valeurs ajustées  $\log S$  sans tendance particulière, ce qui montre la constance de la variance  $\sigma^2$ , c'est-à-dire son indépendance des régresseurs et de la variable dépendante ajustée ; notons également que tous les résidus standardisés sont compris entre les limites  $\pm 2$ . La quasi-linéarité ( $R^2 = 95,0\%$  ;  $R_{critique}^2 = 93,83\%$ )

du diagramme des scores normaux (figure 1-B) est un indice de normalité. La valeur de la statistique de Durbin-Watson ( $d = 2,03$ ) [16] est plus grande que la valeur supérieure donnée par les tables pour 2 régresseurs et pour tout risque raisonnable  $\alpha$ , ce qui établit l'indépendance des résidus.



(A)



(B)

Figure 1 . (A) .Graphe des résidus standardisés en fonction des valeurs ajustées.

(B) : Diagramme des scores normaux.

La colonne  $e_i$  du tableau II montre des résidus ordinaires strictement inférieurs en valeur absolue à 2 fois l'erreur standard ( $e_i < 2ES = 0,9270$ ). Comme  $n = 15$  et  $p = 3$ , on constate que tous les résidus studentisés internes  $r_i$  sont inférieurs en valeur absolue à  $t(0,025 ; n-p) [= 2,1793]$  qui est le 0,975 quantile d'une loi de Student avec  $(n-p)$  degrés de liberté ; pareillement, tous les résidus studentisés externes  $t_i$ , qui sont du même ordre de grandeur que les  $r_i$  correspondants, restent inférieurs en valeur absolue à  $t(0,025 ; n-p-1) [= 2,20]$ .

Ainsi, l'analyse des résidus caractéristiques ne permet pas de détecter d'observation aberrante. La colonne HI donne des termes diagonaux,  $h_{ii}$ , de la matrice de projection, qui sont toutes inférieures à la valeur critique

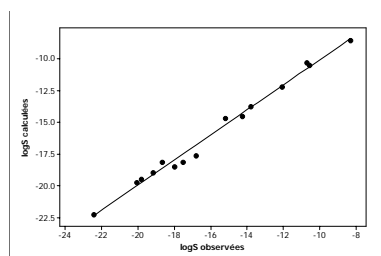
$3p/n = 0,6$ . Comme par ailleurs les valeurs absolues de DFITS sont inférieures à la valeur critique  $2\sqrt{p/n} = 0,89443$ , on peut conclure à l'absence d'observation présentant un levier important, ainsi qu'à l'absence d'observation influençant la détermination des paramètres du modèle ou de la valeur prédite.

**Tableau 2:** Résidus caractéristiques et valeurs ajustées de logS. La numérotation est celle du Tableau I.

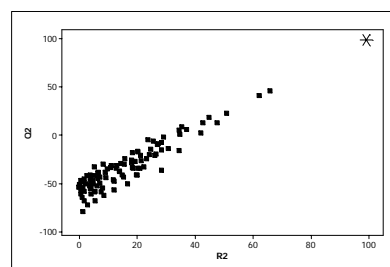
(i)	$e_i$	$r_i$	$t_i$	HI	$D_i$	DFITS <sub>i</sub>	Ajustée
1	0,2413	0,6315	0,6149	0,3044	0,0404	0,4067	-8,5474
2	-0,0718	-0,1747	-0,1674	0,1944	0,0020	-0,0823	-10,4932
3	-0,4339	-1,0700	-1,0772	0,2170	0,0828	-0,5671	-10,2815
4	0,10041	0,2416	0,2319	0,1776	0,0035	0,1078	-12,1914
5	-0,5323	-1,3374	-1,3880	0,2456	0,1465	-0,7921	-14,6615
6	-0,0397	-0,0934	-0,0895	0,1399	0,0004	-0,0361	-13,7560
7	0,2739	0,6471	0,6306	0,1471	0,0205	0,2620	-14,5156
8	0,6529	1,5476	1,6562	0,1525	0,1218	0,7027	-18,1573
9	-0,5209	-1,2478	-1,2806	0,1700	0,0882	-0,5796	-18,1368
10	0,8371	1,9309	2,2266	0,1050	0,1305	0,7626	-17,6424
11	0,5209	1,2085	1,2346	0,1150	0,0560	0,4451	-18,5099
12	-0,1829	-0,4234	-0,4084	0,1114	0,0066	-0,1446	-18,9801
13	-0,1814	-0,4829	-0,4669	0,3279	0,0255	-0,3263	-22,2567
14	-0,3297	-0,9128	-0,9060	0,3790	0,1053	-0,7077	-19,4614
15	-0,3339	-0,8211	-0,8092	0,2128	0,0478	-0,4208	-19,7370

La qualité de l'ajustement a été vérifiée en procédant à une validation croisée par 'leave-one-out'. La figure 2, qui reproduit les valeurs prédites  $\log S_{(i)}$  en fonction de celles observées  $\log S_i$  fait ressortir une faible dispersion caractéristique d'un bon ajustement, d'ailleurs confirmé par la grande valeur de  $Q^2$  ( $=98,38\%$ ).

La figure 3 qui représente le graphe des coefficients statistiques  $Q^2$  et  $R^2$  permet de comparer les résultats obtenus pour les modèles randomisés (carrés noirs) au modèle réel de départ (étoile). Il est clair que les statistiques obtenues pour les vecteurs modifiés de la solubilité sont plus petites que celles du modèle réel, et pour la majeure partie on obtient un  $Q^2 < 0$ . Ceci permet d'assurer qu'une relation structure/ solubilité réelle a été établie.



**Figure 2.** Graphe des valeurs prédites en fonction des valeurs observées.



**Figure 3.** Test de randomisation associé au modèle QSPR

## CONCLUSION

La solubilité dans l'eau des HAP prioritaires considérés a été reliée à des descripteurs moléculaires calculés à l'aide du logiciel DRAGON. L'approche hybride algorithme génétique/régression multilinéaire, qui a été favorisée, conduit à une vingtaine de modèles acceptables à 2 variables explicatives. Les meilleurs font tous intervenir l'indice structural VEA1 qui reflète les différences intramoléculaires, et un descripteur WHIM non dirigé Kw (w représentant l'une des 6 pondérations possibles) associé à la forme globale des molécules.

Nous nous sommes assurés que le modèle présenté exprime une relation structure/solubilité réelle, et qu'il vérifie les hypothèses de base d'un modèle linéaire statistique à effets fixes, sans observation aberrante et / ou influente. Enfin, la qualité de l'ajustement et la robustesse du modèle ont été vérifiées par validation interne.

## REFERENCES

- 1- INERIS- Hydrocarbures Aromatiques Polycycliques - Guide méthodologique- Rapport d'étude N°66244-DESP-R01.2005, pp85.
- 2- US EPA, Chemical Fate rate Constants for SARA Section 113 Chemicals and Superfund Health Evaluation Manual Chemicals. US Environmental Protection Agency, Office of Toxic Substances, Washington, DC 6-02-4245. <http://www.epa.gov>.
- 3- R.S.Pearlman, S.H.Yalkowsky, S.Banerjee, Water Solubilities of Polynuclear Aromatic and Heteroatomic Compounds, J.Phys.Chem.Ref.Data, 13 (02), 1984, 555-562.
- 4- R.Todeschini, V.Consonni, M.Pavan. DRAGON Software for the Calculation of Molecular Descriptors. Release 5.4 for Windows, Milano, 2006.
- 5- D.C. Montgomery, E.A.Peck, Introduction to Linear Regression Analysis, Second Edition, Wiley-Interscience Publication, New York, 1992.
- 6- N.R.Draper, H.Smith, Applied Regression Analysis, Third Edition, Wiley Series in Probability and Statistics, New York, 1998.

## RELATION STRUCTURE/SOLUBILITE AQUEUSE DE HAP PRIORITAIRES

- 7- L.Eriksson, J.Jaworska, A.P.Worth, M.T.D.Cronin, R.M.McDowell, P.Gramatica, Methods for Reliability and Uncertainty Assessment and for Applicability Evaluations of Classification- and Regression-Based QSARs, Environmental Health Perspectives, 111(10), 2003, 1361-1375.
- 8- D.A.Belsley, E.Kuh, R.E.Welsch, Regression Diagnostics : Identifying Influential Data and Sources of Collinearity, Wiley, New York, 1980.
- 9- D.Cook, Detection of Influential Observations in Linear Regression. Technometrics. 19, 1977, 15-18.
- 10- S.Weisberg, Applied Linear Regression; J.Wiley, Inc., New York, 1980.
- 11- EPA Report, EPA -600/8-90/-003, 1990.
- 12- Hyperchem<sup>TM</sup> Release 6.03 for Windows, Molecular Modelling System.2000.
- 13- R.Todeschini. MOBYDIGS Software for Multilinear Regression Analysis and Variable Subset Selection by Genetic Algorithm. Release for Windows. Milano taletè Srl, 2004.
- 14- R.Todeschini, V.Consonni, Handbook of Molecular Descriptors, R.Mannhold, H.Kubinyi, H.Timmerman eds., Wiley –VCH, Verlag G m b H, Weinheim, 2000.
- 15- R.Todeschini, M.Lasagni, E.Marenzo.New Molecular Descriptors for 2D- and 3D- Structures. Theory. J.Chemom., 8, 1994,263-273.Srl.
- 16- J. Durbin, G. S. Watson. Testing for Serial Correlation in Least Squares Regression, Biometrika58, 1971,1-19.

