

RELATION STRUCTURE / RETENTION CHROMATOGRAPHIQUE DE TREIZE ALKYLNAPHTALENES

Mohamed Lotfi Souici, Leïla Lourici et Djelloul Messadi

Laboratoire de Sécurité Environnementale et Alimentaire (LASEA), Université Badji
Mokhtar, Annaba 23 000 , Algérie
d_messadi@voila.fr

(Received 10 December 2005 - Accepted 2 October 2006)

RESUME

Les indices de rétention, I, de 13 alkylnaphtalènes isomères, de masses molaires différentes ont été corrélés avec 2 indices structuraux de type géométrique : le nombre de chaînes moléculaires de longueur 3 (MWC3), et la première composante de l'accessibilité (E1u), un indice du groupe WHIM. Ces 2 indices, qui ne sont pas corrélés ($r = 0,443$ pour une valeur de $p = 0,113$), permettent d'expliquer 99,4 % de la variabilité de I. En outre, le modèle obtenu, très hautement significatif ($F = 778,29$), permet de reproduire l'ordre d'éluion expérimental, l'erreur standard étant de l'ordre d'une unité d'indice.

Si le 1,3- diméthylnaphtalène a une influence relativement grande sur le calcul des valeurs estimées de I, et sur la valeur de certains paramètres estimés, par contre aucune observation aberrante n'a été relevée.

Mots clés : indice de rétention, descripteurs moléculaires, régression linéaire multiple, résidus caractéristiques, observations influentes

ABSTRACT

The retention indices, I, of 13 alkylnaphtalenes isomers, of different molecular weights, were correlated with 2 structural indices of geometrical type : the molecular walk count of length 3 (MWC3), and the first component accessibility (E1u), an index of group WHIM. These 2 indices, are not correlated ($r = 0.443$ for a p -value = 0.113), and make it possible to explain 99.4 % of the variability of I. Moreover, the model obtained, which is very highly significant ($F = 778.29$), makes it possible to reproduce the experimental order of elution, the standard error being of the order of one index unit.

If the 1,3- dimethylnaphtalene has a relatively large influence on the calculation of the estimated values of I, and on the value of certain estimated parameters, on the other hand no outlier was pointed out.

Keywords : retention index, molecular descriptors, multiple linear regression, characteristic residuals, influential observations

INTRODUCTION

Les relations structure – rétention chromatographique sont étudiées pour trois raisons principales : explication des mécanismes des séparations chromatographiques, prédiction des grandeurs de rétention, et caractérisation des propriétés physiques des solutés. Celles-ci sont particulièrement importantes pour estimer la réactivité et la bioactivité des contaminants et / ou des polluants, ce qui peut faciliter le suivi de leur évolution dans la nature.

Le paramètre de forme η (Kaliszan *et al.*, 1979), ou un paramètre légèrement différent appelé rapport de la longueur à la largeur L / B (Wise *et al.*, 1981), en association avec l'indice de connectivité moléculaire $^1\chi$ (Randic¹, 1975) sont utiles pour la prédiction (modèles bi-paramétriques) des différences de rétention de HAP isomères séparés par chromatographie gazeuse, ou par chromatographie liquide à polarité de phase inversée. Ainsi, les indices de rétention de huit diméthylnaphtalènes (DMN) isomères ont-ils pu être corrélés, après élimination du 2,7-DMN, avec les descripteurs η et $^1\chi$, pour un coefficient de détermination R^2 (%) = 88,5, et une erreur standard $S = 9,70$ (Lamparczyk *et al.*, 1986). Cependant, de tels descripteurs moléculaires ne peuvent pas servir pour la comparaison entre rétentions de HAP de masses molaires différentes (Kaliszan, 1987).

Le but de ce travail consiste, dans une première étape, à corréliser les indices de rétention, prélevés dans la littérature (Lee *et al.*, 1979), de seize alkyl naphthalènes isomères (méthyl ; éthyl) substitués de masses molaires différentes, avec des descripteurs moléculaires de définition plus récente en calculant des modèles linéaires statistiques à effets fixes. Dans une deuxième étape, nous analyserons les résultats obtenus en vérifiant successivement les hypothèses de base du modèle et sa validité (analyse des résidus caractéristiques), puis nous essaierons de situer les observations éventuellement influentes sur le modèle de régression pour en apprécier l'impact, et savoir si elles peuvent en affecter l'utilisation (diagnostics d'influence).

COLLECTE DES DONNEES ET METHODES DE CALCULS

Les séparations ont été réalisées sur une colonne capillaire (longueur : 12 m ; diamètre intérieur : 0,3 mm) en verre dont les parois internes sont recouvertes d'un mince film (0,17 μm d'épaisseur) d'une phase stationnaire méthylphénylsilicone SE-52 (5 % phényl), montée sur un chromatographe Varian (modèle 3700) dont la température du four est élevée de 50 à 250 °C à raison de 2 °C/min. (On trouvera dans Lee *et al.*, (1979) des renseignements complets sur les conditions de séparation).

Les composés analysés sont rapportés à des standards de la même famille (naphtalène, phénanthrène, chrysène et picène). Dans ces conditions, en désignant par n le nombre de cycles d'un standard, son indice de rétention est pris égal à 100 n . Les indices de rétention, I_X , sont alors calculés à partir de la relation (Van den Dool & Kratz, 1963) :

$$\frac{I_X}{100} = n + \frac{t_{R_X} - t_{R_n}}{t_{R_{n+1}} - t_{R_n}} \quad (1)$$

où t_R désigne le temps de rétention, alors que les indices X, n et $n+1$ se rapportent au soluté et au nombre de cycles des standards qui l'encadrent sur le chromatogramme.

Les indices de rétention des naphthalènes ont été reliés à des paramètres structuraux (MWC3 ; E1u) qu'on peut rattacher au type géométrique (Karelson, 2000 ; Todeschini & Consonni, 2000) et calculés à l'aide du logiciel DRAGON Professional (Todeschini *et al.*, 2005). Les molécules sont entrées, et leur géométrie optimisée (méthode semi-empirique AM1), à l'aide d'un logiciel de modélisation moléculaire (Hyperchem, 2000).

Le nombre de chaînes moléculaires est lié au degré de ramification (comme $^1\chi$) et à la taille des molécules (Rücker, 1993). Le nombre de chaînes moléculaires de longueur k, MWCK, est donné par la relation :

$$MWCK = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_{ij}^k \quad (2)$$

où les a_{ij}^k désignent les éléments de la matrice d'adjacence, élevés à la puissance k, et n le nombre d'atomes de la molécule.

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 :

$$s_{jk} = \frac{\sum_{i=1}^n w_i (q_{ij} - \bar{q}_j)(q_{ik} - \bar{q}_k)}{\sum_{i=1}^n w_i} \quad (3)$$

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

Six modes de pondération sont proposés (dans le cas le plus simple tous les w_i sont posés égaux à 1), 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 (Todeschini *et al.*, 1994 ; Todeschini & Gramatica, 1997).

La première composante de l'accessibilité (E1u) est un indice du groupe WHIM, dirigé mais non pondéré ($w_i = 1$, pour tout i).

L'estimation par les moindres carrés des coefficients de régression suppose que les données suivent la loi normale, ce qui sera vérifié systématiquement.

Les variables explicatives (MWC3 ; E1u) ont été choisies après une double sélection : application de la technique de sélection régressive à la suite de la technique de sélection pas à pas (MINITAB, 2003), pour différents niveaux de signification.

RESULTATS ET DISCUSSION

Avant de procéder au développement effectif des équations de régression, la qualité statistique des variables dépendantes et explicatives a été vérifiée.

Quelques résultats partiels seront présentés à titre d'illustration, en exploitant les données qui apparaissent dans le Tableau 1.

TABLEAU 1

Valeurs des Indices de Rétention (I) et des Paramètres Structuraux E1u et MWC3
(Nombre de Chaînes Moléculaires de Longueur 3)

Nom	Symbole	(I)	MWC3	E1u	N°	N°
Naphtalène	N	200,00	4,745	0,540	1	-
2-Méthyl	2-MN	218,14	4,860	0,551	2	1
1-Méthyl	1-MN	221,04	4,875	0,551	3	2
2-Ethyl	2-EtN	236,08	4,934	0,567	4	3
1-Ethyl	1-EtN	236,56	4,949	0,551	5	4
2,7-Diméthyl	2,7-DMN	237,71	4,963	0,572	6	-
1,3-Diméthyl	1,3-DMN	240,25	4,977	0,491	7	5
1,7-Diméthyl	1,7-DMN	240,66	4,977	0,537	8	6
1,6-Diméthyl	1,6-DMN	240,72	4,977	0,552	9	7
2,3-Diméthyl	2,3-DMN	243,55	4,977	0,586	10	8
1,4-Diméthyl	1,4-DMN	243,57	4,990	0,543	11	9
1,5-Diméthyl	1,5-DMN	244,98	4,990	0,549	12	10
1,2-Diméthyl	1,2-DMN	246,49	4,990	0,526	13	-
1,8-Diméthyl	1,8-DMN	249,52	4,990	0,604	14	11
2,3,6-Triméthyl	2,3,6-TMN	263,31	5,069	0,586	15	12
2,3,5-Triméthyl	2,3,5-TMN	265,90	5,081	0,620	16	13

Les diagrammes de probabilité établis à partir de ces données montrent que les variables considérées ne se distribuent pas toujours selon la loi normale, puisque les R^2 obtenus ne sont pas systématiquement supérieurs au R^2 critique (R_c^2) donné par les tables, comme le laissent voir les valeurs présentées dans le Tableau 2 pour $n = 16$ individus.

La loi de Laplace – Gauss a été vérifiée, pour le niveau $\alpha = 0,01$ (et pour $\alpha = 0,05$ en acceptant un petit défaut de normalité), après élimination de 3 composés : naphtalène, 2,7-DMN et 1,2-DMN (Tableau 2 ; $n = 13$ individus).

TABLEAU 2

Vérification de la Loi Normale pour un Nombre N d'Individus

n	I	MWC3	E1u	
16	R ² (%)	90,0	86,6	95,2
	R _C ² (%)	94,07 (pour α = 0,05) et 91,46 (pour α = 0,01)		
13	R ² (%)	92,1	92,9	92,7
	R _C ² (%)	93,02 (pour α = 0,05) et 89,88 (pour α = 0,01)		

Parmi les modèles sélectionnés nous avons retenu le plus simple à deux variables explicatives : MWC3 et E1u (de coefficient de corrélation : $r = 0,443$, pour une valeur $p = 0,113$), dont l'équation est :

$$\hat{I} = -793,32 (\pm 28,10) + 202,15 (\pm 6,10) \text{MWC3} + 53,43 (\pm 11,53) \text{E1u} \quad (4)$$

$n = 13$; $S = 1,18$; $R^2 = 99,4\%$; $F = 778,29$

Les valeurs des paramètres statistiques montrent que les deux descripteurs MWC3 et E1u permettent de corrélérer la rétention de 13 dérivés alkyl substitués du naphtalène, constitués de divers isomères de masses molaires différentes. En effet, la valeur du coefficient de détermination (R^2) signifie que 99,4 % de la variabilité de \hat{I} peut être expliquée par MWC3 et E1u, alors que l'erreur standard est de l'ordre d'une unité d'indice ($S = 1,18$). En outre, ce modèle très hautement significatif (grande valeur du paramètre de Fisher : $F = 778,29$) permet de reproduire l'ordre d'éluion expérimental, ce que fait ressortir la dernière colonne du Tableau 3 qui réunit les résidus caractéristiques.

L'analyse de ces résidus permet, en particulier, de vérifier empiriquement le bien-fondé des hypothèses du modèle (Dagnélie, 1998). En effet, la Figure 1-A reproduit le graphe des résidus standardisés (d_i) en fonction des valeurs ajustées (\hat{I}) sans tendance particulière, ce qui montre la constance de la variance σ^2 , c'est-à-dire son indépendance des régresseurs et de la variable dépendante ajustée ; la quasi-linéarité ($r = 0,9695$) du diagramme des scores normaux (Figure 1-B) est un indice de normalité. La valeur ajustée de la statistique de Durbin - Watson ($d = 1,68$) est plus grande que la valeur supérieure donnée par les tables pour 2 régresseurs, et pour tout risque raisonnable α , ce qui établit l'indépendance des résidus.

La colonne (1) du Tableau 3 montre des résidus ordinaires, e_i , peu élevés, dont 4 (e_3, e_5, e_7, e_{11}) dépassent en valeur absolue l'erreur standard, $S = 1,18$. Tous les résidus standardisés d_i de la colonne (2) sont compris entre les limites ± 2 . La colonne (3) rassemble les résidus studentisés internes r_i , qui sont du même ordre de grandeur que les d_i

correspondants. On a ici $p = 3$ et $n = 13$, et on constate que tous les r_i sont inférieurs en valeur absolue à $t(0,025 ; n - p) [= 2,28]$ qui est le 0,975 quantile d'une loi de Student avec $(n - p)$ degrés de liberté. La colonne (4) donne les valeurs de h_{ii} , $i^{\text{ème}}$ terme diagonal de la matrice de projection : $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ où \mathbf{X} est la matrice des valeurs observées des variables explicatives et \mathbf{X}' sa transposée ; ces valeurs sont utiles pour le calcul des résidus caractéristiques. La colonne (5) contient les résidus prédits ; la valeur pour le point 5 est substantiellement plus élevée que le résidu ordinaire correspondant, ce qui indique, certainement, un point où le modèle s'ajuste raisonnablement bien, sans pour autant conduire à de bonnes estimations pour de nouvelles données. La colonne (6) montre le calcul de la somme des carrés des erreurs de prédiction (statistique PRESS) obtenue pour ce modèle. Notons également que les points 5 et 11 contribuent respectivement pour 28 et 20 % à la statistique PRESS, ce qui en fait des points relativement influents. La colonne (7) condense les résidus studentisés externes t_i , pour lesquels l'estimation $S_{(i)}^2$ de σ^2 est calculée (Montgomery & Peck, 1992) selon :

$$S_{(i)}^2 = \frac{(n - p) \text{CME} - e_i^2 / (1 - h_{ii})}{n - p - 1} \quad (5)$$

pour $(n - 1)$ observations, la $i^{\text{ème}}$ étant exclue ; CME est le carré moyen des écarts, et $k = (p - 1)$ le nombre de variables explicatives.

Les plus grandes valeurs de t_i , observées pour les points 5 et 11, ne diffèrent pas substantiellement des valeurs correspondantes des r_i , alors que les valeurs $S_{(5)}^2$ ($= 1,1675$) et $S_{(11)}^2$ ($= 1,0590$) sont inférieures au carré moyen des écarts, $\text{CME} = 1,4$, ce qui confirme la relative influence des points 5 et 11.

Comme les t_i sont inférieurs en valeur absolue à $t(0,025 ; n - p - 1) [= 2,262]$, l'analyse des résidus studentisés internes et externes ne détecte pas d'observation aberrante (outlier).

Enfin, en désignant par SCT la somme des carrés totale, la statistique PRESS conduit à un R^2 de prédiction égal à :

$$R_{\text{prédiction}}^2 = 1 - \frac{\text{PRESS}}{\text{SCT}} = 1 - \frac{27,8084}{2175,2} = 0,9872$$

Ainsi, le modèle permettrait d'expliquer environ 98,72 % de la variabilité de nouvelles observations estimées, en comparaison de la variabilité des données originales expliquées à 99,4 % par un ajustement par les moindres carrés. La capacité prédictive du modèle semble donc globalement satisfaisante.

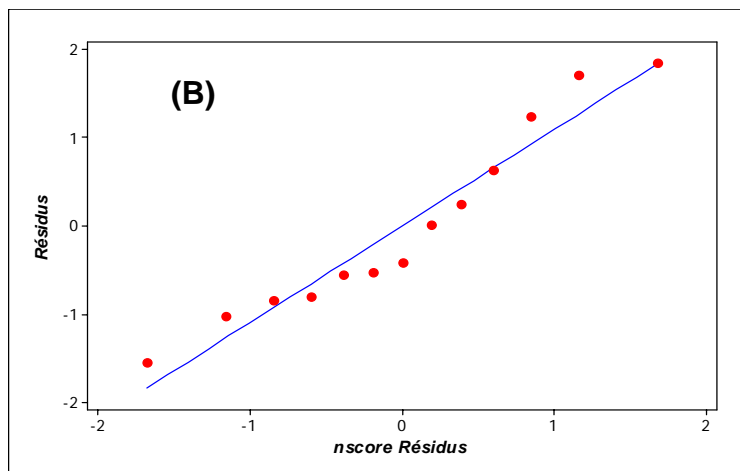
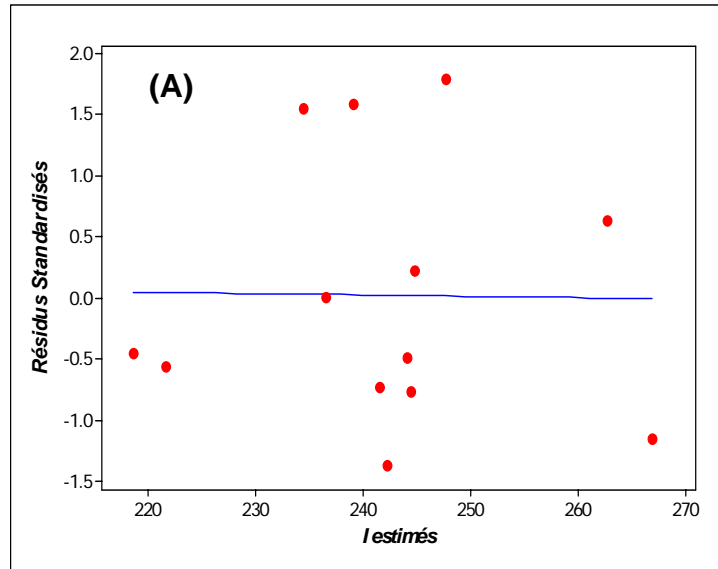


Figure 1. (A) : Graphe des résidus standardisés en fonction des valeurs ajustées.
(B) : Diagramme des scores normaux.

TABLEAU 3
Résidus Caractéristiques (Dagnélie, 1998 ; Montgomery & Peck, 1992) et Valeurs
Prédites (\hat{I}) des Indices de Rétention

<i>i</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	- 0,4180	- 0,4499	- 0,4312	0,37844	- 0,67243	0,4522	- 0,4294	218,558
2	- 0,5502	- 0,5581	- 0,5379	0,30016	- 0,7862	0,6180	- 0,5379	221,590
3	1,7084	1,5564	1,6962	0,13224	1,9687	3,8759	1,6870	234,372
4	0,0108	0,0096	0,0092	0,09072	0,0119	0,0001	0,0091	236,549
5	1,2458	1,5860	1,7392	0,55565	2,8036	7,8602	1,7296	239,004
6	- 0,8014	- 0,7315	- 0,7133	0,13539	- 0,9270	0,8592	- 0,7102	241,461
7	- 1,5427	- 1,3696	- 1,4415	0,08616	- 1,6882	2,8499	- 1,4342	242,263
8	- 0,5289	- 0,4825	- 0,4631	0,13443	- 0,6111	0,3734	- 0,4612	244,079
9	- 0,8399	- 0,7634	- 0,7463	0,12826	- 0,9635	0,9282	- 0,7297	244,410
10	0,2496	0,2241	0,2132	0,10680	0,2795	0,0781	0,2123	244,730
11	1,8516	1,7931	2,0652	0,23204	2,4111	5,8135	2,0532	247,668
12	0,6334	0,6326	0,6125	0,27784	0,8772	0,7694	0,6099	262,677
13	- 1,0186	- 1,1570	- 1,1794	0,44186	- 1,8249	3,3303	- 1,1738	266,919
PRESS = 27,8084								

La numérotation des observations *i* correspond à la dernière colonne du Tableau 1.

$$(1) : e_i = I_i - \hat{I}_i \quad (2) : d_i = e_i / \sqrt{MCE} \quad ; \quad (3) : r_i = e_i / \sqrt{MCE(1 - h_{ii})} \quad ; \quad (4) : h_{ii} \quad ; \quad (5) : e_i / (1 - h_{ii}) \quad ; \quad (6) : [e_i / (1 - h_{ii})]^2 \quad ; \quad (7) : t_i = e_i / \sqrt{S_{(i)}^2 (1 - h_{ii})} \quad ; \quad (8) : \hat{I}.$$

Le Tableau 4 condense d'autres mesures d'influence utiles (Montgomery & Peck, 1992), dont l'étude complète la recherche des observations aberrantes.

En analyse de régression, on appelle point levier une observation *i* qui influence considérablement les estimations parce que ses valeurs par rapport aux variables explicatives diffèrent beaucoup du reste des données. La valeur critique pour déterminer les points leviers correspond à $\frac{2p}{n} = \frac{2 \cdot 3}{13} = 0,46154$. On constate que tous les h_{ii} sont inférieurs à cette valeur critique 0,46154, à l'exception de $h_{55} = 0,55565$. On en déduit que la cinquième observation est un point levier pour le problème étudié ; vient ensuite l'observation 13 qui est la deuxième observation ayant la plus grande valeur de h_{ii} .

La distance de Cook :

$$D_i = \frac{1}{p} \frac{h_{ii}}{1 - h_{ii}} d_i^2 \tag{6}$$

permet d'étudier l'influence d'une observation *i* sur les coefficients de régression estimés par les moindres carrés. Encore une fois, l'observation 5, pour laquelle $D_5 > 1$, semble contribuer fortement à la détermination des paramètres.

TABLEAU 4
Diagnostics d’Influences (Montgomery & Peck, 1992)

i	D _i	DFFIT _i	DFBETAS _{0,i} *	DFBETAS _{1,i} **	DFBETAS _{2,i} ***	COVRATIO _i
1	0,04108	- 0,3365	- 0,3004	0,2954	- 0,0890	2,0767
2	0,04454	- 0,3523	- 0,3045	0,2977	- 0,0827	1,7840
3	0,12305	0,6621	0,3895	- 0,4136	0,2737	0,6933
4	0,00000	0,0029	0,0009	- 0,0008	- 0,0003	1,5086
5	1,04849	1,9448	- 0,4829	0,8375	- 1,7946	1,3051
6	0,02793	- 0,2823	0,0574	- 0,0946	0,1841	1,3471
7	0,05895	- 0,4426	0,0585	- 0,0887	1,1412	0,8095
8	0,01205	- 0,1825	- 0,0213	0,0435	- 0,1184	1,4775
9	0,02858	- 0,2863	0,1007	- 0,1317	0,1686	1,3162
10	0,00200	0,0737	- 0,0252	0,0314	- 0,0344	1,5129
11	0,32384	1,1352	0,0481	- 0,2302	0,9035	0,5636
12	0,05132	0,3799	- 0,3037	0,2790	0,0206	1,6822
13	0,35327	- 1,0494	0,6316	- 0,4776	- 0,5254	1,6030

* : ordonnée à l’origine ; ** : MWC3 ; *** : E1u

La numérotation des observations i correspond à la dernière colonne du Tableau 1.

La statistique représentée par le symbole DFFITS :

$$DFFITS_i = \left(\frac{h_{ii}}{1 - h_{ii}} \right)^{1/2} t_i \tag{7}$$

permet de mesurer l’influence d’une observation i sur la valeur ajustée ou prédite. Les observations 5, 11 et 13 caractérisées par des valeurs absolues de $DFFITS_i > 2\sqrt{p/n} = 2\sqrt{3/13} = 0,96077$, sont inhabituelles.

L’ampleur du changement, en unité d’écart – type, du $j^{ème}$ coefficient de régression estimé, $\hat{\beta}_j$, lorsque la $i^{ème}$ observation est exclue, est indiquée par la statistique :

$$DFBETAS_{j,i} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{\sqrt{S^2_{(i)} C_{jj}}} \tag{8}$$

où C_{jj} est le $j^{ème}$ élément diagonal de la matrice $(X'X)^{-1}$, et $\hat{\beta}_{j(i)}$ le $j^{ème}$ coefficient de régression calculé sans utiliser l’observation i.

L’examen des colonnes $DFBETAS_{j,i}$, en tenant compte de la valeur critique $2\sqrt{n} = 2\sqrt{13} = 0,5547$, fait ressortir immédiatement que le point 5 a une grande influence sur $\hat{\beta}_1$ (MWC3) et $\hat{\beta}_2$ (E1u) ; le point 13 influe sur $\hat{\beta}_0$ (ordonnée à l’origine) et un peu

moins sur $\hat{\beta}_2$ (valeur absolue de $DFBETAS_{2,13}$ proche de 0,5547) ; enfin le point 11 influe grandement sur $\hat{\beta}_2$.

Le $COVRATIO_i$, que l'on calcule pratiquement à partir de la relation :

$$COVRATIO_i = \frac{(S_{(i)}^2)^p}{(CME)^p} \left(\frac{1}{1-h_{ii}} \right) \quad (9)$$

fait ressortir le rôle de la $i^{\text{ème}}$ observation sur la précision de l'estimation.

Les $COVRATIO_i$ sont supérieurs à 1 pour les points 5 et 13, et inférieurs à 1 pour le point 11 ; les deux premiers améliorent la précision, alors que le dernier la détériore.

CONCLUSION

Des logiciels récents (Dragon, Hyperchem) permettent le calcul de nombreux descripteurs moléculaires utilisés pour modéliser une grande variété de propriétés. On trouve dans d'autres (Minitab ..), du moins partiellement, une série d'outils créés pour la validation des modèles de régression, dont l'utilisation permet de mettre en évidence des situations particulières.

Ainsi, les indices de rétention d'alkylnaphtalènes isomères de masses molaires différentes ont pu être corrélés avec 2 indices structuraux de type géométrique, après nettoyage préalable des données. Le modèle hautement significatif obtenu, dont nous avons pu vérifier les hypothèses de départ, permet de reproduire l'ordre d'élution expérimental, et possède une capacité prédictive satisfaisante. On n'a pas relevé d'observation présentant des valeurs extrêmes pour toutes les caractéristiques des Tableaux 3 et 4 et qui, du point de vue statistique, puisse être considérée comme aberrante. Seule l'observation 5 (1,3-DMN) a, par rapport aux autres observations, une influence relativement grande sur le calcul des valeurs estimées de I, et sur la valeur de certains paramètres estimés ($\hat{\beta}_1$, $\hat{\beta}_2$).

REFERENCES

- Dagnélie, P. 1998. *Statistique théorique et appliquée*. Tome 2. De Boeck Université, Paris, Bruxelles, pp. 659.
- HyperchemTM release 6.03 for windows, molecular modelling system 2000.
- Kaliszan, R., Lamparczyk, H., Radecki, A. 1979. A relationship between repression of dimethylnitrosamine – demethylase by polycyclic aromatic hydrocarbons and their shape. *Biochem. Pharmacol.*, 28 : 123-125.
- Kaliszan, R. 1987. *Quantitative structure – chromatographic retention relationships*. J. Wiley, New York, Chichester, Brisbane, Toronto, Singapore, pp. 303.
- Karelson, M. 2000. *Molecular Descriptors in QSAR / QSPR*. Wiley – Interscience, New York, pp. 430.
- Lamparczyk, H., Kaliszan, R., Radecki, A. 1986. Correlation between retention on liquid crystalline phases and chemical structure. *J. Chromatogr.*, 361 : 442 – 444.

- Lee, M., Vassilaros, D.L., White, C.M., Novotny, M. 1979. Retention indices for programmed- temperature gas chromatography of polycyclic aromatic hydrocarbons. *Anal. Chem.*, 51 : 768 – 773.
- MINITAB, release 14.1, statistical software 2003.
- Montgomery, D.C. and Peck, E.A. 1992. *Introduction to Linear Regression Analysis*, 5th Edition. Wiley – Interscience, New York, pp. 527.
- Randic', M. 1975. On characterization of molecular branching. *J. Am. Chem. Soc.*, 97 : 6609 – 6615.
- Rücker, G. 1993. Counts of all walks as atomic and molecular descriptors. *J. Chem. Inf. Comput. Sci.*, 33 : 683 – 695.
- Todeschini, R., Lasagni, M., Marenzo, E. 1994. New molecular descriptors for 2 D- and 3 D-structures. Theory. *J. Chemom.*, 8: 263 – 273.
- Todeschini, R., Gramatica, P. 1997. 3 D- modelling and prediction by WHIM descriptors. Part 5. Theory development and chemical meaning of WHIM descriptors. *Quant. Struct. – Act. Realt.*, 16: 113 – 119.
- Todeschini, R., Consonni, V. 2000. *Handbook of molecular descriptors*. Wiley VCH, Weinheim, pp. 667
- Todeschini, R., Consonni, V., Mauri, A., Pavan, M. 2005. *Logiciel DRAGON, Version 5.3*. Milano.
- Van den Dool, H. and Kratz, P.D. 1963. A Generalization of the retention index system including linear temperature programmed gas - liquid partition chromatography. *J. Chromatogr.*, 11 : 463 - 471.
- Wise, S., Bonnett, W.J., Guenther, F.R., May, W.E. 1981. A relationship between reversed – phase C₁₈ liquid chromatographic retention and the shape of polycyclic aromatic hydrocarbons. *J. Chromatogr. Sci.*, 19: 457 – 465.