

CHROM. 15,114

COMPORTEMENT CHROMATOGRAPHIQUE DE PHASES STATIONNAIRES

II. APPLICATION DE L'ANALYSE FACTORIELLE

R. FELLOUS*

Laboratoire de chimie organique, Institut Polytechnique Méditerranéen, Université de Nice, Parc Valrose, F-06034 Nice Cedex (France)

D. LAFAYE DE MICHEAUX

Laboratoire de signaux et systèmes, E.R.A. 835, 41 Bd. Napoléon III, F-06200 Nice (France)

et

L. LIZZANI-CUVELIER et R. LUFT

Laboratoire de chimie organique, Institut Polytechnique Méditerranéen, Université de Nice, Parc Valrose, F-06034 Nice Cedex (France)

(Reçu le 28 mars 1982; manuscrit modifié reçu le 14 juin 1982)

SUMMARY

Chromatographic behaviour of stationary phases. II. Application of factor analysis

A "principal component analysis" applied to a set of retention times of 17 benzenic monosubstituted compounds on a collection of 21 stationary phases, leads to the following relationship:

$$x_{P,\varphi_i} = a_\varphi \cdot x_{P,\varphi_1} + b_\varphi \cdot x_{P,\varphi_2} + c_\varphi \cdot x_{P,\varphi_3} + d_\varphi$$

Knowing the experimental data x_{P,φ_1} , x_{P,φ_2} and x_{P,φ_3} of any substance P, measured on three phases φ_1 , φ_2 and φ_3 (respectively non-polar, polyfluorinated and polar) one can estimate the gas-liquid chromatographic retention data x_{P,φ_i} of the same substance P on any stationary phase φ_i with the aid of the above relationship. (a_φ , b_φ , c_φ and d_φ are the coefficients of the non-restricted multiparametric relationship.)

INTRODUCTION

La prévision des grandeurs de rétention en chromatographie gaz-liquide a été abordée sous des aspects divers.

Une première série de travaux est basée sur la recherche de relations empiriques entre les grandeurs de rétention et des variables explicatives¹, telles que le point d'ébullition, la masse moléculaire, le nombre d'atomes de carbone d'une molécule ou encore des constantes physicochimiques issues de traitements particuliers, tels que la chaleur de vaporisation¹, les coefficients de partage² ou encore les coefficients d'acti-

té³. On est alors rapidement confronté avec l'interprétation de ces relations et leur limite d'application pour l'analyse des interactions soluté-solvant.

En retenant comme hypothèse fondamentale que chaque élément de structure contribue pour sa part à la rétention et que cette contribution n'est pas affectée par l'environnement, il devient possible, à partir de ces relations empiriques, d'établir des règles d'additivité⁴⁻¹⁴. On assiste ainsi à la naissance de relations de "Topologie-Information"¹⁵ telles qu'on peut les dégager par exemple à l'aide du logiciel DARC^{16,17}. Ces relations permettent de déterminer la contribution à la rétention de chacun des sites d'une molécule et de chiffrer ainsi la participation de chacun de ses éléments structuraux à cette rétention.

Une deuxième approche fait appel aux méthodes de "l'analyse des données" qui permettent entre autres une étude prévisionnelle des grandeurs de rétention, ainsi qu'une analyse globale des interactions soluté-solvant, c'est à dire une analyse qui ne préjuge pas de la nature des solutés et des solvants étudiés.

Parmi les applications de l'analyse des données à la chromatographie gazeuse, signalons en particulier "l'analyse en composantes principales" ou "A.C.P." utilisée surtout par Weiner et Howery¹⁸⁻²¹.

Cette méthode a permis à Malinowski *et al.*²² de prévoir les valeurs des coefficients d'activité, à certains auteurs d'identifier les "facteurs principaux" qui influencent d'une manière prépondérante le phénomène complexe de la rétention^{20,23-27} et enfin à d'autres d'établir des classements des solutés²⁸ et/ou des solvants²⁸⁻³¹. Un tel classement des solvants a été abordé d'autre part à l'aide des méthodes taxonomiques d'agrégation^{32,33} et de la théorie de l'information^{33,34}. Pratiquement, toutes ces méthodes ont été testées à partir des lots de mesures réalisées soit par Rohrschneider³⁵, soit par McReynolds³⁶.

Dans le cadre de nos propres études chromatographiques³⁷ nous avons retenu un autre lot de substances; celles-ci possèdent toutes un même squelette fixe, le noyau benzénique, et des fonctions chimiques diverses. Les données relatives à ces substances nous ont déjà permis, à l'aide d'une étude taxonomique et sur la base du critère du coefficient de corrélation, de regrouper les phases stationnaires en 3 classes et de voir apparaître, à côté des phases dites "polaires" et "apolaires", la famille des phases "fluorées", indépendante des précédentes de par son comportement.

Le but de ce travail est d'étudier les différentes variantes d'A.C.P. et d'envisager un modèle multiparamétrique de prévision de comportement de phases stationnaires pour l'évaluation des grandeurs de rétention.

PARTIE EXPÉRIMENTALE

Les logarithmes de temps réduits qui nous permettent de réaliser cette étude sont issus de mesures que nous avons effectuées, dans des conditions isothermes et pour un ensemble de phases stationnaires, sur une série de dérivés mono- et poly-substitués du benzène et pour lesquelles nous avons déjà défini le détail des conditions expérimentales³⁷. Le Tableau I regroupe les logarithmes de temps réduit ($\log t'$) du benzène sur les différentes phases et permet de déterminer les données de rétention réduites des différents composés, à partir des données publiées antérieurement³⁷.

TABLEAU I
LOGARITHMES DE TEMPS RÉDUIT DU BENZÈNE

<i>Phase</i>	<i>log t' (benzène)</i>	<i>Phase</i>	<i>log t' (benzène)</i>
Apiezon L	0.9906	Carbowax 20M	0.0523
SE-30	0.6180	XF-1150	0.7380
DC-200	0.8170	FFAP	0.9857
PMPE 5R	0.6218	PDEAS	0.6876
Apiezon H	0.9476	DEGA	0.9360
OV-17	0.8100	DEGS	0.4771
UCON 50 HB 2000	0.9868	EGS	0.5211
ULB 550X	0.7482	DEGSéb	1.1009
OV-210	0.4298	QF-1	0.7251
OV-225	0.6907	DCL 5X	-0.2565
NPGS	0.6207		

RÉSULTATS

L'application de l'A.C.P. par diagonalisation de la matrice de covariance, aux logarithmes de temps réduits, montre que trois variables factorielles suffisent à cerner les tendances du phénomène examiné et rendent compte de 99.8 % de la variance totale (la variance liée à la première composante principale représente 96.3 % de la variance totale, celle de la seconde 2.8 % et celle de la troisième 0.7 %). Dans ce type d'A.C.P., les variables initiales qui ont une forte variance ont une importance accrue.

Dans les plans factoriels b_1, b_2 et b_2, b_3 nous avons représenté les projections des vecteurs unitaires portés par les axes des variables initiales (Figs. 1 et 2). Dans chacun des deux plans factoriels les vecteurs de directions voisines correspondent à des phases stationnaires fortement corrélées entre-elles.

En examinant la Fig. 1 on observe un regroupement des phases en deux faisceaux très proches et situés de part et d'autre de l'axe b_1 . Ce regroupement et la proximité au cercle de corrélation de l'ensemble des vecteurs "phases" dénote à la fois l'absence d'un pouvoir discriminant et la forte "représentativité", c'est à dire le poids très élevé du plan factoriel b_1, b_2 . Par contre, le plan factoriel b_2, b_3 (Fig. 2) met en évidence la distribution des phases en trois classes: les phases polaires (groupe C), apolaires (groupe A) et fluorées (groupe B). Nous avons déjà abouti à une distribution semblable à l'aide de la classification ascendante hiérarchique³⁷.

Pour les trois premiers facteurs principaux b_1, b_2 et b_3 les valeurs des coefficients de corrélation f_1, f_2, f_3 , issus de la régression entre les variables factorielles et les variables initiales, ainsi que les variances liées à chacun de ces axes, sont rassemblées dans le Tableau II.

Rappelons que les variables factorielles ne peuvent être rattachées à l'avance à tel ou tel facteur physicochimique. L'étude du Tableau II fait ressortir les faits suivants:

(1) Les variables "phases" sont toutes étroitement corrélées avec le facteur b_1 , les deux faisceaux de vecteurs qui en résultent sont fortement pincés et proches de l'axe b_1 . D'autre part, des valeurs voisines des coefficients de corrélation f_1 expriment le fait que le canevas des grandeurs de rétention d'une population de produits divers

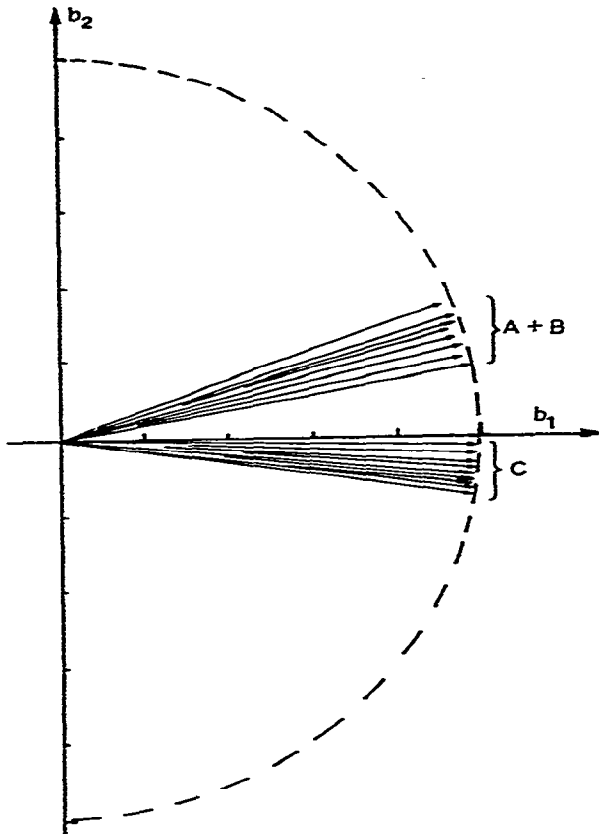


Fig. 1. Projection des vecteurs unitaires portés par les axes des variables dans le plan factoriel b_1, b_2 .

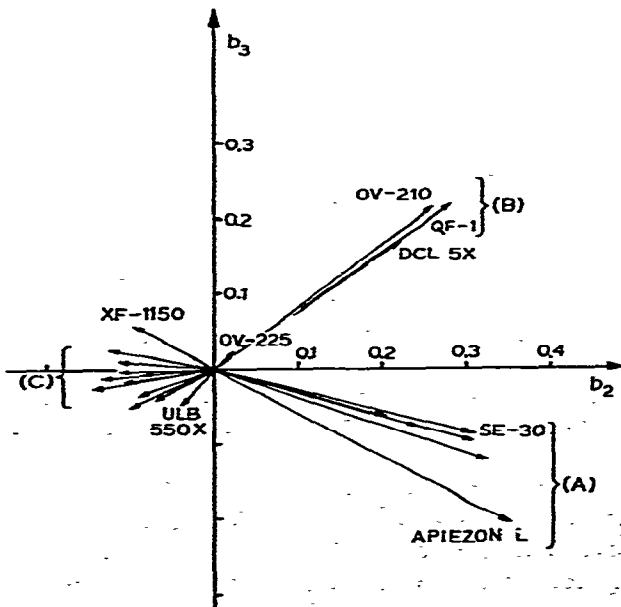


Fig. 2. Projection des vecteurs unitaires portés par les axes des variables dans le plan factoriel b_2, b_3 .

TABLEAU II

COEFFICIENTS DE CORRÉLATION DES TROIS PREMIERS FACTEURS PRINCIPAUX b_1 , b_2 , b_3 AVEC LES VARIABLES INITIALES

φ_i	f_1	f_2	f_3
Apiezon L	0.9119	0.3521	-0.2060
SE-30	0.9471	0.3014	-0.0823
DC-200	0.9427	0.3064	-0.0971
PMPE 5R	0.9771	0.1980	-0.0581
Apiezon H	0.9397	0.3185	-0.1221
OV-17	0.9672	0.2370	-0.0762
UCON 50 HB 2000	0.9922	-0.1037	-0.0535
ULB 550X	0.9959	-0.0369	-0.0454
OV-225	0.9988	0.0177	0.0280
NPGS	0.9960	-0.0779	0.0064
Carbowax 20M	0.9897	-0.1325	-0.0124
XF-1150	0.9931	-0.0907	0.0661
FFAP	0.9876	-0.1472	-0.0248
PDEAS	0.9932	-0.1072	-0.0166
DEGA	0.9938	-0.1097	-0.0054
DEGS	0.9913	-0.1175	0.0090
EGS	0.9907	-0.1224	0.0284
DEGSeb	0.9961	-0.0694	-0.0358
OV-210	0.9418	0.2603	0.2075
DCL 5X	0.9596	0.2141	0.1654
QF-1	0.9322	0.2834	0.2182
Variance liée à chacun des axes (%)	96.3	2.8	0.7

est le même pour l'ensemble des variables "phases", indépendamment des plages des valeurs des grandeurs numériques.

En clair, si sur une phase φ_1 le temps de rétention du corps A est "court" et celui du corps B "long", la situation sera qualitativement analogue pour les temps de rétention sur une phase φ_2 . Les raisons qui font que A "sorte" plus rapidement que B sont liées à la nature des facteurs $b_2, b_3, b_4 \dots b_n$.

(2) Les variables "phases" se répartissent en deux groupes, en fonction de la valeur du facteur b_2 ; celles pour lesquelles f_2 est négative présentent un caractère "polaire" et la capacité de s'engager dans des liaisons hydrogène.

On peut mettre en évidence des relations linéaires entre les valeurs de f_2 et les "échelles de polarité" P de McReynolds³⁹, Z_1 de Wold²⁹ et X de Lowry³¹. On aboutit aux relations

$$f_2 = -0.161 P \cdot 10^{-3} + 0.373 \quad r = 0.901 \quad p = 18 \quad (1)$$

$$f_2 = -2.452 Z_1 + 0.083 \quad r = 0.935 \quad p = 14 \quad (2)$$

$$f_2 = -0.056 X + 0.120 \quad r = 0.944 \quad p = 14 \quad (3)$$

Ces résultats montrent bien les parallélismes existant entre les "échelles de polarité" et permettent de préciser le rôle de l'axe b_2 que l'on peut considérer comme discriminant vis à vis de la "polarité" d'une phase.

(3) Le facteur b_3 permet d'approfondir l'analyse des propriétés des phases dites "non polaires" qui se scindent en deux groupes distincts dont l'un rassemble les phases fluorées. Celles-ci constituent donc bien un groupe à part entière.

En définitive, la variance absorbée par les trois axes ($b_1 + b_2 + b_3$) est de 99.8 %; aussi un espace à trois dimensions suffit pour décrire l'information significative de nos données, comme le confirme l'application de la méthode Varimax⁴⁰. Cette dernière consiste en une analyse en composantes principales suivie d'une rotation orthogonale des axes factoriels. On remarque, après cette rotation, que les valeurs les plus élevées de f_1 concernent la classe des variables "phases" la plus fortement représentée (groupe C), et que celles de f_2 sont les plus élevées pour les phases du groupe A. Quant aux valeurs de f_3 , elles ne sont importantes que pour les phases fluorées. Pour les axes au-delà de b_3 , les composantes principales sont très faibles et dispersées positivement et négativement, ce qui montre bien qu'il s'agit d'un "facteur de bruit" non significatif.

Les résultats de l'A.C.P. par diagonalisation de la matrice de covariance sont certainement fortement influencés par la représentativité inégale des trois classes de phases mises en évidence. En particulier, le fait que les phases polaires soient mieux représentées par l'axe b_1 que les autres phases pourrait être dû, soit à une cause intrinsèque restant à expliquer, soit à leur nombre élevé dans l'échantillon des phases retenues. Aussi, pour confirmer la représentativité du groupe des phases fluorées, nous avons procédé à une A.C.P. sur la matrice de corrélation avec un jeu de pondération des variables⁴¹ qui donne une importance égale à chacun des trois groupes de phases considérées: tout se passe comme si chaque groupe comportait le même nombre de représentants. Dans notre cas, où le groupe le plus important rassemble 12 termes, on fait intervenir chaque phase apolaire deux fois et chaque phase fluorée quatre fois. Les résultats de cette A.C.P. permettent de constater: (1) que l'axe b_1 n'est pas plus représentatif des phases polaires que des autres phases; (2) que la distinction première entre phases apolaires/fluorées et polaires est maintenue; et (3) que l'axe b_3 rend compte d'une rôle particulier des phases fluorées.

Il en résulte qu'un modèle prévisionnel des grandeurs de rétention, réalisé uniquement à l'aide des données relevées sur seulement des phases polaires et apolaires ne peut pas rendre compte avec précision des grandeurs de rétention sur l'ensemble des phases. La déviation devient importante dans le cas de phases fluorées. Ce phénomène avait été déjà constaté par Rohrschneider⁴².

L'ensemble de nos analyses en composantes principales nous amène à proposer un modèle de prévision des données de rétention linéaire par régression faisant intervenir trois variables "phases" correspondant à trois phases stationnaires prises chacune dans l'un des trois groupes.

Modèle de régression à trois variables explicatives

Plusieurs critères permettent la sélection des trois phases, destinées chacune à représenter son groupe. Parmi ces critères, nous avons envisagé:

(1) *la régression multiple*³⁷; elle conduit à retenir les phases SE-30 (groupe A), OV-210 (groupe B) et NPGS (groupe C);

(2) *L'espace factoriel* issu de l'analyse en composantes principales. Dans notre cas cet espace est tridimensionnel et amène à retenir comme meilleur choix les phases Apiezon H (groupe A), OV-210 (groupe B) et DEGA (groupe C);

(3) *L'analyse discriminante*⁴⁰ dans laquelle on recherche parmi un ensemble de variables "phases" celles qui conduisent à la meilleure discrimination des produits. Contrairement aux critères précédents, l'analyse discriminante établit un ordre indépendant de la classification en groupes et basé sur l'efficacité de la discrimination. Les phases placées en tête sont cette fois NPGS (groupe C), DC-200 (groupe A) et QF-1 (groupe B).

En effectuant des régressions affines du type

$$x_{P,\varphi_i} = a_\varphi \cdot x_{P,\varphi_1} + b_\varphi \cdot x_{P,\varphi_2} + c_\varphi \cdot x_{P,\varphi_3} + d_\varphi \quad (4)$$

où x_{P,φ_i} représente la grandeur de rétention d'un produit P sur une phase φ_i .

Dans chacun des cas ci-dessus, on observe que les écarts entre les grandeurs de rétention expérimentales et les grandeurs recalculées sont très voisins ($\sigma_{e,\varphi} \approx 0.04$). En réalisant par ailleurs des régressions avec les données relatives aux phases les plus usuelles (SE-30, OV-210, DEGS), la valeur $\sigma_{e,\varphi}$ obtenue dans la reconstitution des données ne diffère guère de la précédente. Cette concordance provient de la bonne corrélation observée entre phases d'un même groupe taxonomique.

Avec ces dernières phases, l'application de la relation 4 aux dérivés monofonctionnels du benzène conduit aux résultats du Tableau III. Si, pour la plupart des phases, le coefficient b_φ est peu différent de zéro, il faut noter qu'il devient réelle-

TABLEAU III

APPLICATION DE LA RELATION 4 AUX DÉRIVÉS MONOFONCTIONNELS DU BENZÈNE

$$\log r_{\varphi Z,\varphi} = a_\varphi \cdot \log r_{\varphi Z,SE-30} + b_\varphi \cdot \log r_{\varphi Z,OV-210} + c_\varphi \cdot \log r_{\varphi Z,DEGS} + d_\varphi$$

r_φ = coefficient de corrélation; $\sigma_{e,\varphi}$ = écart type moyen de l'erreur d'estimation.

φ_i	a_φ	b_φ	c_φ	d_φ	r_φ	$\sigma_{e,\varphi}$
Apiezon L	1.753	-0.268*	-0.047*	-0.012	0.9916	0.05
DC-200	1.112	-0.025*	-0.005*	0.149	0.9993	0.01
PMPE 5R	1.158	0.092*	0.229	-0.334	0.9953	0.05
Apiezon H	1.472	-0.029*	-0.007*	-0.040	0.9926	0.05
OV-17	1.141	0.067*	0.127*	-0.061	0.9940	0.05
UCON 50 HB 2000	0.410	-0.233*	0.808	0.428	0.9967	0.04
ULB 550X	0.536	-0.124*	0.614	0.149	0.9968	0.04
OV-225	0.256	0.252	0.541	0.116	0.9988	0.03
NPGS	0.160*	0.043*	0.782	0.137	0.9984	0.03
Carbowax 20M	0.046*	-0.057*	0.934	-0.472	0.9960	0.05
XF-1150	-0.315	0.342	0.832	0.336	0.9991	0.02
FFAP	0.122*	-0.174*	0.980	0.511	0.9979	0.04
PDEAS	0.125*	-0.049*	0.913	0.123	0.9975	0.04
DEGA	0.085*	-0.016*	0.901	0.380	0.9982	0.04
EGS	-0.149*	0.105*	0.975	0.060	0.9994	0.02
DEGSéb	0.375	-0.099*	0.754	0.472	0.9975	0.04
QF-1	-0.019*	0.935	-0.030*	0.339	0.9966	0.03
DCL 5X	0.156*	0.879	0.103	-0.712	0.9975	0.03

* Valeur des coefficients de la régression pour laquelle le test de Fischer est supérieur à 2%.

TABLEAU IV

APPLICATION DE LA RELATION 4 AUX DÉRIVÉS POLYFONCTIONNELS DU BENZÈNE

Benzènes disubstitués	φ_i	a_φ	b_φ	c_φ	d_φ	r_φ	$\sigma_{c,\varphi}$
<i>ortho</i>	Carbowax 20M	-0.045*	0.029*	0.972	-0.016	0.979	0.08
	ULB 550X	0.482	-0.278	0.682	0.363	0.993	0.04
	OV-225	0.221	0.290	0.510	0.062	0.997	0.02
	PDEAS	0.247*	0.017*	0.813	0.058	0.994	0.04
	Apiezon L	1.385	-0.139*	-0.094*	0.221	0.996	0.03
	PMPE 5R	1.066	0.108	0.140	-0.181	0.999	0.02
<i>meta</i>	Carbowax 20M	0.142*	-0.056*	1.030	-0.421	0.995	0.04
	ULB 550X	0.811	-0.170*	0.510	-0.072	0.995	0.03
	OV-225	0.256	0.269	0.543	-0.056	0.999	0.01
	PDEAS	0.245*	0.059*	0.851	-0.111	0.998	0.03
	Apiezon L	1.393	-0.185*	-0.039*	0.170	0.997	0.02
	PMPE 5R	0.760	0.327*	0.253*	-0.210	0.976	0.07
<i>para</i>	Carbowax 20M	-0.026*	-0.115*	1.117	-0.210	0.995	0.04
	ULB 550X	0.157*	-0.085*	0.768	0.431	0.985	0.06
	OV-225	0.210	0.244	0.574	0.005	0.999	0.02
	PDEAS	0.263	-0.006*	0.872	-0.059	0.996	0.03
	Apiezon L	1.530	-0.178	-0.120	0.096	0.995	0.03
	PMPE 5R	1.294	0.002*	0.119*	-0.376	0.995	0.04
Benzènes tri- substitués	Carbowax 20M	-1.460*	-1.080	2.832	0.511	0.992	0.06
	ULB 550X	0.987	-0.002*	0.247	0.178	0.999	0.01
	OV-225	0.391*	0.443	0.293	0.105	0.999	0.02
	PDEAS	0.421	0.178	0.707	-0.084	0.999	0.01
	Apiezon L	1.631	-0.213	-0.053*	0.031	0.998	0.01
	PMPE 5R	0.912	0.036*	0.342	-0.080	0.999	0.01

ment significatif non seulement dans le cas des phases fluorées (QF-1 et DCL 5X), mais aussi dans celui des phases cyanées (OV-225 et XF-1150).

Les coefficients a_φ , b_φ et c_φ de la relation 4 ont été déterminés à partir d'une famille de dérivés monofonctionnels du benzène. Il restait à vérifier si la relation 4 pouvait constituer un modèle de portée plus générale. A cet effet, nous l'avons appliquée aux populations suivantes: benzènes disubstitués en *o*, benzènes disubstitués en *m*, benzènes disubstitués en *p*, benzènes trisubstitués³⁷. Les résultats, regroupés dans le Tableau IV, montrent que le modèle représenté par la relation 4 permet une bonne reconstitution de l'ensemble des données de rétention.

Enfin, en appliquant (Tableau V) la relation 4 à une population de 27 composés RZ (où R correspond à une chaîne aliphatique et Z à des fonctions diverses) pour laquelle nous disposons des indices de rétention déterminés par McReynolds³⁶ sur les phases SE-30, DC-1265 F (très voisine de la phase OV-210) et DEGS, nous obtenons des valeurs recalculées dont l'écart $\sigma_{c,\varphi}$ est inférieur à 16 unités d'indice de rétention (sauf en ce qui concerne la phase Apiezon L). La bonne concordance entre indices de rétention expérimentaux et recalculés, atteinte à partir d'une population restreinte, nous a semblé assez prometteuse pour tenter d'appliquer la relation 4 à l'ensemble des données de rétention déterminées par Rohrschneider³⁵ (Tableau VI) pour la mise au

point de son propre système de prévision. Ce dernier est basé sur cinq substances judicieusement choisies et censées représenter les différents types d'interactions pouvant coexister dans la chromatographie en phase gazeuse. Compte-tenu de la variété des produits examinés et des phases, l'écart $\sigma_{e,\varphi}$ qui se situe presque toujours en-dessous de 20 pèse peu sur les valeurs recalculées. Cette valeur de $\sigma_{e,\varphi}$ est du même ordre de grandeur que celle que l'on observe dans le système de prévision établi par Rohrschneider.

Une analyse en composantes principales par diagonalisation de la matrice de covariance que nous avons réalisée par ailleurs à partir des indices de rétention de Rohrschneider confirme d'une part la qualité de notre relation 4 et d'autre part la légitimité de notre choix d'un espace factoriel à trois dimensions.

DISCUSSION

L'A.C.P. est une méthode d'analyse de données qui, à travers ses différentes variantes, répond à de nombreuses approches visant toutes à condenser des informations groupées généralement dans un tableau. Celui-ci représente la "matrice des données de rétention étudiées", composée de n lignes "produits" — indicées par P — et de m colonnes "phases" — indicées par φ . Un élément de cette matrice, situé dans une colonne φ et dans une ligne P , est noté $x_{P,\varphi}$. Il représente la grandeur de rétention $Gr(P,\varphi)$ du produit P sur la phase φ . $Gr(P,\varphi)$ peut exprimer, selon le cas, soit un logarithme de temps réduit ($\log t'$), soit un logarithme de temps réduit relatif ($\log t'_H$), soit un logarithme de volume de rétention ($\log V_g, \log V_N$), soit enfin un indice de rétention (I) ou encore une différence d'indice (δI), considérée comme un incrément de fonction.

Le tableau, constitué d'observations interdépendantes, permet d'analyser des relations entre "produits" — les lignes — ou entre "phases" — les colonnes —. L'objectif consiste alors à condenser l'essentiel des informations résultant de l'observation en un nombre plus restreint de nouvelles informations, non observables directement et appelées "variables factorielles" ou "composantes principales".

Ces variables factorielles, notées b_1, \dots, b_q permettent, par des régressions multiples, de reconstituer de façon approchée la matrice des données initiales. Le choix de b_1, \dots, b_q est déterminé par l'optimisation d'un critère mesurant la qualité de cette reconstitution. Le pourcentage d'inertie associé aux q premières variables factorielles b_1, \dots, b_q , constitue un indice de qualité de la reconstitution de la matrice. Si cet indice est proche de 100 %, cela signifie généralement que (1) les variables factorielles situées au-delà de b_q décrivent en fait un "bruit" sans signification, issue en général des erreurs de mesures, et (2) les variables initiales x_{φ} , aux erreurs de mesures près, sont situées dans un espace de dimension réduite, cette situation étant due à leurs interdépendances.

On peut donc, à l'aide de l'A.C.P., déterminer un nombre minimal de variables observées qui permettent une reconstitution du tableau de données sans perte sensible d'information. Par ailleurs, la représentation des variables initiales dans l'espace de dimension réduite engendré par les q variables factorielles retenues permet de mettre en évidence les similitudes et les différences de comportement des phases étudiées.

Les différentes variantes de l'A.C.P. classiquement utilisées dans "l'Analyse de Données" sont au nombre de 4 et se distinguent entre-elles par :

le type de matrice à diagonaliser (produits scalaires, produits scalaires normalisés, covariances, corrélations);

le type de régression (affine ou linéaire) intervenant dans la reconstitution des données;

les paramètres dont les valeurs sont à estimer;

le critère d'optimisation (somme des carrés des écarts entre grandeurs mesurées et valeurs recalculées avec ou sans normalisation).

En chimiométrie il convient de déterminer les relations qui existent entre ces différentes variantes et de choisir le (ou les) modèle(s) le(s) mieux adapté(s).

Dans le cas de la chromatographie en phase gazeuse, nous avons mis en évidence³⁸ les résultats suivants:

I.A.C.P. par diagonalisation de la matrice de corrélation présente l'avantage de donner des résultats identiques, quelle que soit la grandeur de rétention utilisée. Par ailleurs, pour la détermination des axes factoriels, cette A.C.P. attribue la même importance aux variables à fortes ou faibles variances, c'est-à-dire aux phases stationnaires qui retiennent fortement ou peu les produits étudiés;

selon les grandeurs de rétention choisies, les A.C.P. par diagonalisation de la matrice de covariance donnent des résultats différents; cependant, ce type d'A.C.P. a l'avantage de présenter un critère d'optimisation identique à celui utilisé en régression multiple. Ce dernier type d'A.C.P. appliqué aux indices de rétention (I) (respectivement au logarithme de temps réduit, $\log t'$) donne des résultats pratiquement identiques à ceux d'une A.C.P. par diagonalisation des produits scalaires sur des différences d'indices δI (ou sur des logarithmes de temps réduits relatifs $\log t'_R$). L'identité serait totale si le composé de référence (le benzène dans notre cas) était parfaitement recalculé dans l'A.C.P. initiale.

Dans la pratique, le temps de rétention réduit du benzène ne peut être déterminé avec une très grande précision, compte-tenu des caractéristiques techniques de notre appareillage et des températures de travail qui sont liées aux propriétés des substances analysées.

Cette imprécision se répercutant sur les valeurs de rétention réduites relatives des composés étudiés, il nous a semblé préférable d'appliquer l'A.C.P. par diagonalisation de la matrice de covariance aux logarithmes de temps réduit ($\log t'$); on observe alors que la précision des valeurs recalculées reste compatible avec la précision expérimentale.

CONCLUSION

Les méthodes de l'analyse des données permettent d'aborder avec fruit le problème de la prévision des grandeurs de rétention en chromatographie en phase gazeuse. Wold⁴³ a réalisé des A.C.P. à l'aide des indices de Kovats par diagonalisation de la matrice de corrélation; les régressions résultant de ce traitement sont en général affines. Nous-mêmes avons pu montrer que l'analyse en composantes principales (A.C.P.), et plus particulièrement sa variante issue de la diagonalisation de la matrice de covariance, conduit à un espace tridimensionnel. Nous avons ainsi été amenés à proposer le modèle

$$x_{P,\varphi_i} = a_\varphi \cdot x_{P,\varphi_1} + b_\varphi \cdot x_{P,\varphi_2} + c_\varphi \cdot x_{P,\varphi_3} + d_\varphi$$

TABLEAU V

APPLICATION DE LA RELATION 4 AUX DONNÉES DE McREYNOLDS

$$I_{RZ,\varphi} = a_{1,\varphi} \cdot I_{RZ,SE-30} + b_{1,\varphi} \cdot I_{RZ,DC1265F} + c_{1,\varphi} \cdot I_{RZ,DEGS} + d_{1,\varphi}$$

R = *tert.*-C₄H₉, *iso*-C₄H₉, *sec.*-C₄H₉, *n*-C₄H₉, *n*-C₅H₁₁, *n*-C₆H₁₃ et Z = OH, CHO, COCH₃, COOCH₃, OCH₃

φ_i	$a_{1,\varphi}$	$b_{1,\varphi}$	$c_{1,\varphi}$	$d_{1,\varphi}$	r_φ	$\sigma_{r,\varphi}$
Apiezon L	1.255	-0.010*	-0.119*	- 57.63	0.9814	29.0
Carbowax 20M	0.518	-0.324	0.856	- 141.23	0.9986	9.2
DEGA	0.277	-0.178	0.917	- 95.89	0.9993	6.7
DEGSeb	0.565	-0.163	0.618	- 78.11	0.9988	7.7
Diéthylhexylsébacate	0.773	-0.061*	0.224	- 13.70	0.9986	6.9
Didecyl phtalate	0.711	0.058*	0.198	- 22.38	0.9982	8.5
EGS	0.636	-0.186	0.581	- 84.18	0.9986	8.4
NPGS	0.372	-0.054*	0.679	- 81.54	0.9990	7.3
PMPE 5R	0.693	0.199	0.144	- 15.41	0.9989	7.2
UCON 50 HB 2000	0.688	-0.292	0.640	- 124.63	0.9991	6.6
XF-1150	-0.117*	0.359	0.738	- 108.30	0.9966	15.5
Tricrésyl phosphate	0.589	-0.011*	0.423	- 54.97	0.9980	9.8

TABLEAU VI

APPLICATION DE LA RELATION 4 AUX DONNÉES DE ROHRSCHEIDER

$$I_{RZ,\varphi} = a_{1,\varphi} \cdot I_{RZ,DC-200} + b_{1,\varphi} \cdot I_{RZ,QF-1} + c_{1,\varphi} \cdot I_{RZ,DEGS} + d_{1,\varphi}$$

φ_i	$a_{1,\varphi}$	$b_{1,\varphi}$	$c_{1,\varphi}$	$d_{1,\varphi}$	r_φ	$\sigma_{r,\varphi}$
Squalane	1.144	-0.071*	-0.059	- 4.013	0.9962	13.4
Apiezon L	1.110	-0.109*	0.012*	- 5.080	0.9927	18.1
Diéthylhexylsébacate	0.866	-0.100*	0.185	21.228	0.9957	12.4
Célanèse ester NR9	0.842	-0.081*	0.202	11.886	0.9965	11.3
Diisodécylphtalate	0.736	0.074*	0.160	10.033	0.9965	11.1
DC-710	0.705	0.201	0.121	- 11.308	0.9949	14.1
Polypropylèneglycol	0.815	-0.225	0.345	22.350	0.9922	17.3
Acétyltributylcitrate	0.612	0.065*	0.297	3.807	0.9947	14.3
Tricrésylphosphate	0.562	0.040*	0.371	8.969	0.9959	13.1
PMPE	0.639	0.206	0.229	- 36.985	0.9911	19.9
Marlophène 87	0.603	-0.140*	0.472	19.780	0.9931	17.6
Polypropylènesébacate	0.630	-0.081*	0.434	- 2.142	0.9973	11.1
Marlophène 814	0.552	-0.174	0.560	19.480	0.9938	17.9
NPGS	0.321	0.040*	0.605	5.300	0.9957	16.3
XF-1150	-0.015*	0.433	0.543	4.811	0.9961	15.9
Carbowax 20M	0.442	-0.210	0.755	- 13.486	0.9923	24.0
Carbowax 4000	0.433	-0.222	0.775	- 14.974	0.9931	23.1
Réoplex 400	0.334	-0.106	0.785	- 22.634	0.9981	12.5
Éthylèneglycol-cyanoéthyléther	-0.368	0.492	0.953	- 64.168	0.9988	13.1
Tris1,2,3cyano-2-éthoxypropane	-0.510	0.462	1.096	- 35.029	0.9988	14.3

comme outil de prévision des grandeurs de rétention. Ce modèle, mis au point à partir des données relatives à des benzènes monosubstitués, présente un caractère plus général; en effet, son domaine de validité peut être étendu non seulement à des benzènes polysubstitués, mais même à des populations de structure quelconque.

Par ailleurs, les résultats obtenus montrent très nettement que les phases polyfluorées se distinguent dans leur comportement chromatographique tant des phases dites "polaires" que de celles qui sont "apolaires". Elles constituent par conséquent une classe particulière.

RÉSUMÉ

L'analyse en composantes principales appliquée à un ensemble de données chromatographiques recueillies sur 21 phases stationnaires et relatives à une population de 17 benzènes monosubstitués permet de déduire un modèle multiparamétrique simple. La relation qui en découle

$$x_{P,\varphi_i} = a_\varphi \cdot x_{P,\varphi_1} + b_\varphi \cdot x_{P,\varphi_2} + c_\varphi \cdot x_{P,\varphi_3} + d_\varphi$$

présente un caractère de généralité. Elle permet d'estimer la grandeur de rétention x_{P,φ_i} d'un produit P sur une phase φ_i à partir des valeurs expérimentales des grandeurs de rétention x_{P,φ_1} , x_{P,φ_2} et x_{P,φ_3} de ce même produit P, relevées sur trois phases stationnaires φ_1 , φ_2 et φ_3 , respectivement apolaire, polyfluorée et polaire.

BIBLIOGRAPHIE

- 1 G. Castello, M. Lunardelli et M. Berg, *J. Chromatogr.*, 76 (1973) 31; et réf. citées.
- 2 K. Boček, *J. Chromatogr.*, 162 (1979) 209; et réf. citées.
- 3 G. Castello et G. D'Amato, *J. Chromatogr.*, 107 (1975) 1.
- 4 B. L. Karger, Y. Elmeirik et R. L. Stern, *Anal. Chem.*, 40 (1968) 1227.
- 5 B. L. Karger, Y. Elmeirik et W. Andrade, *J. Chromatogr. Sci.*, 7 (1969) 209.
- 6 D. Bonchev, Ov. Mekenjan, G. Pročić et N. Trinajstić, *J. Chromatogr.*, 176 (1979) 149.
- 7 J. N. Tejedor, *J. Chromatogr.*, 177 (1979) 279.
- 8 M. Gassiot-Matas et G. Firpo-Pamies, *J. Chromatogr.*, 187 (1980) 1.
- 9 G. H. E. Nieuwdorp, C. L. de Ligny et N. G. van der Veen, *J. Chromatogr.*, 154 (1978) 133.
- 10 S. D. West et R. C. Hall, *J. Chromatogr. Sci.*, 14 (1976) 339.
- 11 E. Bojti, M. Mihók, J. Borbély, J. Bárkai et J. M. Takács, *J. Chromatogr.*, 119 (1976) 321.
- 12 T. S. Lombosi, E. R. Lombosi, I. Bernát, ZS. SZ. Bernát, E. C. Takács et J. M. Takács, *J. Chromatogr.*, 119 (1976) 307.
- 13 M. Yamada, N. Nomura et Den-Itsu Shiho, *J. Chromatogr.*, 64 (1972) 253.
- 14 R. Fellous, L. Lizzani-Cuvelier, R. Luft et J. P. Rabine, *J. Chromatogr.*, 110 (1975) 13.
- 15 M. Chastrette, G. Lenfant, A. Remy et M. Cohen-Makabeh, *J. Chromatogr.*, 84 (1973) 275.
- 16 J. E. Dubois, D. Laurent et A. Aranda, *J. Chim. Phys.*, (1973) 1608; et réf. citées.
- 17 J. E. Dubois, D. Laurent et A. Aranda, *J. Chim. Phys.*, (1973) 1616; et réf. citées.
- 18 P. H. Weiner et D. G. Howery, *Can. J. Chem.*, 50 (1972) 448.
- 19 P. H. Weiner et D. G. Howery, *Anal. Chem.*, 44 (1972) 1189.
- 20 P. H. Weiner, C. J. Dack et D. G. Howery, *J. Chromatogr.*, 69 (1972) 249.
- 21 D. G. Howery, P. H. Weiner et J. S. Bliuder, *J. Chromatogr. Sci.*, 12 (1974) 366.
- 22 P. T. Funke, E. R. Malinowski, D. E. Martire et L. Z. Pollara, *Sep. Sci.*, 1 (1966) 661.
- 23 R. B. Selzer et D. G. Howery, *J. Chromatogr.*, 115 (1975) 139.
- 24 J. H. Kindsvater, P. H. Weiner et T. J. Klíngen, *Anal. Chem.*, 46 (1974) 982.
- 25 D. G. Howery, *Anal. Chem.*, 46 (1974) 829.
- 26 J. F. K. Huber, E. T. Alderlieste, H. Harren et H. Poppe, *Anal. Chem.*, 45 (1973) 1337.
- 27 P. H. Weiner et J. F. Parcher, *Anal. Chem.*, 45 (1973) 302.

- 28 M. Chastrette, *J. Chromatogr. Sci.*, 14 (1976) 357.
- 29 S. Wold et K. Andersson, *J. Chromatogr.*, 80 (1973) 43.
- 30 D. H. Mac Closkey et S. J. Hawkes, *J. Chromatogr. Sci.*, 13 (1975) 1.
- 31 S. R. Lowry, G. L. Ritter, H. S. Woodruff et T. L. Isenhour, *J. Chromatogr. Sci.*, 14 (1976) 126.
- 32 D. L. Massart, P. Lenders et M. Lauwereys, *J. Chromatogr. Sci.*, 12 (1974) 617.
- 33 A. Eskes, F. Dupuis, A. Dijkstra, H. de Clercq et D. L. Massart, *Anal. Chem.*, 47 (1975) 2168.
- 34 F. Dupuis et A. Dijkstra, *Anal. Chem.*, 47 (1975) 379.
- 35 L. Rohrschneider, *J. Chromatogr.*, 22 (1966) 6.
- 36 W. O. McReynolds, *Gas Chromatographic Retention Data*, Preston Technical Abstracts Company, Evanston, IL, 1966.
- 37 R. Fellous, D. Lafaye de Micheaux, L. Lizzani-Cuvelier et R. Luft, *J. Chromatogr.*, 213 (1981) 223.
- 38 L. Lizzani-Cuvelier, résultats non publiés, disponibles sur demande.
- 39 W. O. McReynolds, *J. Chromatogr. Sci.*, 8 (1970) 685.
- 40 W. J. Dixon, *Biomedical Computer Programs*, University of California Press, Berkeley, CA, Version 1981.
- 41 *Chaîne d'analyse de données "Microchadoc"*, Laboratoire d'Informatique, Nice, 1981.
- 42 L. Rohrschneider, *Advan. Chromatogr.*, 4 (1967) 333.
- 43 S. Wold, *Technometrics*, 20 (1978) 397.