Journal of Chromatography, 609 (1992) 261–267
Elsevier Science Publishers B.V., Amsterdam

CHROM. 24 372

Factor analysis and experimental design in high-performance liquid chromatography

XIII.[☆] Correspondence factor analysis of the heterogeneity of the behaviour of congeneric solutes series studied on homogeneous OV phases in gas chromatography

Jacques R. Chrétien

Laboratoire de Chimiométrie, Université d'Orléans, 45046 Orléans Cédex (France) and Institut de Topologie et de Dynamique des Systèmes, Associé au CNRS UA 34, Université Paris VII, 1 Rue Guy de la Brosse, 75005 Paris (France)

Michel Righezza

Laboratoire de Chimiométrie, Université d'Orléans, 45046 Orléans Cédex (France)

Aicha Hassani and Brahim Y. Meklati

Laboratoire d'Analyse Organique Fonctionnelle Systématique, Institut de Chimie, USTHB, El-Alia BP 32, Bab Ezzouar, Algiers (Algeria)

(Received March 13th, 1992)

ABSTRACT

A correspondence factor analysis (CFA) of gas chromatographic retention data, *i.e.*, Kováts retention indices, is presented. The data matrix correspond to aromatic congeneric solutes series, monosubstituted benzenes, benzaldehydes or acetophenones, studied on the same type of stationary phases, namely phenylmethylsilicone with various percentages of the phenyl substituent (0–75%). CFA of three reduced submatrices gives a trend analysis of the relative behaviour of solutes and of the relative influence of stationary phases and temperatures. CFA was found to be a sensitive detector of the relative heterogeneity between solutes and/or chromatographic systems, even for the study of congeneric compound series studied on stationary phases of the same type.

INTRODUCTION

Most often principal component analysis (PCA) is used to search for similarities in the chromatographic behaviour of heterogeneous series of compounds [1]. However, PCA is not suitable for the analysis of closely similar data, such as chromatographic retention data relative to homogeneous or congeneric series of compounds. When there is a major factor that governs the chromatographic retention, *e.g.*, the vapour pressure in gas-phase chromatography, the contributions of the remaining factors are hidden. In such types of factor analysis,

---

Correspondence to: Dr. J. R. Chrétien, Laboratoire de Chimiométrie, Université d'Orléans, 45046 Orléans Cedex, France.

[☆] For part XII see B. Walczak, M. Dreux and J. R. Chrétien, *Chromatographia*, 31 (1991) 575–582 and for part. XI see ref. 11.

0021-9673/92/$05.00

about 99% of the information content of the data matrix is taken into account with the first axis [2] and all the variables are strongly correlated.

The less often used correspondence factor analysis (CFA) is more appropriate for the analysis of homogeneous populations [2,3]. It can be used just as a classical clustering technique or as a detector of fine physico-chemical effects influencing the chemical activity of the studied compounds and/or of the selectivity of the stationary phases. CFA has been used, for example, for the analysis of a large set of Kováts retention indices of esters [2] and hydrocarbons [3]. In the latter instance, interesting results were obtained for simulation of the catalysis of hydrocarbons, for comparison of the interactions of zinc and nickel ions with alkenes, the data later being made more precise by topological analysis [4–6]. Such studies partially explain the interest devoted to factor analysis and to CFA in chromatographic structure–activity relationships (CSAR) with their extension to the field of liquid chromatography [4,7].

It must be emphasized that, like polarity, the heterogeneity of solutes and/or stationary phases is a relative concept. CFA has proved to be a valuable tool for evaluating the heterogeneities nested in a chromatographic retention data matrix [3], but it must be questioned whether this technique is able also to detect potential heterogeneity of more homogeneous chromatographic systems. For example, let us consider the previous example of esters [2]. This involves a large set of 1450 Kováts retention indices of 175 saturated and unsaturated aliphatic esters determined of fourteen polysiloxane stationary phases which were analysed by factor analysis. The series of compounds was congeneric, but variation in the polarity of the stationary phases was obtained with different percentages of different types of substituent groups, phenyl, cyanoethyl, trifluoroalkyl, trifluorophenyl, etc. The question now arises of whether the application of CFA could have been extended to the case of a homogeneous series of stationary phases by varying only the percentage of the same substituent to obtain phases of different polarities.

To go beyond the practical limit of application of CFA presented so far, for a homogeneous series of compounds, we selected a data matrix of Kováts retention indices of aromatic compounds determined on a homogeneous series of seven OV phenyl stationary phases. This matrix corresponds to 35 unsubstituted or monosubstituted benzene, aldehyde and acetophenone compounds, studied at different temperatures. A systematic chemometric analysis of heterogeneity inside homogeneous systems is presented with progressive CFA analysis of this large data set. The study of three complementary matrices will offer the possibility of explaining more subtle chromatographic factors.

DATA

The 35 compounds, presented Fig. 1, were eleven monosubstituted benzenes [including the unsubstituted benzaldehyde (AH) and acetophenone (CH)], fourteen *o*-, *m*- and *p*-substituted aldehydes and ten *o*-, *m*- and *p*-substituted acetophenones. These compounds were studied on seven OV methylphenylsilicone phases. The phenyl compounds were studied in the range 80–140°C and the disubstituted phenyl compounds in the range 120–160°C [8,9].

Three different data matrices were studied. The first corresponds to 33 compounds of the three subseries: benzenes, aldehydes and acetophenones derivatives, that is, except the two compounds labelled BF and AoOM. These compounds were studied on all the OV stationary phases, except OV-22. The temperatures considered were 120°C for all these six phases plus 130 and 140°C for OV-11 and 130°C for OV-25. Hence this first data matrix regroups the data relative to 33 compounds studied in nine different chromatographic systems, including six different stationary phases and three temperatures effects.

The other two data matrices correspond to more homogeneous data sets. In the second matrix, the temperature is limited to 120°C. This matrix corresponds to all 35 compounds indicated in Fig. 1. In the third matrix, the subseries of compounds is limited to all the aldehydes and acetophenones, *i.e.*, to 26 compounds including the unsubstituted aldehyde (AH) and acetophenone (AC). This last case offers the possibility of treating the largest set of the chromatographic system, the seven OV phases considered at four different temperatures, 120, 130, 140 and 160°C, minus OV-25 at 140°C, that is, 27 different chromatographic systems.

a

Substituents:

X	H	F	CH_3	Cl	OCH_3	Br	OH	OC_2H_5	NO_2
Label:	H	F	M	Cl	OM	Br	OH	OE	NO

Compounds:

B :	BH	BF	BM	BCl	BOM	BBr	BOH	BOE	BNO
A :	AH	AoF AmF ApF	AoM AmM ApM	AoCl AmCl ApCl	AoOM AmOM ApOM	AoBr AmBr			
C :	CH	CoF CpF	CoM CmM CpM	CoCl CpCl	 CpOM	 CpBr	CoOH		

B = $X-C_6H_5$ **A** = $X-C_6H_4-CHO$ **C** = $X-C_6H_4-CO-CH_3$

b

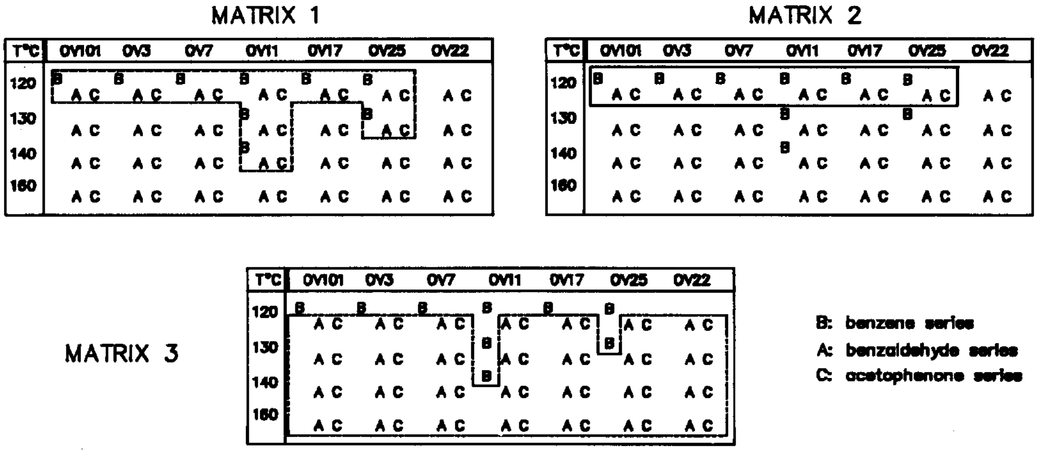

Fig. 1. (a) The 35 substituted phenyl compounds: benzenes (B), benzaldehydes (A) and acetophenones (C). The label for a compound consists of the label of its substructure A, B or C, plus the position of the substituent, *ortho* (*o*), *meta* (*m*) or *para* (*p*) for the disubstituted ones, plus the label of the substituent X. (b) Table of the experimental design: chromatographic system and considered data matrices. The columns of the matrix indicate the stationary phase type and the rows indicate the temperature. The compounds series are indicated by their labels A, B or C. The submatrices submitted to CFA are indicated by a dashed-line box for the first, a solid-line box for the second and a dot-dashed-line box for the third.

DATA PROCESSING

Factor analysis is particularly effective as a method of extracting information from large data sets. Among the variants of this technique CFA, developed by Benzécri [10], is more suitable for the analysis of relatively homogeneous series of data. The basic principles of CFA have been presented previously [3] and will not be repeated here. Owing to the normalizations used in this method, the compounds and chromatographic variables play a symmetric role which allows the superposition of both projections on the same graph (Fig. 2–4). Relative proximities of compounds and chromatographic systems in the considered factorial planes are informative, even if they must be handled with caution [11].

RESULTS AND DISCUSSION

The CFA of the first data matrix gives rise to the first factorial plane presented Fig. 2. This plane includes 68% of the information content, 47% according to the first and 21% according to the second axis. This first axis is determined by the nonpolar OV-101, a 0% phenylsilicone, OV-3, only a 10% phenylsilicone, and the most polar OV-25, a 75% phenylsilicone, all studied at 120°C. This axis reflects the increase in the average polarity of solutes and phases.

The second best axis of inertia is determined by OV-17 at 120°C and OV-11 at 140°C. OV-11 at 120°C shows an average behaviour and is projected near the centre of the graph. Axis 2 reflects the temperature effect for medium-polarity phases. The temperature effects on OV-11 show greater variations according to axis 2 than with the polar OV-25.

The compounds are projected on axis 1 according to their polarity. Methylbenzene is on the left-hand side, whereas nitrobenzene is on the right, according to the variation of the electronegativity of the corresponding substituents.

The monosubstituted benzenes offer a larger variation of dipolar moments than their disubstituted acetophenone and aldehyde analogues. A larger cloud is observed for the representative points of the monosubstituted benzene derivatives. These compounds contribute the most to axes 1 and 2. Two restricted clouds are observed for the carbonylated series which show a restricted variation of their dipolar moments.

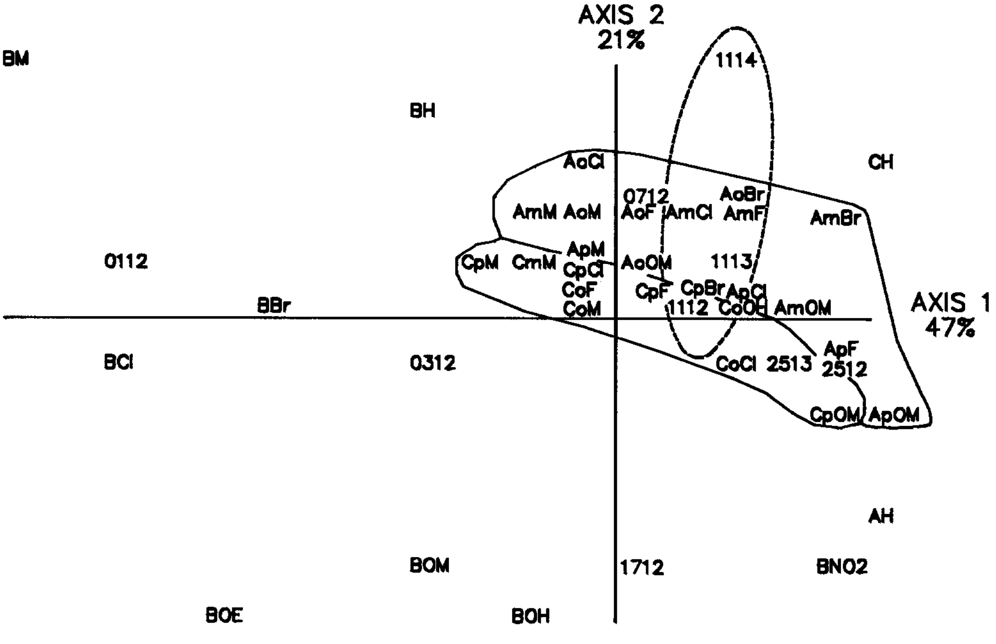

Fig. 2. CFA of 34 compounds of the three subseries with three temperatures effects on OV-11 and OV-25. Projection of the first matrix on the first factorial plane. The two disubstituted series acetophenone (C) and benzaldehyde (A) are circled with solid lines. In CFA projections, the label of the chromatographic systems is composed of four figures: the first two characterize the stationary phase and correspond to the value *n* taken from the name OV-*n*, *i.e.*, 01, 03, 07, 11, 17, 22, 25 for OV-101, OV-3, ..., up to OV-25, respectively; the two last figures characterize the temperature and correspond to the number of ten degree steps, *i.e.*, 12, 13, 14 and 16 for 120, 130, 140 and 160°C, respectively. The temperature effects are circled with a dashed line for OV-11.

The study of this large set of compounds and phases used with some temperatures effects shows that differences in behaviour can be detected at different levels. The CFA detects differences with an average behaviour calculated from the experimental data matrix. The compounds or phases that have this average behaviour are projected at the centre of the graph. The compounds or phases that can be deduced from each other by a proportionality relationship have the same projection. In Fig. 2, the compounds or phases that have different projections, *i.e.*, with no proportionality relationships between them, are significantly different.

The projection area of the benzene derivatives is more scattered than that for the other two subseries of carbonylated compounds. However, the temperature effects must also be taken into account. These temperature effects partially hide the potential structural variations between aldehydes and acetophenones.

A second matrix without temperature effects must be considered. This matrix includes the six OV phases used at the same temperature of 120°C. The corresponding CFA is presented Fig. 3 in the form of the first factorial plane. This factorial plane includes 79% of the information content, 57% for axis 1 and 22% for axis 2. The information content of the similar first factorial plane for the preceding matrix was 67% owing to a greater dispersion of the information induced by two supplementary temperatures corresponding to three supplementary chromatographic systems (OV-11 at 130 and 140°C and OV-25 at 130°C). This graph, corresponding only to Kováts retention indices determined at 120°C, shows the same trends. Axis 1, determined by OV-101 and OV-25, reflects polarity. The projection order of compounds on the first axis is similar. The greatest variations of dipolar moment are for the substituted benzenes. They induce a greater dispersion of the representative points of the compounds. Axis 2 offers a better discrimination of the three series of compounds. The most deactivated compounds, *i.e.*, benzaldehydes, acetophenones and nitrobenzene, are on the right-hand part of the diagram and the less polar compounds, with donor substituents, are on the left.

To improve the sensitivity of the CFA analysis for the partially hidden disubstituted compounds, a new matrix will be considered. The set of monosubstituted benzenes which contribute the most to

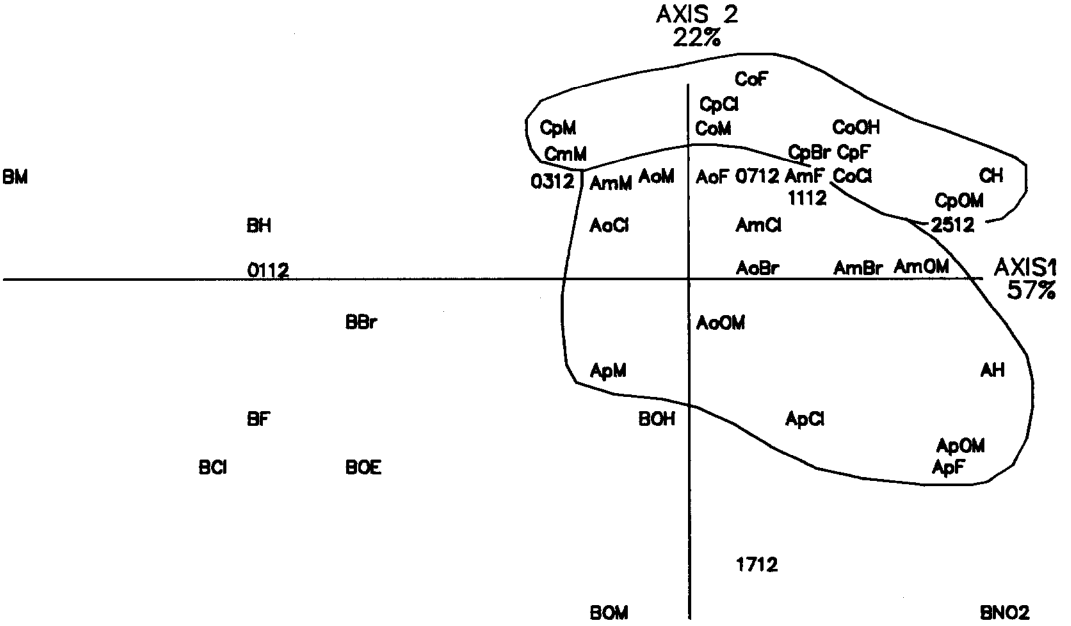

Fig. 3. CFA of the second matrix including the three subseries without temperature effects. The projection on the first factorial plane shows a better discrimination between the two substituted series, acetophenone and benzaldehyde.

the information content in the two previous analyses are removed. Then only the two more similar carbonylated subseries remain. A more extensive study of temperature effects can be envisaged simultaneously.

The third matrix is then based on a more homogeneous population including only carbonylated compounds, *i.e.*, substituted benzaldehydes or acetophenones. It includes a wider heterogeneity dealing with temperature effects by using the seven stationary phases OV-101, OV-3, OV-7, OV-11, OV-17, OV-22 and OV-25 at four temperatures, 120, 130, 140 and 160°C, except for OV-25 at 140°C. This represents a set of 26 compounds studied in 27 different chromatographic systems.

Projection on the first factorial plane is given Fig. 4. The plane includes 72% of the information content, 45% according to the first and 27% according to the second axis. This graph presents an interesting relationship order. A clear separation of the two series of compounds is observed. Comparison with Fig. 3, corresponding to the second matrix, shows an apparent rotation of about 45° of the compound clusters due to the absence of monosubstituted phenyl compounds. The comparison of Figs. 3 and 4 shows that the first axis is rotated counterclockwise. The better separation of the two series is due to the absence of the levelling effect of the monosubstituted phenyl series, with its important variation of the electronic effects of the substituent reflected in the variation of the dipolar moments.

The two populations, aldehydes and acetophenones, constitute two clouds admitting the origin of the axes 1 and 2 as a centre of symmetry, or, more precisely, the first bisector as an axis of symmetry. The better separation of the two chemical series according to their corresponding cloud is given by the first bisector. The variation of the polarity of the stationary phases is introduced by the second bisector, from the non-polar OV-101 up to the most po-

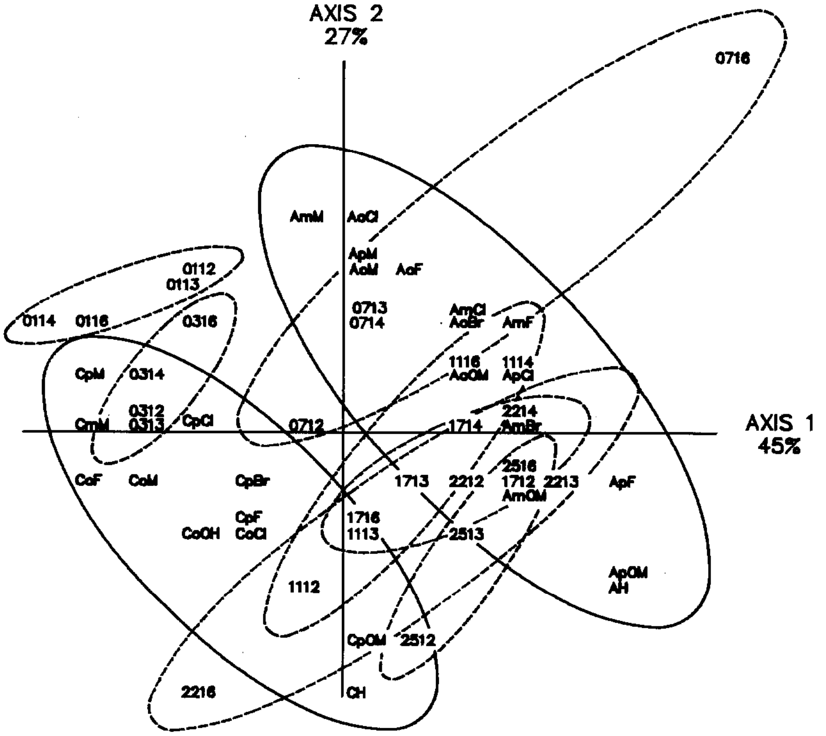

Fig. 4. CFA of the third matrix limited to carbonylated compounds with maximum temperature effects. Projection on the first factorial plane. A total discrimination is observed between the two disubstituted series acetophenone and benzaldehyde, circled with solid lines. Temperature effects, circled with dashed lines, are oriented according to the first bisector.

lar OV-25. Comparison of Figs. 2 and 4 shows that axes 1 and 2 in Fig. 2 play a similar role to the first and second bisectors, respectively. We have shown in Fig. 2 that the temperature effect with OV-11 was parallel to axis 2. In Fig. 4, the temperature effect is parallel to the corresponding first bisector. The general trends of the temperature effects are directed according to the first bisector. Increasing temperatures are noticed in this direction except for the non-phenyl-substituted OV-101, a methylsilicone, and for OV-22 studied at 160°C, labelled 2216, which shows a relatively large DI/dT variation in the range 140–160°C. For these two exceptions the directions of these temperatures effects are opposite to the preceding ones.

We could have considered also a "supplementary" Fig. 2, not given here, which corresponds to a CFA of the first matrix but with a projection on the second factorial plane, defined by factorial axes 3 and 4. The analogy between this "supplementary" figure and Fig. 4 must be stressed. In Fig. 2, the major contribution of the monosubstituted phenyl derivatives to the information content is taken into account by axes 1 and 2. Hence the resulting information corresponding to the two disubstituted phenyl derivative series is given chiefly and more clearly in the "supplementary" Fig. 2. In Fig. 4 the corresponding data matrix is strictly independent of the monosubstituted benzene series. This first factorial plane is a pure contribution of the aldehydes and acetophenones studied at different temperatures. Despite three supplementary temperatures (OV-11 at 130 and 140°C and OV-25 at 130°C) in Fig. 2, the level of similarities between the CFA of matrices 1 and 3, that is, between the "supplementary" Fig. 2 and Fig. 4, was not so evident, *a priori*.

The influence of a relatively limited effect on Kováts retention indices, such as the temperature effect, introduced only by three supplementary temperature effects can produce non-negligible perturbation. On comparing Fig. 2 and 3, a modification of the ordering of the three chemical series is observed. In Fig. 3, the cluster of acetophenones is lost among the aldehydes. The contribution of axis 3, *i.e.*, factorial plane 2-3, must be considered also to observe a clear separation between these three chemical series.

CONCLUSION

Despite the strong homogeneity of the data matrix, due to the choice of congeneric series of aromatic solutes analysed on the same type of phenylsilicone stationary phases, the present study demonstrate that CFA is a sensitive detector of the relative heterogeneity between solutes and/or stationary phases. The so-called heterogeneity corresponds to deviations from proportionality between rows or columns of the data matrix, that is, between particular compounds or between particular chromatographic variables.

The posibility of taking homogeneous systems, such as congeneric compound series and stationary phases of the same type, offers the possibility of identifying weak structural, electronic, polarity or temperature effects. To emphasise this point, let us consider the case of another experimental design, with a higher degree of heterogeneity of the compounds and/or of the stationary phases. The projection of the compounds considered here would have been much less scattered and the corresponding factors, detected here, more or less hidden.

A progressive study with CFA of particularly chosen submatrices helps to delineate more precisely the abstract and/or the real factors which govern the chromatographic process.

REFERENCES

1 G. Musumarra, G. Scarlata, G. Romano, S. Clementi and S. Wold, *J. Chromatogr. Sci.*, 22 (1984) 538–547.
2 K. Szymoniak and J. R. Chrétien, *J. Chromatogr.*, 404 (1987) 11–22.
3 R. F. Hirsch, R. J. Gaydosh and J. R. Chrétien, *Anal. Chem.*, 52 (1980) 723.
4 J. R. Chrétien, *Trends Anal. Chem.*, 6 (1987) 275–278.
5 J. R. Chrétien, J. E. Dubois, R. F. Hirsch and R. J. Gaydosh, *J. Chromatogr.*, 207 (1981) 115–130.
6 J. R. Chrétien, K. Szymoniak, J. E. Dubois, R. F. Hirsch and R. J. Gaydosh, *J. Chromatogr.*, 294 (1984) 1–14.
7 R. J. Kalishan, *Quantitative Structure Chromatographic Retention Relationships*, Wiley, New York, 1987, p. 224.
8 A. Hassani and B. Y. Meklati, *Chromatographia*, 33 (1992) 267–272.
9 A. Hassani and B. Y. Meklati, in preparation.
10 J. P. Benzécri, *L'Analyse des Données*, Vol. 2, Dunod, Paris, 1973.
11 M. Righezza and J. R. Chrétien, *J. Chromatogr.*, 556 (1991) 169–180.