

This article was downloaded by:

On: 24 January 2011

Access details: *Access Details: Free Access*

Publisher *Taylor & Francis*

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of Liquid Chromatography & Related Technologies

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t713597273>

DATA ANALYSIS. A BRIDGE BETWEEN FACTOR ANALYSIS AND CHROMATOGRAPHY

S. Ounnar^a; M. Righezza^a

^a Institut de Chimie Organique et Analytique, CNRS UPRES-A 6005, Université d'Orléans, Orléans, Cedex 2, France

Online publication date: 10 May 1999

To cite this Article Ounnar, S. and Righezza, M.(1999) 'DATA ANALYSIS. A BRIDGE BETWEEN FACTOR ANALYSIS AND CHROMATOGRAPHY', *Journal of Liquid Chromatography & Related Technologies*, 22: 17, 2575 – 2594

To link to this Article: DOI: 10.1081/JLC-100102045

URL: <http://dx.doi.org/10.1081/JLC-100102045>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

DATA ANALYSIS. A BRIDGE BETWEEN FACTOR ANALYSIS AND CHROMATOGRAPHY

S. Ounnar, M. Righezza*

Institut de Chimie Organique et Analytique
CNRS UPRES-A 6005
Université d'Orléans
B. P. 6759
45067 Orléans Cedex 2, France

ABSTRACT

Factor analysis has long been applied to the systematic investigation of chromatographic data. In this paper, two factor analyses, Principal Component Analysis and Correspondence Factor Analysis, are explained. The two approaches are applied to a trial set of chromatographic data. In the first part, the mathematical processes are shown. Particular attention is paid to the data pre-processing with correspondence factor analysis. Next, the results obtained from the two factor analyses are illustrated with two graphs, which are compared to the corresponding tridimensional graphs. Finally, the physico-chemical meaning of the two factor analyses are explained with the use of physico-chemical properties of solutes.

INTRODUCTION

Analytical chemistry has changed profoundly over the last few decades. The advent of intelligent instruments and laboratory automation has led analytical chemistry to information science.

Larger amounts of data are being collected and the rate will continue to accelerate in coming decades. It is now necessary to work on bigger data sets and to extract from them all accessible information.

One way to work on large data sets is to apply suitable data analysis. Data analysis methods have been proposed for a long time under the generic name of chemometrics. The earliest work, by K. Pearson¹ and C. Spearman,² dates from the beginning of the 19th century. Since this time, chemometrics has evolved and developed powerful tools for the investigation of large and complex data sets.³

Now, many chemometrics applications have been published in various domains such as analytical chemistry,⁴⁻⁷ therapeutic chemistry,⁸ biologic chemistry,^{9,10} geochemistry,^{11,12} food chemistry,¹³ spectroscopy,¹⁴ and geology.¹⁵

Chemometric techniques are based on mathematical and statistical processes. They aim to condense large assembly projects into more manageable time frames. Their modeling capability allows users to speed up methods development and the interpretation of complex data. For example, chemometrics can be used to accomplish a variety of goals in chromatography such as accelerating methods development, giving more effective multivariate calibration, or improving the detection and the monitoring of impurities.

One of the chemometric techniques, successfully applied in chromatography, is factor analysis. Factorial analysis was first used as a prospective implement^{16,17} to establish relationships between retention and various physico-chemical properties of solutes. Now, the use of factorial analysis methods are more widely applied and affects various studies about selectivity of stationary phases,^{18,19} identification of drugs (or pharmaceuticals) with thin layer chromatography,²⁰ or peaks deconvolution in liquid chromatography.²¹

Several different techniques of factor analysis used in the field of chromatography have been developed. In the thirties, H. Hotelling laid the foundations of principal component analysis (PCA).²² J. P. Benzécri's work^{23,24} has provided the bases for the development of correspondence factor analysis. Until the sixties, these methods were perfected but remained inaccessible because of the huge computing power necessary. The appearance and development of both hardware and software have permitted the popularization of these techniques.

Now, chromatographers may be eager to investigate factor analysis in the study of chromatographic processes (solute behavior, solvent effects...). First, they are faced with great difficulty in understanding and in applying chemometric methods to their own data. Next, chromatographers may have trouble interpreting the chemometric results. Some chromatographers have suggested "building a bridge between factor analysis and chromatography."

The present paper aims to explain factor analysis especially for chromatographers. The main purpose is to detail graphically the mathematical process of factor analysis through the study of chromatographic data.

Finally, the factor analysis results are interpreted with the help of well known physico-chemical properties such as adsorption energy or hydrophobic parameters.

EXPERIMENTAL

The considered data were obtained from the literature.²⁵⁻²⁷ A series of substituted chalcones in position 4 or 4', corresponding to the E-s-cis and Z-s-cis isomers were studied with various chromatographic systems. The general structure is $X-C_6H_4-CH=CH-CO-C_6H_4-Y$. The substituents X and Y are listed in Table 1.

Throughout this paper, the E-s-cis isomers are referred to as X-Y and the Z-s-cis isomers as X-Y*. The original data set corresponds to 63 compounds studied with 43 chromatographic systems.

From the original data matrix of capacity factors (k'),²⁵⁻²⁷ a subset has been extracted. The reduced matrix is a set of 66 capacity factors of 22 compounds eluted with two normal phases (Diol, DNA) and one reversed phase (ODS) chromatographic system. The characteristics of the three chromatographic systems, the packing, the eluents, and the corresponding identifiers of the systems, are presented in Table 1.

DISCUSSION

Factor Analysis Methods

For the chromatographer, it is necessary to grasp the main information nested in huge series of data to give a physico-chemical meaning to the observed phenomenon. The principal interest in applying factor analysis in chromatography is to condense, extract, and represent the most important information dispersed over the content of a large data set.

The chromatographic measurements are made on several objects (i.e. solutes) under different experimental conditions (i.e. eluents, stationary phases) which are called variables. The results (i.e. retention, capacity factors...) are arranged in a table which is called a data matrix in which objects are associated with rows and variables with columns.

Table 1
Capacity Factors, k' , for 22 E-s-cis and Z-s-cis Chalcones Separated on Three Chromatographic Systems

Packing	Diol	3-(2,4-dinitro-anilinopropyl)	Zorbax ODS
Eluent	Heptane/Ethanol	Heptane/THF	Water/Methanol
	(0.5/99.5) (v/v)	(3/97) (v/v)	(30/70) (v/v)
Identifiers	DIOL	DNA	ODS
H-H	0.78	6.14	5.83
H-CF ₃	0.62	2.84	11.23
H-tBu	0.64	7.01	24.66
H-iPr	0.66	7.07	18.72
H-Et	0.70	7.77	12.80
MeO-MeO	3.44	82.46	5.83
MeO-Me	1.78	29.57	8.76
Me-MeO	1.75	31.88	9.65
F-MeO	2.03	22.57	5.36
NO ₂ -F	3.52	19.34	4.32
NO ₂ -H	2.68	18.91	4.02
H-H	0.62	3.12	3.22
H-CF ₃	0.50	1.63	7.19
H-tBu	0.53	3.65	16.63
H-iPr	0.55	3.69	11.68
H-Et	0.58	3.99	7.85
MeO-MeO	2.05	23.19	3.74
MeO-Me	1.06	9.09	5.84
Me-MeO	1.28	10.94	5.45
F-MeO	1.29	8.05	3.56
NO ₂ -F	2.43	13.67	2.93
NO ₂ -H	1.90	13.23	2.77

Normal phases labelled (DIOL, DNA) and reversed phase labelled (ODS).
 Composition of the eluent for DIOL is Heptane/Ethanol (0.5/99.5) (v/v), for DNA, Heptane/THF (3/97) (v/v) and for ODS, Water/Methanol (30/70) (v/v).
 Me = CH₃; tBu = tertiary butyl; iPr = iso propyl; Et = ethyl.

As long as data matrices are two-dimensional (two rows or two columns), they can be visualized with a two-dimensional plot in a Cartesian coordinate system. For multidimensional data matrices, this is no longer possible because of the abundance of entries and the complexity of data structure due to the interdependency between the variables and the objects. Methods of multivariate statistics are required to understand such data in their entirety and to deal

adequately with their mathematical properties. Factor analysis methods,²⁸⁻⁴¹ such as principal component analysis or correspondence factor analysis, play an important role here. Their main objectives are to display multidimensional data in a space of smaller dimensions with a minimal loss of information and to extract the basic features behind the data with the ultimate goal of interpretation and/or prediction.

Principal component analysis and correspondence factor analysis represent the data in a space extracted from the data matrix. In the present paper, both methods are applied to the chosen data set. For each method, data are first analyzed graphically in Cartesian space. Next, factor analysis is applied and the projection of the data in the factorial space is drawn. Finally both graphs are compared.

Principal Component Analysis

Principal component analysis (PCA)^{31, 41-45} is a linear algebra technique which attempts to describe the observed data on the basis of a small number of unique underlying factors. The starting point of PCA is data normalization followed by the calculation of the principal components, which define the factorial axes. The general process of PCA is given in Figure 1.

Data normalization is necessary for the analysis of multivariate data when the measurements differ greatly in scale. In this case, the results are strongly affected by the measurements which are largest in magnitude. To avoid this effect, various normalizations can be used.

The data can be scaled before any PCA processing to attribute the same a priori importance to the variables. So, subtracting each value from the average value of the concerned variable first centers the initial data matrix (Figure 1a). Then, the centered data matrix is reduced (Figure 1b) by dividing each element obtained by the standard deviation of the variable concerned. In the case of chromatographic studies, it is not necessary to reduce the data matrix because measurements do not differ in scale.

The study of the data matrix is divided into two parts. In the first one (Figure 1c), the objects are examined. The data matrix is considered as a juxtaposition of rows. The objects are represented as points in the vectorial space R^j defined by the j variables (Figure 1e).

In the vectorial space R^j , the objects constitute a cloud of data points, which is called N_j . The orthogonal directions given by the eigenvectors of the data matrix represent the inertia of the cloud N_j . These abstract factors, called principal components, define a new space in which the cloud has a new representation.

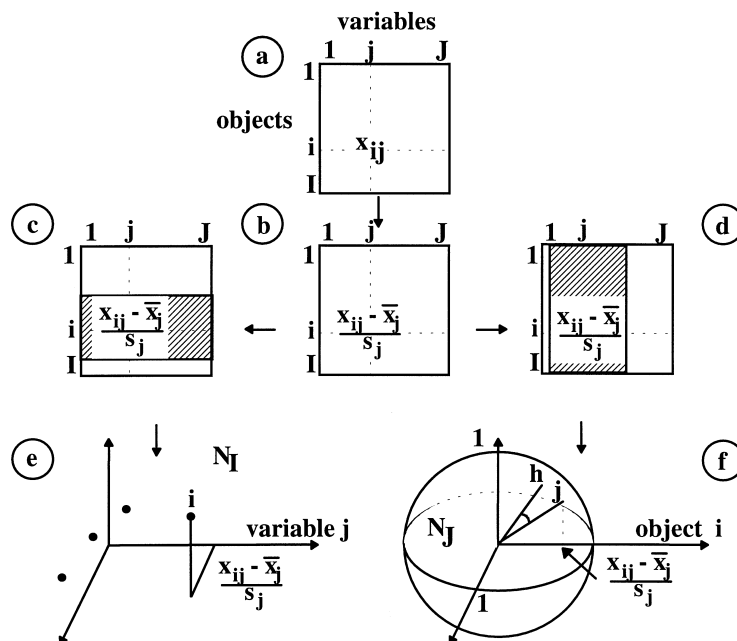


Figure 1. The general processing of the X data matrix with principal component analysis.

The principal components reproduce the original data matrix exactly. It is theoretically possible to determine as many principal components as original variables. However, only a few sets of principal components remain necessary since they retain most of the variance of the data. So, the factorial space is a reduced space of the vectorial R^J space. The first axis corresponds to the main trend; often it represents the largest part of the information content, which can be related to physico-chemical parameters. The interpretation of the other axes becomes more difficult as the information content of the axis decreases. This is due to the mathematical definition of the method; only the first axis is strictly determined by the maximum of inertia of the raw data matrix, whereas the other axes are conventionally chosen perpendicular to the preceding ones. In the R^J space, the distances between objects must be interpreted as resemblance.

The second part concerns the variables (Figure 1d). The data matrix is considered as a juxtaposition of columns. The variables are represented as vectors in vectorial space R^I defined by the i objects (Figure 1f). In this space, the extremities of vectors constitute the cloud N_J . The PCA of the N_J cloud, as for the objects, consists in a series of orthogonal directions for which the inertia of the projection of the cloud on these directions is maximal. The coordinates of

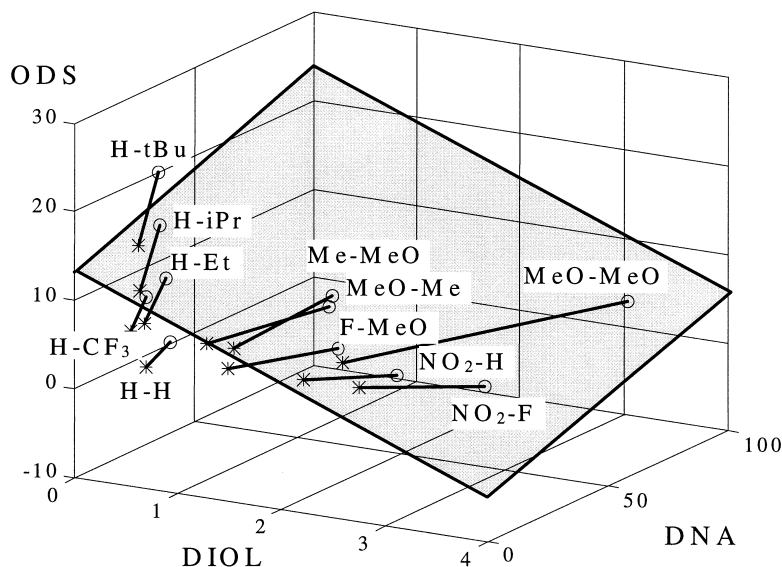


Figure 2. Plot of capacity factors (○ E-s-Cis and * Z-s-Cis isomers) in three-dimensional space defined by the columns DIOL, DNA and ODS which represent the variables.

each variable are interpreted as correlation coefficient (r) with the principal components. The correlation coefficients calculated between the original variables and the principal components are used to interpret the meaning of the principal components. They are essential to interpret the principal components meaning. The representation of the variable is called a correlation circle.

Let us now apply PCA to the sample data matrix corresponding to the chromatographic study of the substituted chalcones.

The 22 solutes, called the objects, were studied with three chromatographic systems, which are the variables. Two graphical approaches are investigated. Chromatographers usually use the first one and the second is PCA.

The most natural way to exploit chromatographic data is to draw a diagram, which represents the variations of retention. The capacity factors are plotted (Figure 2) in the three dimensional space defined by DIOL, DNA, and ODS variables.

In this space, each compound has three coordinates, which are the capacity factors measured for the three chromatographic systems. In this graph, the solutes make a cloud. The main directions of the cloud defined by the

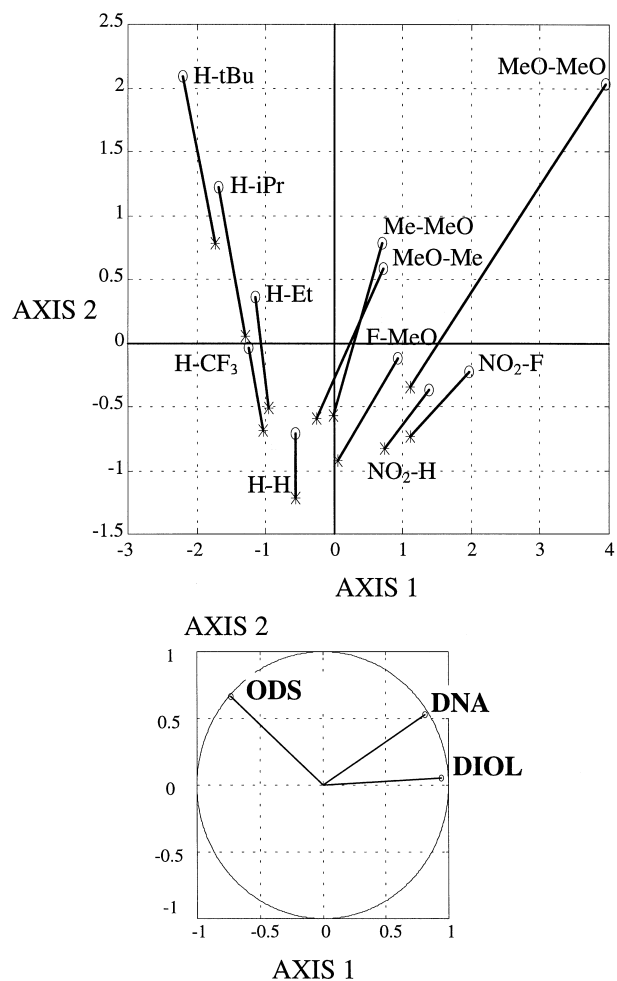


Figure 3. Principal Component Analysis of 66 capacity factors (○ E-s-Cis and * Z-s-Cis isomers) in normal and reversed phases. a) The compounds are projected in the plane defined by the first and the second best axes of inertia, which represent 93% of the information content. b) Circle of Correlation. The independence of the two chromatographic modes appears clearly.

orthogonal axes of inertia, can be represented by the plane drawn in dark in the Figure. A better representation of this Figure will just draw the cloud of points in the previously defined plane. The orthogonal directions of the plane must be calculated. That is the role of PCA.

The PCA of the retention data of the substituted chalcones, considered as the objects, leads to Figure 3. This graph represents the data matrix of capacity factors in a space defined by the first two principal components. The two graphs (Figure 2 and 3) represent the same information, one in three dimensional space and the other in the reduced factorial space calculated by PCA. The projections seem identical.

In Figure 2, the solutes which have the lowest retention with DIOL and DNA systems are in the left of the graph and those with the highest retention on the right. In Figure 3a, the solutes are projected from the left to the right. The PCA graph is similar to the representation of the solutes projected on the dark plane in Figure 2.

Let us now consider PCA of the variables. The correlation circle given in Figure 3b shows that the principal direction defines the three variables projected along the first principal component. On the left side of the graph, the direction is given by ODS and, on the opposite side, by the DIOL and DNA variables. Interpretation of both graphs (3a and 3b) indicates that the solutes which have the highest retention with ODS systems are projected in the direction defined by ODS. Those which have the highest retention with DIOL or DNA systems are projected in the direction defined by the two corresponding variables.

Both usual and PCA graphs give a representation of the same information contained in the data matrix. The chromatographic approach is always limited to a three-dimensional graph taking into account, at the most, only three variables. PCA can be applied to matrices which have higher dimensions.

Correspondence Factor Analysis

Correspondence factor analysis (CFA),^{23,24,35,41,43} follows the same process as PCA, except that the preprocessing of the data matrix is more complex. This preprocessing aims to weight the data in order to change the inertia of the cloud, giving more importance to the small values of the data matrix (Figure 4).

Due to the calculations involved in the preprocessing step, the data table (Figure 4.a) must be a matrix of a non-negative number x_{ij} , in which the row and the column sums are non-zero. A matrix F of relative frequencies (Figure 4.b) is calculated by dividing each element x_{ij} by the sum of the data ($n = \sum_i \sum_j x_{ij}$).

The data matrix of x_{ij} values becomes a data matrix of $\left(f_{ij} = \frac{x_{ij}}{n} \right)$ values. For the study of the objects, each row is transformed (Figure 4.c) into a row profile by dividing each f_{ij} value of the row by its corresponding marginal value

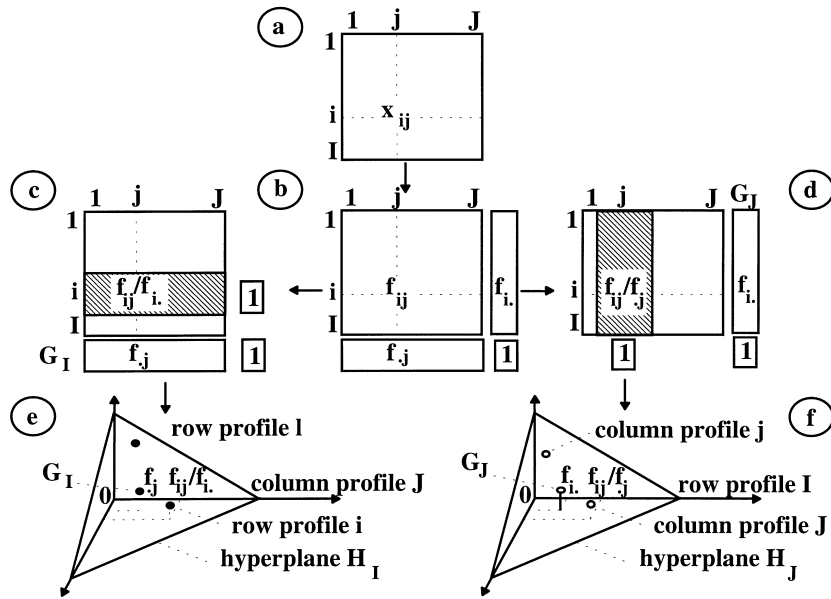


Figure 4. The general processing of the X data matrix with correspondence factor analysis.

$(f_i = \sum_j f_{ij})$. The same transformation (Figure 4d) is applied to the variables.

Each column is transformed into a column profile by dividing each f_{ij} value of the column by its corresponding marginal value $(f_j = \sum_i f_{ij})$. The sums of the

row or column elements are now equal to 1. The initial sums of the row or column represent the weight of the row or column; thus, no information is lost during this transformation. Now, the weak values have a higher weight and the strong values a lower weight than in the initial data matrix.

The elements of the two matrices obtained (Figure 4c and Figure 4d) are represented in their corresponding space. Each object or variable, also called row or column profile, has j or i numerical coordinates.

The objects can be represented (Figure 4e) in the space R^J and the variables in the space R^I . Due to the symmetry between rows and columns in CFA, the construction of the cloud of column profiles (Figure 4f) is performed according to the same methodology.

As explained for PCA, the points constitute a cloud defined by several axes of inertia. The process must search for a series of orthogonal directions for which the inertia of the projection of the cloud on these directions is maximal. The first axis corresponds to the maximum of inertia of the cloud of data points. Other axes of this space are calculated perpendicular to the first and to each other. The factorial axes can reproduce the raw data matrix.

Concerning the spaces R^j and R^l , it must be underlined that the similarity between two rows or two columns is defined by means of the distance between their profiles. This distance is termed chi-square distance (χ^2). For the rows (i.e., i, l row profiles), the following formula is used:

$$d\chi^2(\text{row profile } i, \text{row profile } l) = \sum_j \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_i} - \frac{f_{lj}}{f_l} \right)^2$$

Similarly, for the columns (i.e., j, k column profiles) the distance is calculated as:

$$d\chi^2(\text{column profile } j, \text{column profile } k) = \sum_i \frac{1}{f_i} \left(\frac{f_{ij}}{f_{.j}} - \frac{f_{ik}}{f_{.k}} \right)^2$$

The chi-square distance has the properties of Euclidean distance and confers on R^j and R^l the structure of Euclidean spaces. This distance amounts to allocating to the j th dimensions of R^j the weight $1/f_{.j}$. The sum of the coordinates of each row profile is 1 and therefore, the cloud N_i belongs to the plane H_i .

Each row profile has a weight equal to f_i . The barycentre of the weighted cloud N_i is G_i . Its j th coordinate is the marginal value $f_{.j}$. The barycentre can be interpreted as a mean profile.

Applied to N_i , factor analysis gives a first trivial direction linking the origin O to G_i and orthogonal to H_i . For further direction, G_i is projected at the origin of the axes. The directions are the maximal lengthening directions of N_i . The same demonstration can be made for N_j .

Two factor spaces are obtained, one for the rows and one for the columns. Because of the geometric correspondence of the two clouds of points, the graphical display of the row and column profiles may be merged.

One factor space is obtained, which is simultaneously the factor space for the rows and the columns. CFA can be viewed as finding the best simultaneous representation of two data sets that include the rows and columns of a data matrix.

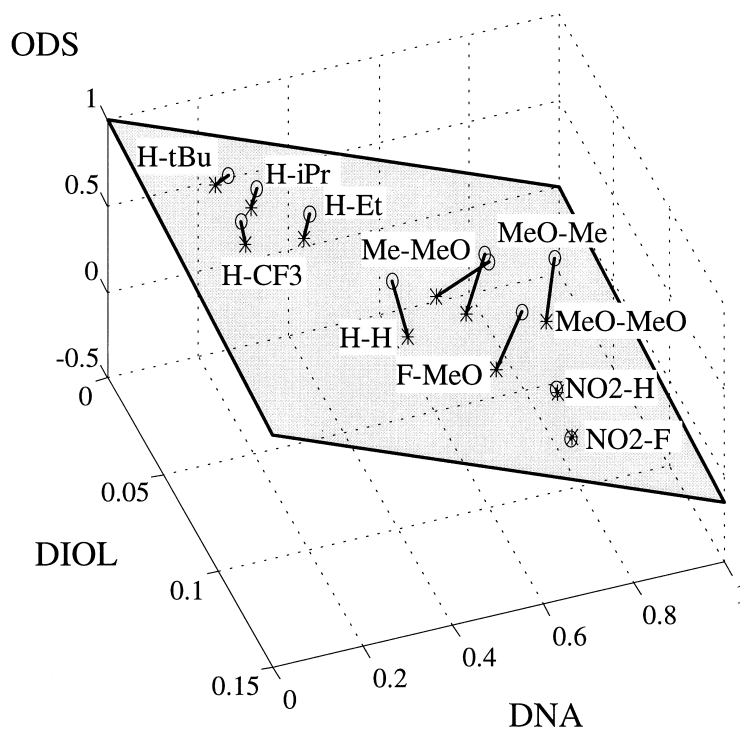


Figure 5. Plot of the row profiles $\left(\frac{f_{ij}}{f_{i.}}\right)$ (O E-s-Cis and * Z-s-Cis isomers) in space defined by the three columns DIOL, DNA and ODS.

To exploit the CFA graphs, the angle θ between the true profile point vector and the principal axis must be examined. The squared cosine of this angle, $\cos^2\theta$, indicates the axis contribution to the inertia of the point. When $\cos^2\theta$ is high, the axis explains the point's inertia very well; in other words when θ is low, the profile vector is said to lie in the direction of the axis, or "correlate" with the axis.

Let us apply CFA to the study of the series of substituted chalcones. To give a comprehensive representation, only the objects are considered. The preprocessing of the data matrix of the capacity factors is applied starting from Figure 4.a until Figure 4.c. Then, the objects are drawn in the R^1 space defined by the columns (Figure 5). The three-dimensional plot is drawn with the three columns DNA and DIOL as x and y axes and ODS as z -axis. On this Figure, the objects are on the plane H_1 represented by the dark rectangle.

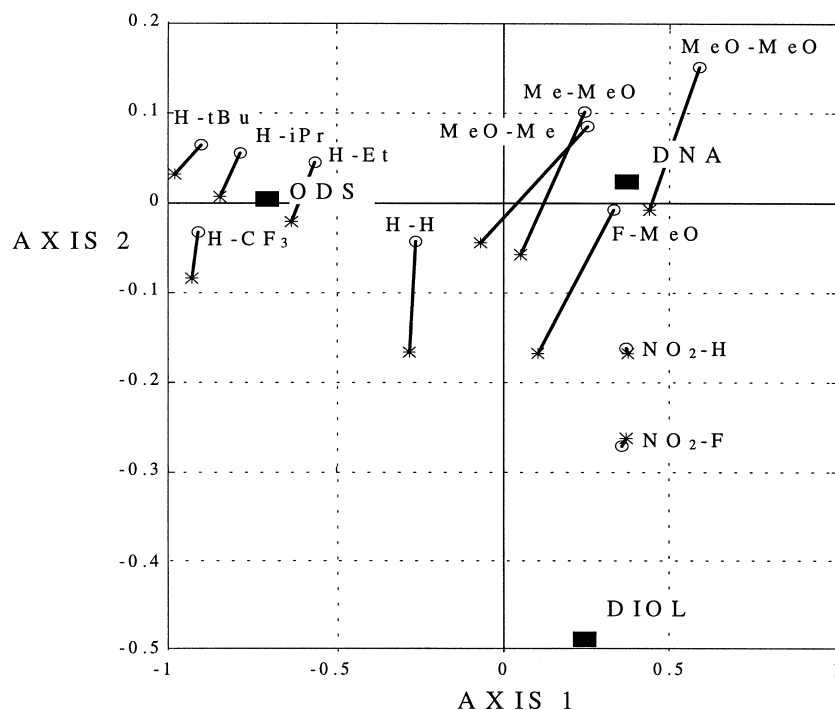


Figure 6. Correspondence Factor Analysis. Simultaneous projection of chromatographic systems (■ ODS, DNA and Diol) and compounds (○ E-s-Cis and * Z-s-Cis isomers) on the plane defined by the first and the second best axes of inertia which respectively represent 95% and 5% of the information content.

The factorial axes, which represent the inertia of the cloud, are defined on the same plane. Now, the completed data process is applied to the capacity factor data matrix. The CFA graph is given in Figure 6. This graph represents the cloud of data points in factorial space. The first two factorial axes are defined in the plane according to the results of Figure 5. The non-polar solutes are projected on the left side of the graphs and the polar solutes on the right side. The CFA graph reproduces the same information in two-dimensional space as Figure 5 in the three dimensional space.

It is interesting to compare the PCA and CFA graphs (Figures 3 and 6). Considering first only the first factorial axis, the two graphs are approximately identical. This means that the information given by the graphs is the same. No information is lost with CFA. The essential differences between the two graphs are the objects which have the highest capacity factors i.e; H-tBu, H-iPr for the ODS variables and MeO-MeO, NO₂-F for DIOL and DNA. With PCA these

objects are not weighted, whereas they are with CFA. Depending on the weight given by CFA, the importance of the highest values is decreased and reciprocally the importance of the lowest values is increased. This is especially evident on the second factorial axes of PCA and CFA. The H-H, H-H* and H-tBu, H-tBu* projections show clearly the effect of the weighting. The distance between H-H and H-H*, objects with low data values, are higher with CFA than with PCA. The distance between H-tBu and H-tBu* which have high data values, are smaller with CFA than with PCA.

The importance of the higher values in the inertia of the cloud is decreased. From these two graphs, chromatographers may have some difficulty determining which is the best graph. In fact, both graphs give specific and interesting information. The PCA graph, which may be easier to understand, gives the main trends defined by the highest values. The CFA graph gives more information to the objects which have small values with respect to the main trends.

The CFA graph increases the discrimination of the small values in cloud by increasing their importance on the inertia of the cloud. This effect on the inertia has often led to the conclusion that CFA is a more discriminating technique compared to PCA. The understanding of the weighting effect is may be more obvious by considering both projections of the objects and variables on the CFA graph.

Axis 1 of CFA is related, as with PCA, to the different systems. On the left side of the graph, the direction is given by ODS and, on the opposite side, by the DIOL and DNA systems. The two variables DIOL and DNA define axis 2 from the bottom to the top. These two directions, which are not really well defined by PCA are now evident. They make the resemblance or the affinity of the objects and the variables explicit. The differences between the solutes which have a high value with the two variables DIOL and DNA are better discriminated with CFA than with PCA (eg. MeO-MeO, NO₂-H, NO₂-F).

The high discrimination given by CFA seems to indicate that CFA is the best factorial analysis. In fact, the weighting effect responsible for the discrimination often gives more difficult graphs. The interpretation of these CFA graphs must be done very carefully.

INTERPRETATION

The abstract factors (axes) extracted from PCA or CFA are not recognizable as physical or chemical parameters, as they are generated to yield a purely mathematical solution. To evaluate ideas concerning the nature of these factors, potential physical or chemical parameters of the chromatographic systems or of the solutes are tested.

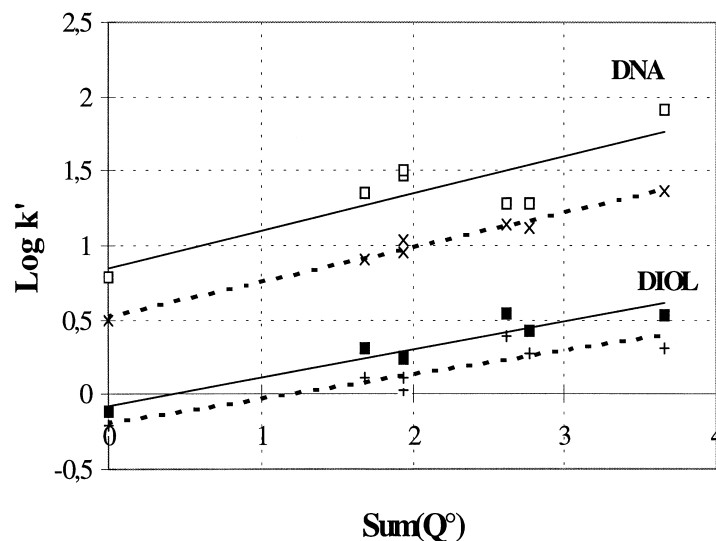


Figure 7. Plot of the logarithm of capacity factors versus sum the adsorption energy (Q°) for the Diol (■ E-s-Cis and + Z-s-Cis isomers) and DNA (□ E-s-Cis and x Z-s-Cis isomers) normal phase systems.

The PCA of the matrix of the series of substituted chalcones eluted from the three normal and reversed phase chromatographic systems is presented in Figure 3. The projection of the 22 compounds is given on the plane defined by the first and the second best principal component. The first two principal components account for 67% and 26%, respectively, of total variance. The PCA of this matrix also gives the correlation circle of the chromatographic systems (Figure 3b).

Two main groups of systems define the two perpendicular directions, which clearly show their independence. These groups correspond to the normal (DIOL, DNA) and reversed phase (ODS) systems.

PCA gives trends about the affinity of compounds for the system belonging to the two chromatographic modes. In fact, the two directions described previously are related to the main parameters which govern the retention. Each group of systems gives an average direction, which presents a better dispersion for sets of compounds. In the center of the two directions the compounds with weak retention in comparison with the system concerned are projected. The more the compounds move off the center, the more they have a better affinity for the system concerned. For example, solutes MeO-MeO, Me-MeO and MeO-Me have a better affinity for NP than for RP systems.

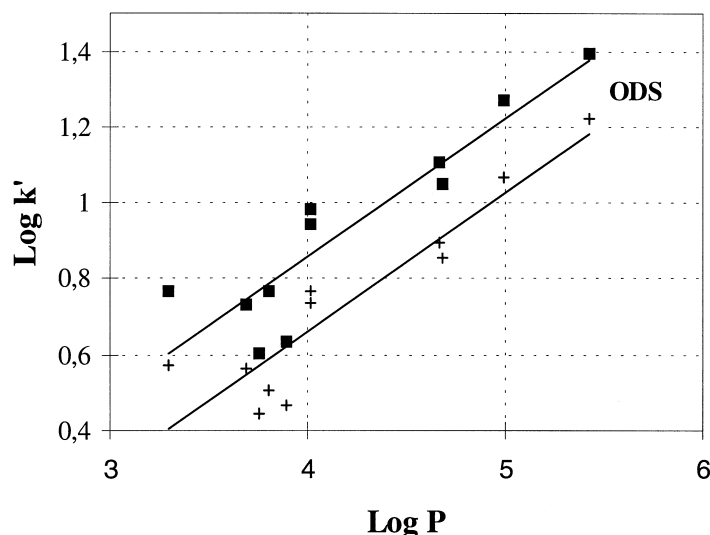


Figure 8. Plot of The logarithm of capacity factors versus the logarithm of the hydrophobic parameter (P) for the ODS (■ E-s-Cis and + Z-s-Cis isomers) reversed phase system.

Similarly, compounds H-tBu, H-iPr and H-Et have a better affinity for RP system. The capacity factors increase depending on the main directions defined by the NP and RP systems. So, trends in affinity are deduced from the proximity of compounds and system projections.

Both directions can be related to physico-chemical parameters. The physico-chemical parameters most commonly used in NP systems is the adsorption energy Q° and the hydrophobic parameter $\text{Log}(P)$ in RP systems. The logarithm of capacity factors is plotted versus sum Q° for the NP system (Figure 7) and versus $\text{Log}(P)$ for the RP system (Figure 8). The good correlation confirms the important part played in the retention mechanism by the adsorption energy Q° in the NP system and by the hydrophobic parameter $\text{Log}(P)$ in the RP system.

With CFA, the simultaneous projection of the chromatographic systems and compounds on planes defined by factorial axes 1 and 2 are given in Figure 6. The two best axes of inertia represent 95% and 5%, respectively, of the information content. The CFA graph shows that the main trend represented on axis 1 is described by the chromatographic modes. Positive values on the abscissa correspond here to the NP mode and negative values to the RP mode. On the NP side one finds packing materials such as DIOL and DNA and, on the RP side, ODS bonded phases. The solutes located on the right side of the graph

have a low affinity for the RP system. These solutes have a better affinity for NP systems (DIOL, DNA). The compounds that have the greatest k' values for RP systems (ODS) are located on the left side of the graph. For example, let us examine three solutes, MeO-MeO, H-tBu and H-H. Solute MeO-MeO has a high k' value in the NP mode and a small k' value in the RP mode. Solute MeO-MeO has a greater affinity for NP systems than for RP systems. Its abscissa is around 0.60. Solute H-tBu has a high k' value in the RP mode and a small k' value in the NP mode. Its abscissa is equal to -0.904.

The affinity of solute H-tBu is higher for RP system. Solute H-H has an abscissa around zero. The affinity of this solute is equivalent in both systems. Consequently, the projection of solutes on axis 1 represents the chromatographic behavior of the solutes with respect to a single chromatographic system or to a set of chromatographic systems.

DNA and DIOL define axis 2 of CFA. The dispersion of solutes can be related to the dispersion of NP systems (DIOL, DNA). For example solute MeO-MeO has a great affinity for DNA. Compound NO₂-F has a great affinity for DIOL.

CONCLUSION

To analyze chromatographic data, it is often necessary to use chemometric methods. Chemometric methods, such as factor analysis, can effectively be used to run a rough preliminary investigation.

The main objectives of factor analysis are, from formal considerations, the determination of linear combinations of variables and the dimensionality reduction of the data. For chromatographers, it can give a useful visualization of multidimensional data and help in the identification of underlying variables, groups of objects, and the determination of relations between variables and objects.

PCA puts the stress on the chromatographic affinity between the solutes and the chromatographic systems. Due to the weighting introduced in the data processing, CFA is more discriminating. The weighting effect underlines both the affinity and the selectivity of the chromatographic systems.

Nevertheless, CFA must be interpreted carefully. With homogeneous data, the weighting effect can modify the projections to give a representation of the selectivity of the chromatographic systems. In these cases, the part of information related to the affinity between the solutes and the chromatographic systems may not be represented on the map.

REFERENCES

1. K. Pearson, *Phil. Mag.*, **2**, 559-572 (1901).
2. C. Spearman, *Amer. J. Psyc.*, **15**, 201-292 (1904).
3. M. A. Sharaf, D. L. Illman, B. R. Kowalski, **Chemometrics in Chemical Analysis**, Vol. **82**, P. J. Elving, J. D. Winefordner, eds., John Wiley and Sons, New York, 1986.
4. E. R. Malinowski, D. G. Howery, **Factor Analysis in Chemistry**, John Wiley and Sons, New York, 1991.
5. M. P. Derde, D. L. Massart, W. Doghe, A. De Waele, *J. Automatic Chemistry.*, **5**, 136-145 (1983).
6. D. L. Massart, L. Kaufman, K. H. Esbensen, *Anal. Chem.*, **54**, 911-917 (1982).
7. C. Baiocchi, E. Marengo, G. Saini, M. A. Roggero, D. Giacosa, *J. Chromatogr.*, **644**, 259-267 (1993).
8. J. Ch. Doré, J. F. Miguel, *C. R. Acad. Sci., Paris*, **293-III**, 1061-1064 (1981).
9. R. Labia, M. Guionie, A. Morand, M. Barthelemy, *J. Ch. Doré, Ann. Microbiol. (Institut Pasteur)*, **134-A**, 29-38 (1983).
10. J. Ch. Doré, J. Gilbert, A. Crontes de Paulet, F. Michel, J. F. Miquel, *C. R. Acad. Sci., Paris*, **294-III**, 731-734 (1982).
11. A. T. Miesch, *J. Int. Assoc. Math. Geol.*, **15**, 501-504 (1983).
12. L. Viallefond, Ph.D., B.R.G.M., Vol. 95-8, pp. 125-135, 1985.
13. J. N. Jaubert, Ph.D., Université Paul Sabatier, Toulouse, 1983.
14. G. Musumarra, S. Wold, Granowitz, *Org. Magn. Resultat.*, **17**, 118-126 (1981).
15. J. C. Davis, **Statistics and Data Analysis in Geology**, John Wiley and Sons, New York, 2nd ed., 1986.
16. P. H. Weiner, J. F. Parcher, *Anal. Chem.*, **45**, 302-307 (1973).
17. D. G. Howery, *Anal. Chem.*, **46**, 829-833 (1974).

18. D. H. McCloskey, S. J. Hawkes, *J. Chromatogr. Sci.*, **13**, 1-5 (1975).
19. M. Chastrette, *J. Chromatogr. Sci.*, **14**, 357-359 (1976).
20. G. Musumarra, G. Scarlata, G. Romano, S. Clementi, S. Wold, J. *Chromatogr. Sci.*, **22**, 538-547 (1984).
21. P. J. Gemperline, *J. Chem. Inf. Comput. Sci.*, **24**, 206-212 (1984).
22. H. Hotelling, *J. Educ. Psych.*, **24**, 417-441 (1933).
23. J. P. Benzécri, *L'Analyse des Données*, 2nd ed., Dunod, Paris, 1980.
24. J. P. Benzécri, *Histoire et Préhistoire de l'Analyse des Données*, Dunod, Paris, 1982.
25. B. Walczak, M. Dreux, J. R. Chrétien, K. Szymoniak, M. Lafosse, L. Morin-Allory, J. P. Doucet, *J. Chromatogr.*, **353**, 109-121 (1986).
26. B. Walczak, J. R. Chrétien, M. Dreux, L. Morin-Allory, M. Lafosse, K. Szymoniak, F. Membrey, *J. Chromatogr.*, **353**, 123-137 (1986).
27. B. Walczak, L. Morin-Allory, J. R. Chrétien, M. Lafosse, M. Dreux, *Chemom. Intell. Lab. Syst.*, **1**, 79-90 (1986).
28. S. Wold, C. Albano, W. J. DunnIII, U. Edlund, K. Esbensen, P. Geladi, S. Hellberg, E. Johansson, W. Lindberg, M. Sjostrom, in **Chemometrics. Mathematics and Statistics in Chemistry**, B. R. Kowalski, ed., Reidel, Dordrecht, 1984.
29. P. J. Lewi, **Multivariate Data Analysis in Industrial Practice**, Research Studies Press, Chichester, 1982.
30. P. P. Mager, **Multidimensional Pharmacology: Design for Safer Drugs**, Academic Press, New York, 1983.
31. S. Wold, K. Esbensen, P. Geladi, *Chemom. Intell. Lab. Syst.*, **2**, 37-52 (1987).
32. B. Escofier, J. Pages, in, **Applied Multivariate Analysis in SAR and Environmental Studies**, J. Devillers, W. Karcher, eds., Kluwer, Academic Publishers, Dordrecht, 1991.
33. J. B. Lebreton, R. Sabatier, G. Banco, A. M. Bancou, in, **Applied Multivariate Analysis in SAR and Environmental Studies**, J. Devillers, W. Karcher, eds., Kluwer, Academic Publishers, Dordrecht, 1991.

34. W. R. Dillon, M. Goldstein, **Multivariate Analysis Methods and Applications**, John Wiley and Sons, New York, 1984.
35. M. J. Greenacre, **Theory and Application of Correspondence Analysis**, Academic Press, London, 1984.
36. T. W. Anderson, **An Introduction to Multivariate Descriptive Statistical Analysis**, John Wiley and Sons, New York, 1984.
37. L. Lebart, A. Morineau, K. M. Warwick, **Multivariate Descriptive Statistical Analysis**, John Wiley and Sons, New York, 1984.
38. B. Escofier, J. Pagès, **Analyses Factorielles Simples et Multiples. Objectifs, Méthodes et Interprétation**, Dunod, Paris, 1988.
39. D. L. Massart, B. G. M. Vandeginste, S. N. Deming, Y. Michotte, L. Kaufman, **Chemometrics: A Textbook**, Elsevier, Amsterdam, 1988.
40. M. Mellinger, **Multivariate Data Analysis: Its Methods**, Chemom. Intell. Lab. Syst, **2**, 29-36 (1987).
41. **Chemometric Methods in Molecular Design**, H. van der Waterbeemd, ed., VCH Publishers, Weinheim, Germany 1995.
42. J. Joliffe, **Principal Component Analysis**, Springer, Berlin, 1986.
43. L. Lebart, A. Morineau, K. M. Warwick, **Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for Large Matrices**, Wiley, New York, 1984.
44. C. Chatfield, A. J. Collins, **Introduction to Multivariate Analysis**, Chapman & Hall, London, 1980.
45. B. S. Everitt, **Cluster Analysis**, Edward Arnold, London, 1993.

Received October 28, 1998

Accepted December 21, 1998

Manuscript 4935

Request Permission or Order Reprints Instantly!

Interested in copying and sharing this article? In most cases, U.S. Copyright Law requires that you get permission from the article's rightsholder before using copyrighted content.

All information and materials found in this article, including but not limited to text, trademarks, patents, logos, graphics and images (the "Materials"), are the copyrighted works and other forms of intellectual property of Marcel Dekker, Inc., or its licensors. All rights not expressly granted are reserved.

Get permission to lawfully reproduce and distribute the Materials or order reprints quickly and painlessly. Simply click on the "Request Permission/Reprints Here" link below and follow the instructions. Visit the [U.S. Copyright Office](#) for information on Fair Use limitations of U.S. copyright law. Please refer to The Association of American Publishers' (AAP) website for guidelines on [Fair Use in the Classroom](#).

The Materials are for your personal use only and cannot be reformatted, reposted, resold or distributed by electronic means or otherwise without permission from Marcel Dekker, Inc. Marcel Dekker, Inc. grants you the limited right to display the Materials only on your personal computer or personal wireless device, and to copy and download single copies of such Materials provided that any copyright, trademark or other notice appearing on such Materials is also retained by, displayed, copied or downloaded as part of the Materials and is not removed or obscured, and provided you do not edit, modify, alter or enhance the Materials. Please refer to our [Website User Agreement](#) for more details.

[Order now!](#)

Reprints of this article can also be ordered at

<http://www.dekker.com/servlet/product/DOI/101081JLC100102045>