ELSEVIER

Journal of Chromatography A, 913 (2001) 23–48

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Identification of the factors that influence the reproducibility of chromatographic retention data

Attila Felinger[a], Marianna Kele[b,c], Georges Guiochon[b,c,*]

[a]*Department of Analytical Chemistry, University of Veszprém, Egyetem utca 10, H-8200 Veszprém, Hungary*
[b]*Department of Chemistry, 552 Buehler Hall, The University of Tennessee, Knoxville, TN 37996-1600, USA*
[c]*Division of Chemical and Analytical Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6120, USA*

Abstract

Principal component analysis was used to identify the parameters that influence the column-to-column and batch-to-batch reproducibility of retention times and retention factors measured on Symmetry C_{18}, Kromasil C_{18}, Luna C_{18} (2) and Vydac RP C_{18}, all reversed-phase silica columns. We devised a procedure that allows the determination of the differences in column volume and packing density between two columns, provided that these columns are packed with identical stationary phases (i.e., phases that originate from the same batch). Principal component analysis of the retention times confirmed that the column-to-column variations of the column volume and the total porosity of the bed are the factors that influence the reproducibility of the retention times, the column volume being the major factor. For the fluctuations of the retention factors, the column phase ratios (or the bed porosities) and some specific, secondary retention mechanisms are responsible. All the C_{18} columns investigated proved to behave in a very similar fashion. Two principal components were always sufficient to characterize the variations of either the retention times or the retention factors. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Reproducibility; Retention data; Principal component analysis; Stationary phases, LC

1. Introduction

Recently, two of us (M.K. and G.G.) undertook a rigorous, systematic investigation of the precision achievable in state-of-the-art reversed-phase high-performance liquid chromatography (HPLC) separations [1–4]. The aim of that investigation was the determination of the short-term and long-term repeatability of chromatographic data acquired with one column as well as the column-to-column and the batch-to-batch reproducibility of the characteristics of chromatographic peaks measured with different brands of commercially available C_{18}-bonded reversed-phase silica columns. The chromatographic parameters investigated were the retention parameters, the hydrophobic interaction selectivity, the steric selectivity, the relative retention of basic–neutral compounds pairs, the column efficiency, and the peak asymmetry.

The measurements made on columns of various brands demonstrated that nowadays the reproducibility of certain porous-silica-based C_{18}-packing materials is quite remarkable. For instance, the relative standard deviation (RSD) of the retention factors is usually of the order of 0.15–1.5% for column-to-column and 1.3–3.0% for batch-to-batch reproducibility. The fluctuations of the experimental data are

*Corresponding author. Department of Chemistry, 552 Buehler Hall, The University of Tennessee, Knoxville, TN 37996-1600, USA. Tel.: +1-865-9740-733; fax: +1-865-9742-667.

E-mail address: guiochon@utk.edu (G. Guiochon).

0021-9673/01/$ – see front matter © 2001 Elsevier Science B.V. All rights reserved.
PII: S0021-9673(00)01044-X

small but systematic and consistent. It would be useful, to spawn further progress, to understand what their origin is. In this study we analyze the retention data measured in our previous studies [1–4]. Our aim is to identify the sources of column-to-column, batch-to-batch, and brand-to-brand fluctuations of the retention times and retention factors for the high quality columns which are now available.

Principal component analysis (PCA) was chosen to determine the ultimate factors that influence the retention data. PCA is used with various objectives in chromatography. It was extensively applied to classify stationary phases and samples and to relate physico-chemical properties of analytes to retention properties. Delaney et al. used PCA for the characterization and classification of HPLC stationary phases [5]. Schmitz et al. used PCA to classify 26 stationary phases and to determine the degree of importance of the test compounds in the Engelhardt test mixture [6]. They concluded that predominantly polar and hydrophobic forces affect the separation process. Seventeen silica-based C_{18} stationary phases were classified by Olsen and Sullivan [7]. PCA was applied by Vervoort et al. to reduce the number of test compounds necessary to characterize a stationary phase [8].

Cruz et al. compared 30 different brands of silica-based C_{18} columns by PCA [9]. They correlated the first principal component (PC) with steric and methylene selectivity and with the retention factor of amylbenzene. The second PC was correlated with the relative retentions of the pairs caffeine–phenol and benzylamine–phenol, the former characterizing the hydrogen bonding capacity, the latter the ion-exchange capacity of the stationary phase [10].

Sándi and Szepesy compared stationary phases with different functionality (C_{18}, C_8, C_4, CN) and pore sizes using one eluent composition and 34 solutes [11]. They found that three PCs describe almost all the differences between the columns. The first PC was assigned to the hydrophobic character of the stationary phases, the other two to the hydrogen-bond acceptor basicity and acidity, respectively.

Brereton and McCalley performed PCA on a data set composed of retention factors, column efficiencies, and asymmetry factors measured on eight commercial reversed-phase columns, using 10 basic compounds and three mobile phase modifiers at pH 3 and 7 [12]. They concluded that out of the 10 compounds initially used, five gave unrelated results so the number of test compounds could have been reduced. The first two PCs explained 65% of the total variance, the first one being twice as significant as the second one. The authors assigned the pH as the major factor contributing to the first PC. The results confirmed an earlier observation of Vervoort and co-workers [8,13] that the position of the columns on the score plot was different at pH 3 and 7.

Welsh et al. tested the applicability of several chemometric classifying methods – including PCA – in pharmaceutical fingerprinting, based on HPLC trace organic impurity analyses [14]. On the score plots obtained, they observed brand-to-brand segregation of the data. For some of the brands, there was no lot-to-lot clustering of the objects, whereas no segregation at all was observed for the short time repeatability.

All these results indicate that PCA is a useful tool in the interpretation of column test results and in the classification of stationary phases. However, one has to choose the initial data matrix with extreme caution when attempts are made to correlate the abstract factors with some actual physical parameters. When PCA is applied to classify stationary phases or to understand the molecular mechanism of retention, a set of very diverse stationary phases is usually chosen and each analyte is injected onto every column. Most of the time no more than one column of any brand is used.

In this study, we do not aspire to classify columns nor to relate retention parameters and molecular properties. We will use PCA to recognize why retention times and retention factors vary from column to column and from batch to batch. We will also discuss the factors which may explain the differences between the behavior of the brands studied. Therefore, our study differs from all previous ones in that we are mostly studying small differences between numerous columns packed with identical stationary phase (when coming from the same batch) or very close ones (when coming from different batches of the same brand).

2. Theory

PCA is a statistical multivariate data analysis

method which allows data reduction, the determination of hidden relationships, data simplification, outlier detection, variable selection, data classification, or data prediction [15,16]. The principal components represent a linear combination of the original variables. The purpose of PCA is to transform the original data set into a smaller set of uncorrelated variables.

The starting point of PCA is a data matrix. The rows of this matrix represent different samples while the matrix columns represent the different sample properties. Most often the original data matrix undergoes some pre-treatment before PCA calculations. Usually the mean value, $\bar{d}_j$, and the standard deviation, σ_j, of each column are calculated, then the corresponding data are centered, by subtracting the mean from each data point in the column, and scaled, by dividing each data point in the column with the standard deviation:

$$d_{i,j} = \frac{d_{i,j} - \bar{d}_j}{\sigma_j} \quad (1)$$

After this data pre-treatment, each property has a zero mean and a unit variance. The original data matrix is then multiplied with its transpose and a covariance matrix is obtained:

$$\mathbf{Z} = \mathbf{D}^{\mathbf{T}}\mathbf{D} \quad (2)$$

The determination of the eigenvalues and the eigenvectors of the covariance matrix can be made with several methods, such as PCA, principal factor analysis, or singular value decomposition. All these methods produce practically the same result. The eigenvectors are mutually orthonormal unit vectors. The eigenvectors and eigenvalues of the matrix satisfy the following set of linear equations:

$$\mathbf{Z}\mathbf{q}_j = \lambda_j \mathbf{q}_j \quad (3)$$

where $\mathbf{q}_j$ is the jth eigenvector and λ_j is the corresponding eigenvalue.

The first eigenvector points into the direction of the largest variability within the original data set. The first eigenvalue gives a measure of that variation. The next eigenvector is orthogonal to the first one and indicates the direction of the remaining largest variation not explained by the first eigenvector, and so forth. Thus, the first factor is the most significant one, and each subsequent factor holds less and less information. The minimum number of eigenvectors necessary to reconstruct the original data matrix within the limits of experimental errors is then determined. A factor matrix is composed from the m necessary eigenvectors:

$$\mathbf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_m] \quad (4)$$

The original data matrix can be reconstructed as the product of two matrices:

$$\mathbf{D} = \mathbf{P}\mathbf{Q}^{\mathbf{T}} \quad (5)$$

where $\mathbf{P}$ is calculated as $\mathbf{P} = \mathbf{D}\mathbf{Q}$. Matrix $\mathbf{P}$ is known as the score matrix whilst matrix $\mathbf{Q}^{\mathbf{T}}$ is often referred to as the loading matrix. Both matrices are abstract matrices, often without any physical significance. Most often some type of transformation is necessary in order to decompose the data matrix into physically interpretable score and loading matrices.

When the number of factors is significantly smaller than the original dimension of the data matrix, a remarkable simplification is achieved via mapping the original data points of usually high dimensionality into an m-dimensional space. It is common in PCA to convert the data matrix into a few informative plots. By plotting the score matrix columns against each other (score plot) one can envision the samples and their configuration in the reduced space. This plot helps to identify outliers, shows groupings and other strong patterns in the data. The plot of the vectors of the loading matrix (loading plot) assists in recognizing which variables cause an object to be an outlier or which variables are responsible for the separation of the different clusters.

3. Experimental

The detailed experimental set-up used for the data collection was published elsewhere [1–4]. Suffice to state here that (1) the column temperature was maintained constant at 25.0°C, measured periodically with an accuracy of 0.1°C, and found to remain constant within these limits; (2) the mobile phase was prepared by pumping the two solvents (methanol and water) using the solvent delivery system of the HP 1100 liquid chromatograph (Hewlett-Packard, Palo Alto, CA, USA) used to collect the data; and (3) the flow-rate stability was found to be better than

0.05% within a few hours and between 0.05 and 0.20% for longer periods of time [1].

The data set included the retention times and retention factors of 17 compounds, nine of them in test 1 and eight in test 2. In test 1, thiourea, phenol, 1-Cl-4-nitrobenzene, toluene, ethylbenzene, butylbenzene, *o*-terphenyl, amylbenzene and triphenylene were used with methanol–water (80:20, v/v) as the mobile phase (except for the Luna-C_{18} columns, for which amylbenzene was replaced by propylbenzene to avoid some interference). In test 2, thiourea, aniline, phenol, toluidines, ethyl-benzoate, *N*,*N*-dimethylaniline, toluene and ethylbenzene were used with methanol–water (55:45, v/v) as the mobile phase. The columns were equilibrated with the required mobile phase for 5 h before the first injection of the test samples. Five successive injections of each sample were made on each column.

Fifty-one columns of four different brands [Symmetry C_{18} (Waters, Milford, MA, USA), Kromasil-C_{18} (Eka Chemicals, Bohus, Sweden), Luna-C_{18} (2) (Phenomenex, Torrance, CA, USA) and Vydac-C_{18} 218TP54 (Vydac, Hesperia, CA, USA)] were used in the study. For each brand, five columns were packed with packing material belonging to the same batch and 10, six, nine or six columns for the respective four brands listed above were packed with material from as many different batches. The columns were packed by the manufacturers and used as received.

4. Results and discussion

4.1. Error analysis of the data

The fluctuations of the retention times measured for a given compound on columns packed with nearly identical aliquot of the same batch of a stationary phase are affected by the fluctuations of the column size (or more exactly of the size of the tube in which the column is packed), the packing density, the flow-rate, the eluent composition and the column temperature.

In our experiments, the flow-rate, the eluent composition, and the temperature were kept constant (see Experimental section). The stability of these parameters was sufficient to eliminate their contribution to the experimental errors, as confirmed later. Therefore, the fluctuations of the retention time observed between the different columns packed with the same lot of packing material can be attributed to the column-to-column variations of the packing density and of the column volume. We can express the retention time as:

$$t_R = \frac{V_0}{F_v} \cdot (1 + k') = \frac{V_c \epsilon}{F_v} \cdot (1 + k') \quad (6)$$

where V_0 and V_c are the void volume and the total volume of the column, respectively; F_v is the mobile phase flow-rate, k' is the retention factor, and ϵ is the total porosity of the column, defined as the fraction of the column volume occupied by the mobile phase. Using the conventional relationship between the retention factor, the porosity, and the thermodynamic constant of the phase equilibrium:

$$k' = \frac{1 - \epsilon}{\epsilon} \cdot K \quad (7)$$

we can re-express the retention time as a function of two constant terms (the flow-rate and the thermodynamic constant) and two variable terms (the total column volume and the total porosity):

$$t_R = \frac{V_c}{F_v} \cdot [\epsilon + (1 - \epsilon)K] \quad (8)$$

Calculating the full differential of the retention time and rearranging gives:

$$\frac{dt_R}{t_R} = \frac{dV_c}{V_c} + \frac{1 - K}{\epsilon + (1 - \epsilon)K} d\epsilon \quad (9)$$

For an unretained compound, the above equation simplifies to:

$$\frac{dt_R}{t_R} = \frac{dV_c}{V_c} + \frac{d\epsilon}{\epsilon} \quad (10)$$

These equations allow the separate determination of the contributions of the fluctuations of the column volume and the total column porosity to the fluctuations of the retention times and factors. To compare two columns, we calculate for each compound of a test mixture the relative difference of their retention times on the two columns and plot dt_R/t_R against $(1-K)/[\epsilon+(1-\epsilon)K]$ (Fig. 1). For each compound, this latter term is calculated as an average for the two columns, using the average total porosity of the columns. The use of these average values in the

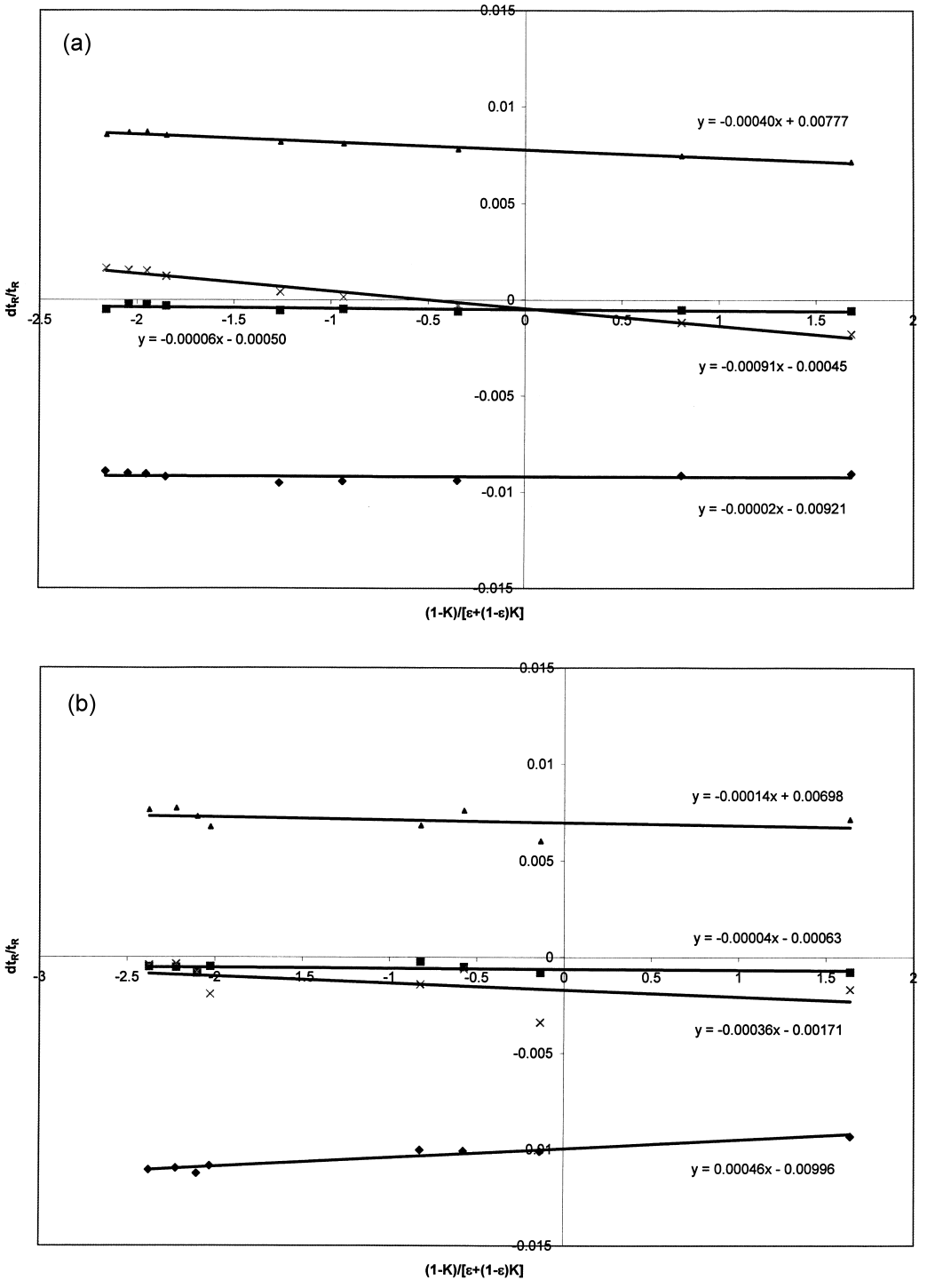


Fig. 1. Plot of dt_R/t_R against $(1-K)/[\epsilon+(1-\epsilon)K]$ for the five Symmetry C_{18} columns packed with packing material coming from the same batch for (a) test 1 and (b) test 2.

calculations and the fact that the equilibrium constant is derived for each analyte by using Eq. (7) introduces some uncertainty in the abscissa of the corresponding data point. This uncertainty, however, is rather small and the information we derive is hardly affected.

The intercept of the straight lines fitted to the plot corresponding to two columns gives the relative column volume difference between these two columns. Their slope provides the packing density difference. Any arbitrarily selected column can be used as a reference in this calculation. In the graph (Fig. 1), this column is then represented by the *x*-axis. The consistency of the results can be checked by carrying out the same determinations with the same set of columns but using the data obtained with different test mixtures (and mobile phase composition). Fig. 1 shows the results obtained with the five Symmetry C_{18} columns packed with material from the same batch. Fig. 1a and b were derived from the data acquired with tests 1 and 2, respectively. These figures show that the compositions of the test mixture and of the eluent have a negligible effect on the slope and intercept of the straight lines. The differences observed are small and due to experimental errors. The column-to-column fluctuations observed are not due to experimental errors but arise from systematic column-to-column variations of the column volume and its total porosity.

Similar calculations were repeated on the five Kromasil C_{18}, Luna C_{18} (2) and Vydac C_{18} columns. The results obtained for the four brands are summarized in Fig. 2. This figure shows plots of the column-to-column differences of the total porosity and total column volume. The columns are identified by their (arbitrary) rank (1 to 5) in the series. Out of five Symmetry C_{18} columns, three have almost the same size. The difference between the volumes of the smallest and the largest column is 1.7%. The relative difference between the total porosities of these five columns is an order of magnitude smaller than the relative difference of the column size, the largest such difference being 0.15%. The largest column volume difference among the five Kromasil

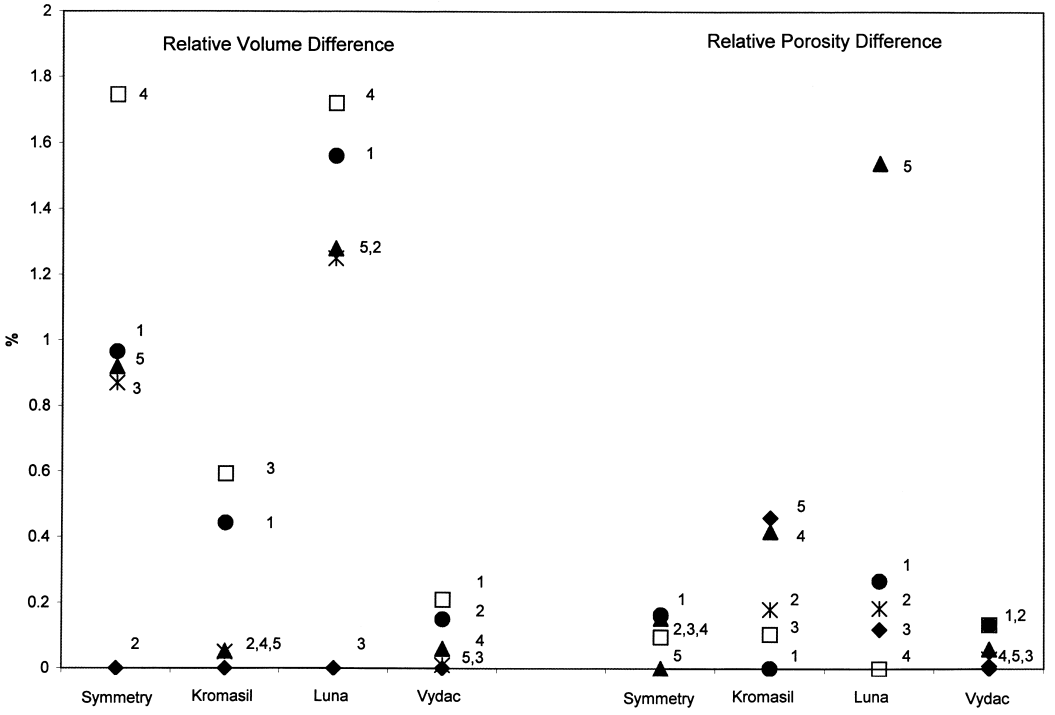


Fig. 2. Relative column volume and total porosity differences on five Symmetry, Kromasil, Luna and Vydac columns packed with packing material coming from the same batch.

columns was 0.58%, the largest total porosity difference was 0.46%. The highest volume difference among the five Luna columns was 1.7%, one column being much smaller than the other four. These columns show also a high variation of their total porosity, the largest difference being 1.54%, mainly because the total porosity of one column is much higher than the total porosity of the other four columns. Excluding this value from the comparison, the largest total porosity difference drops to 0.26%. Among the Vydac columns both the column volume and the total porosity differences are small (0.21 and 0.14%, respectively, as the largest difference).

In order to see whether the calculated column volume and porosity differences are significant, the standard deviations of the slope and intercept parameters were calculated for each fitted line (such as those in Fig. 1a and b) and *t*-tests were performed to assess the significance of the difference between parameters [17]. The RSD of the intercepts is roughly 7–10%, while that of the slopes fluctuates more, in the 3–25% range. For the Symmetry columns, for instance, we found that the volume of column 2 is significantly smaller and that of column 4 is significantly larger than the volume of columns 1, 3 and 5 while the statistical analysis failed to distinguish between the volumes of columns 1, 3 and 5. When the total porosity of the Symmetry columns are considered, we can state that the porosity of column 5 is significantly smaller than that of the rest, while we cannot see a significant difference between the other four columns. In general, we can conclude that there is no significant difference between those columns that are grouped together in Fig. 2, and we can confirm that columns situated farther away from each other or from groups of columns are significantly different regarding either column volume or total porosity.

The range of fluctuations observed for the column volumes are smaller than the value which can be derived from the tubing specifications. These are $\pm$0.003 in. ($\pm$0.0076 cm) for the inner diameter and $\pm$0.01 in. ($\pm$0.0254 cm) for the column length for Vydac and Kromasil columns, $\pm$0.002 in. for the inner diameter and $\pm$0.005 in. for the length for Symmetry columns and $\pm$0.0015 in. for the inner diameter and $\pm$0.01 in. for the length for Luna columns. From these values we can derive a volume fluctuation of the tube volume of 3.5% for Vydac and Kromasil, 2.7% for Symmetry and 1.9% for Luna columns. There are several possible explanations of the observed smaller fluctuations. The specifications given might not have been updated. They correspond to size fluctuations over a large batch (so-called ''heat'') of stainless steel tube and the manufacturing of column tubing does not lead to complete randomization over such a large batch. Column manufacturers do not seem to have paid much attention to this factor, yet.

For any brand, the packing material contained in any set of columns packed with material from the same batch should have identical physico-chemical properties since a batch is intensely mixed during the bonding process. Then, only the column packing procedure should be responsible for the variations of the phase ratio. From Eq. (9), we concluded that the relative range of variations of the total porosity of the five columns packed with the same batch of packing material was 0.15% for Symmetry, 0.46% for Kromasil, 1.54% for Luna and 0.14% for Vydac. These numbers can be related to the range of fluctuations of the retention factors on these columns. This range can be calculated by differentiating Eq. (7):

$$\frac{dk'}{k'} = \frac{-1}{1-\epsilon} \cdot \frac{d\epsilon}{\epsilon} \tag{11}$$

On the basis of test 1, the largest relative differences of the retention factors are 0.37% for Symmetry, 1.16% for Kromasil, 3.60% for Luna and 0.40% for Vydac columns [2–4]. Using these numbers, we can calculate the relative differences in the total porosity of the columns, with Eq. (11). The results are 0.15% for Symmetry, 0.47% for Kromasil, 1.28% for Luna and 0.15% for Vydac columns. These values are in excellent agreement with those calculated by fitting Eq. (9) for Symmetry, Kromasil and Vydac columns and give a fair estimate for the Luna columns.

4.2. Principal component analysis of the retention times on a single batch

PCA performed on the same data set (retention times of test compounds in test 1 and test 2) gave

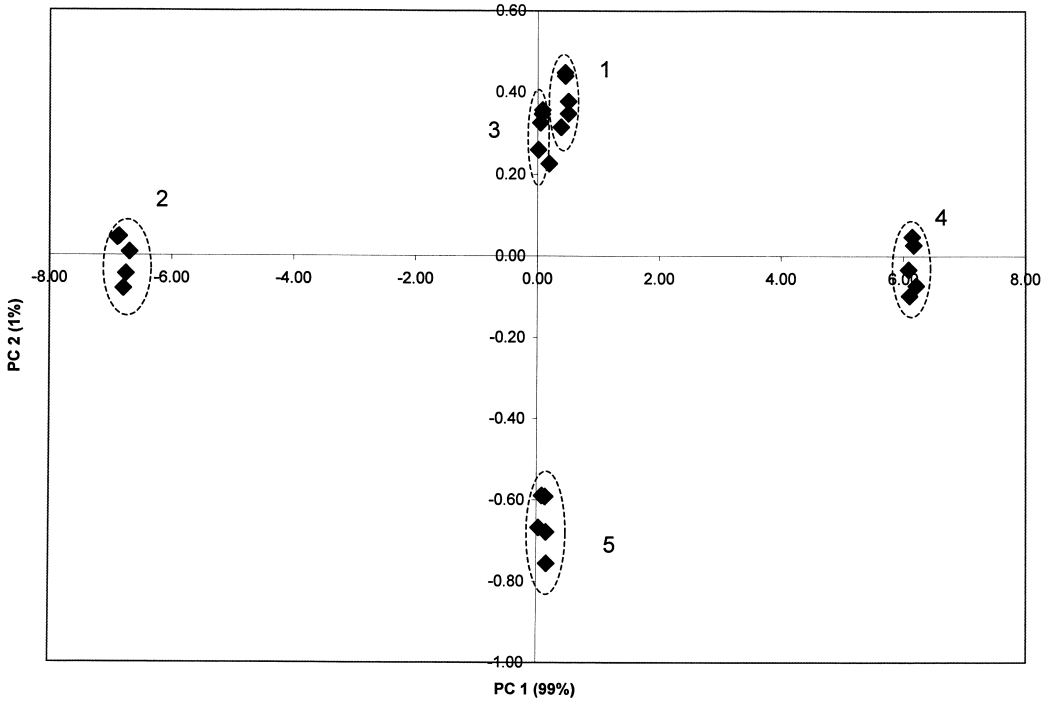


Fig. 3. Score plot of PCA performed on retention times obtained by test 1 on five Symmetry C_{18} columns packed with packing material coming from the same batch.

results consistent with those of the above calculations. The score plot for the five Symmetry columns is shown in Fig. 3. Since five replicate measurements were made on each column, each column is represented by five symbols. Two factors are necessary to describe the retention time differences observed on the five Symmetry columns. The first factor explains 99% of these differences, the second one, the remaining 1%. Comparing the position of the five columns on the score plot along the abscissa (first PC) and the ordinate (second PC) with the relative volume and total porosity differences (Fig. 2) allows a meaningful physical assignment of the two PCs. The positions of the five Symmetry columns along the abscissa in Fig. 3 are in excellent agreement with the order of their total column volumes as plotted in Fig. 2. The volume of the second column is smaller and that of the fourth column is larger than those of the other three columns which are almost the same. This indicates that the first factor characterizes the column volume differences.

Although the second factor has a small effect and, in most cases, a factor with a weight of only 1% could be ignored, we see that the position of the columns along the ordinate in Fig. 3 is consistent with their rank when listed by order of increasing total porosity (Fig. 2). For the second factor as for the porosity, the largest difference is between columns 1 and 5. This suggests that the second factor correlates with the total porosity differences. If this assumption is correct, we can predict the results of the PCA of the retention times obtained on the five Kromasil columns. We should obtain similar first factor values for columns 1 and 3 on the one hand and for columns 2, 4 and 5 on the other hand. This is indeed what is observed in Fig. 4 which shows the score plot of the PCA performed on the retention times measured on the five Kromasil columns. The position of the columns along the abscissa correlates with the column volume rankings and so does their ordinate with the total porosity rankings, the only exception being the third column in this last case. From Fig. 2, we conclude that the total porosities of the Kromasil columns 2 and 3 are close. The score

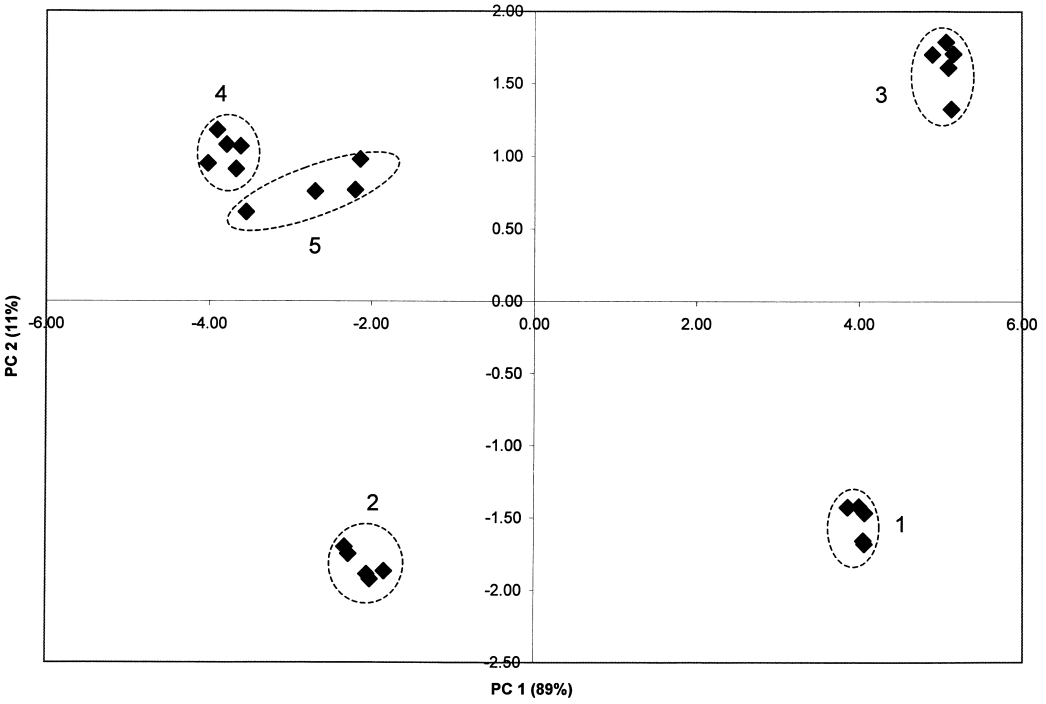


Fig. 4. Score plot of PCA performed on retention times obtained by test 1 on five Kromasil C_{18} columns packed with packing material coming from the same batch.

plot of PCA, on the other hand (Fig. 4) indicates that the total porosity of columns 1 and 2 are nearly identical and markedly different from those of columns 3, 4 and 5, with these last three columns having very close total porosities.

The PCA of the retention times on the Luna C_{18} (2) columns gives results which are roughly consistent with those derived from the ordering of columns by their calculated volume and packing density. The score plot (Fig. 5) indicates that the third column is the smallest and the fourth one the largest, a result in agreement with our previous calculations (Fig. 2). The ranking based on the total porosity matches with our earlier result for the columns with the smallest and the largest total porosity but there are some minor discrepancies for the columns of intermediate total porosities. If the two axes are slightly rotated counter-clockwise, into the position of the arrows in Fig. 5, the column volume and the porosity order of the columns is restored.

Factor axes are often rotated in PCA, in order to ease the physical interpretation of some of the factors. The most popular transformation method is the Varimax rotation [18]. This and other algorithms, however, try to find such a combination of factor axes that results in a simple structure; they attempt to locate the objects as close to the final factor axes as possible. The rationale for such an arrangement is that one usually wishes to maximize the effect of one factor on any object. Nevertheless, such an algorithm does not necessarily increase the chemical and/or physical meaning of a factor. For this reason, and because in this study the abstract factors calculated could always be related to physical properties calculated with independent methods, we omitted the rotation of the factor axes.

The weight of the second factor is 1% on Symmetry, 12% on Kromasil, 25% on Luna and 6% on Vydac columns. This is related to our independent finding that the relative total porosity difference on Symmetry columns is an order of magnitude smaller than the volume differences. Accordingly, the factor, representing the variations of the column volume is

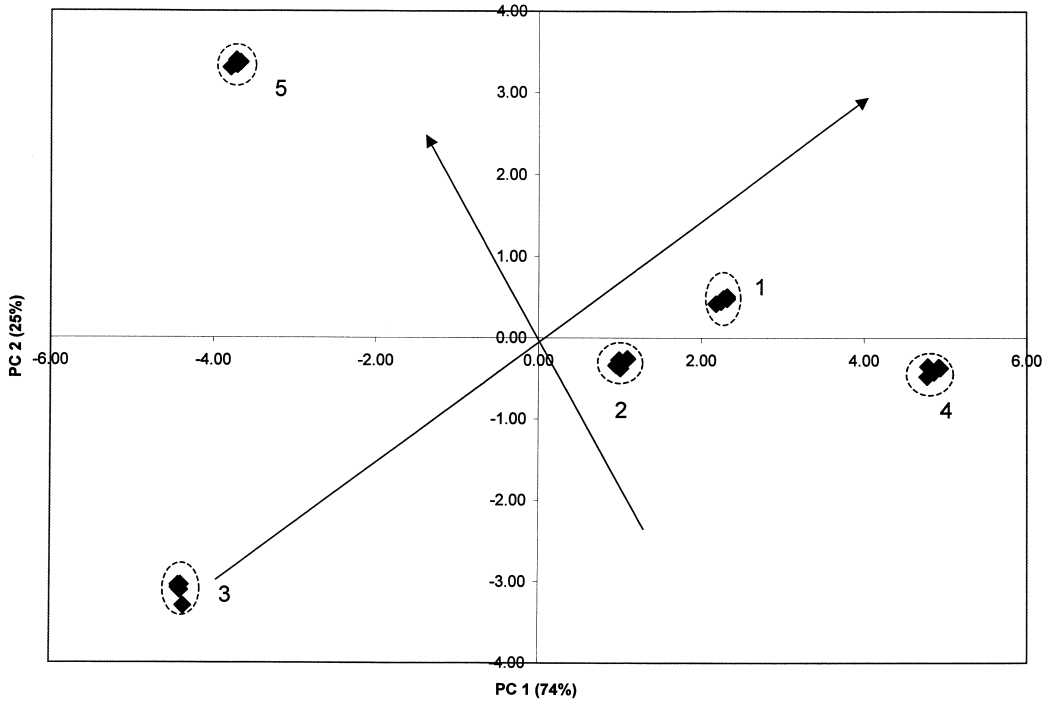


Fig. 5. Score plot of PCA performed on retention times obtained by test 1 on five Luna C_{18} (2) columns packed with packing material coming from the same batch.

sufficient to explain almost all the retention time variations observed. On the Kromasil, Luna and Vydac columns, the range of total porosity variations is only slightly smaller than that of the column volume. These results confirm that most probably the two factors are closely related to the column volume and the total bed porosity, respectively.

The score plots (Figs. 3–5) allow also a visual evaluation of the repeatability and the reproducibility of the measurements. It is obvious from these figures that the short-term repeatability of the measurements is far better than the column-to-column reproducibility of the batch. The distance spanned by the five symbols corresponding to one column is proportional to the short-term repeatability whereas the distance between the centers of the respective five symbols is proportional to the column-to-column reproducibility. It is unusual that the clusters of data points for two columns are close to each other in the two-dimensional factor space. This happens only for Symmetry columns 1 and 3 and for Kromasil columns 4 and 5. The separate score plots, however, do not allow the direct comparison of the short-term repeatability achieved on columns from different brands, because the data matrices were subjected to centering and scaling which affects the proportionality coefficient between cluster size and retention time RSD. For the purpose of direct comparison, a large data matrix must be created with all the retention times of all the brands and the PCA performed on that matrix (see later).

4.3. Principal component analysis of the retention times on all batches

PCA was also performed on the data matrix composed of the retention times of the compounds of the first test mixture on all the columns from the same brand. The score plots obtained for the different brands are plotted in Figs. 6–8. The plots show that two PCs always describe completely the retention time variations among the different batches. In Figs. 6–8, the numbered columns are the same as in Figs. 3–5. The introduction in the set of the data

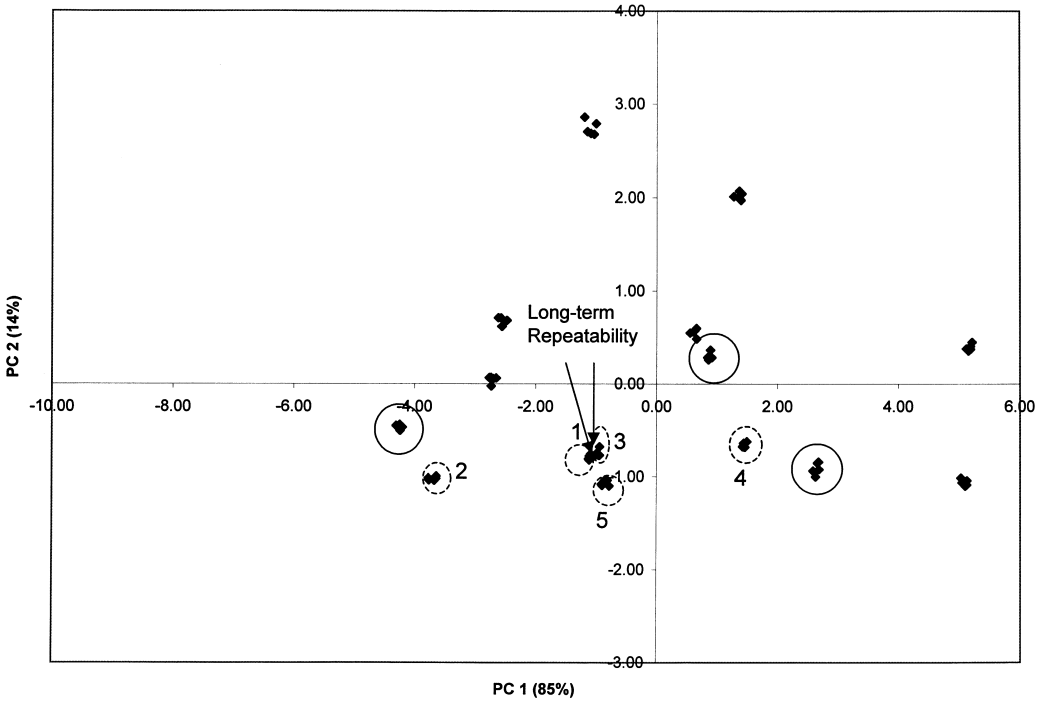


Fig. 6. Score plot of PCA performed on retention times obtained by test 1 on five Symmetry C_{18} columns packed with packing material coming from the same batch and on 10 columns from 10 different batches.

obtained with columns from different batches did not change the relative position of the five columns packed with material of the same batch on the score plots (compare Figs. 3 and 6; Figs. 4 and 7; Figs. 5 and 8). This indicates that the two PCs correspond to the same properties as before; the first factor can be correlated with the column volume, the second one with the total porosity.

Since the columns studied are now packed with material coming from different batches, the actual meaning of the ''total porosity'' is less clear. Its fluctuations may originate not only from variations in the packing density but also from batch-to-batch variations of the pore size, the carbon content, and the surface coverage of the stationary phase. Evidently, all these parameters influence the phase ratio of the columns to some extent. We will refer to this effect in a generic manner. The retention of nonpolar, nonionic compounds is governed only by dispersive interactions. The nonselective retentivity (hydrophobicity) expresses the retention of these analytes on the column. It is less influenced by batch-to-batch chemistry differences than the retention of polar or basic compounds. Since the relative position of the columns originally plotted in Figs. 3–5 remains unaffected by the introduction of the data pertaining to different batches in the data matrix, we conclude that column hydrophobicity and total bed porosity are strongly correlated in the set studied here, in large part because different batches of a given brand of packing material are still C_{18}-bonded silicas with very similar stationary phase properties.

The dashed-line circles in Figs. 6–8 correspond to the five columns packed with material of the same batch. The solid-line circles surrounding some other object clusters in Fig. 6 indicate the three different batches of stationary phases that were prepared using the same lot of neat silica. From the score plot, it is obvious that these three batches do not cluster in the factor space. Accordingly, the silica itself has no or little identifiable effect on the variations of the retention times. One should recall, however, that the first factor is the column volume – a parameter

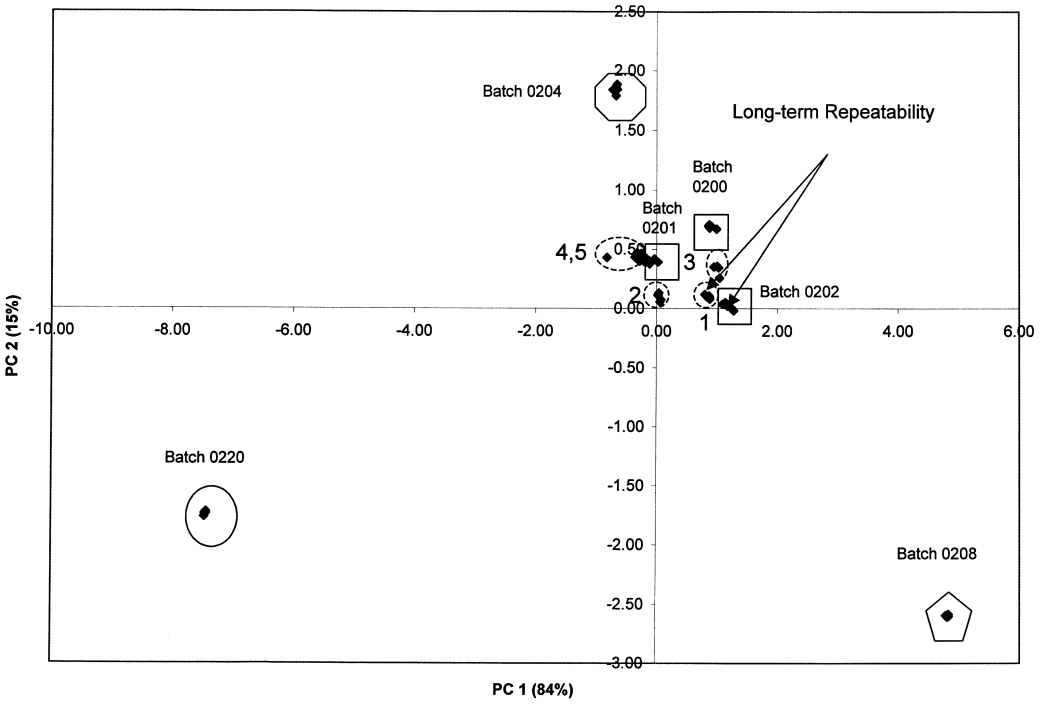


Fig. 7. Score plot of PCA performed on retention times obtained by test 1 on five Kromasil C_{18} columns packed with packing material coming from the same batch and on six columns from six different batches.

which is completely unrelated to the nature of the underlying silica. The factor along the ordinate is the bed porosity, which is closely related to the hydrophobicity of the columns. It is not surprising then that the packing materials made from the same silica lot are very close to each other in the direction of this factor.

For the Kromasil columns, the three batches of stationary phase that are based on the same lot of silica are marked with squares in Fig. 7. They show very similar properties. The variations of both the column volume and the hydrophobicity are much smaller for them than the total variations observed for the whole data set.

In Fig. 8, the different batches of Luna C_{18} based on the same lot of silica are marked with the same symbol (squares, circles or rhombuses). The results indicate that the underlying silica has no or little influence on the variations of the retention times. In this instance, even the values of the second factor are very different for these batches of material made from the same lot of silica.

The plots in Figs. 6–8 indicate also that the contribution of the repeatability or error of measurement on the phenomenon studied is negligible. The values derived from repeated experiments (indicated as long-term repeatability in Figs. 6–8) are much closer than the values obtained with the data measured on the sets of five columns or on those of all the columns from the different batches (see later). By contrast with the results obtained with the PCA of the data measured on the sets of five columns, the two PCs now have almost the same weight on the three different brands. The weight of the second factor increased on the Symmetry columns, remained the same on the Kromasil columns and decreased on the Luna columns compared to the results of the PCA of the five-column data sets. The plots indicate that the scatter of the data points along the abscissa for the ten Symmetry columns packed with material coming from different batches is twice as large than that for the five columns packed with material from the same batch. The corresponding increase of the scattering of the data points is sixfold for the Kromasil columns and threefold for the Luna columns.

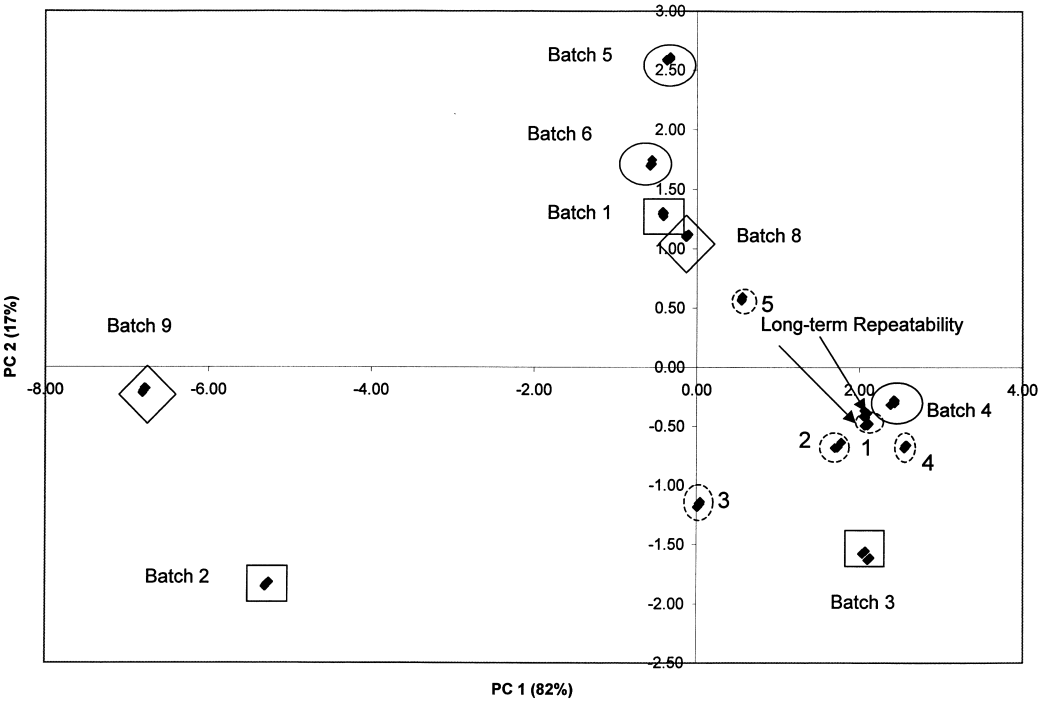


Fig. 8. Score plot of PCA performed on retention times obtained by test 1 on five Luna C_{18} (2) columns packed with packing material coming from the same batch and on nine columns from nine different batches.

However, this increase cannot be attributed unambiguously to increased column volume differences because different batches of silica and of chemical modifications can lead to materials having slightly different physico-chemical properties.

The fact, that the significance of nonselective retentivity (second PC, scattering along the ordinate) is much higher on the different batches than on the columns packed with packing material from the same batch is not surprising. The surface area and pore volume differences between the batches and the fact that the different batches have different particle size distribution predict this result. Nevertheless, Figs. 6–8 allow a simple visual evaluation of short-term repeatability, as well as column-to-column and batch-to-batch reproducibility of retention time.

4.4. Principal component analysis of the retention factors

The retention factors should be independent of the column volume since these are relative data, normalized by the hold-up time (Eqs. (9) and (10)). Thus, the retention factors were also subjected to PCA. As expected, only one factor affects the fluctuations of the retention factors measured for the components of the first test mixture (neutral compounds) on the set of five columns packed with material from the same batch. This factor must be related to the phase ratio differences among the five columns. The score plots on the Symmetry, Kromasil, and Luna columns are shown in Figs. 9–11.

The PCA of the retention times showed that the column volume is the dominant factor. With its elimination, we have to deal with very small differences. In the case of Symmetry C_{18}, the first factor found in the PCA of the retention times was the column volume and it explained 99% of the variations. The variations of the retention factor are much smaller than those of the retention times and the first factor explains at least 99% of these variations. Thus, we must consider as irrelevant the ordinate in Figs. 9–11. The significant information in these figures is the position of the data points along the abscissa. The relative position of the columns along this axis in Figs. 9–11 is essentially the same as their position

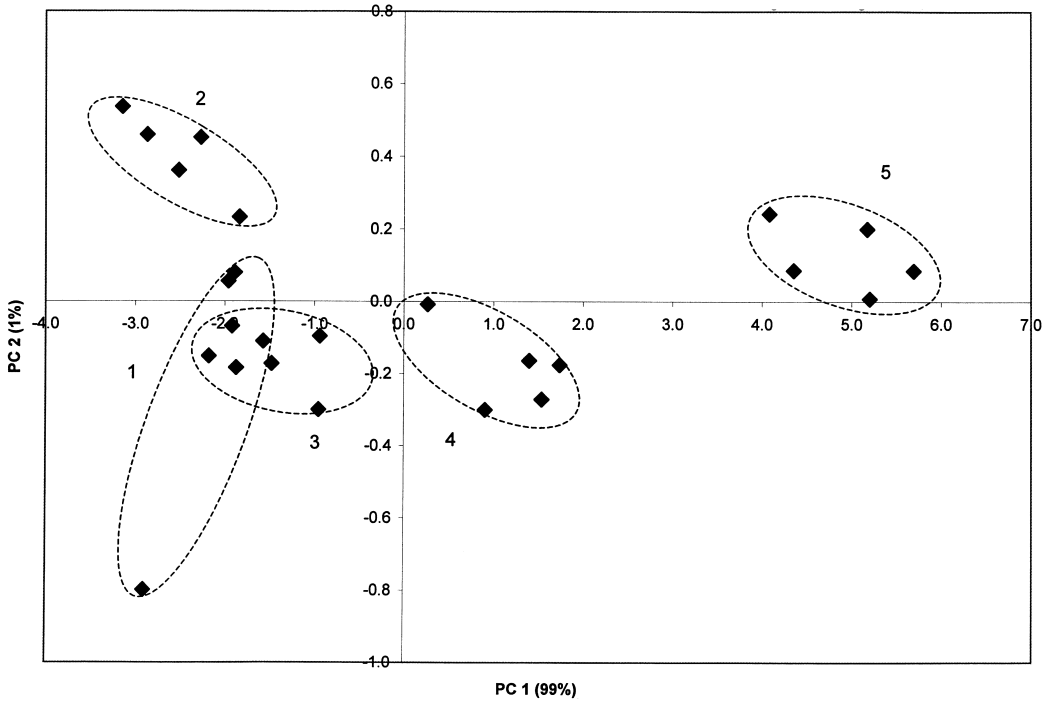


Fig. 9. Score plot of PCA performed on retention factors obtained by test 1 on five Symmetry C_{18} columns packed with packing material coming from the same batch.

along the ordinate in Figs. 3–5 and it matches the order of total bed porosity of the corresponding columns as plotted in Fig. 2. Accordingly, we can conclude that the total porosity, which is the second factor in the PCA of the retention times, becomes the first and only factor in the PCA of the retention factors.

Then, the PCA of the retention factors of test mixture 2, that contains both neutral and basic compounds, was carried out. For the Kromasil and Luna sets of five columns, the weight of the first factor was the same as with test 1, at least 99%. For the five Symmetry columns, however, an unexpected result was obtained. Two PCs were found, with weights of 75% for the first factor and 23% for the second (Fig. 12a). This result should be related to the fact that the three basic compounds in this test mixture (aniline, toluidines, *N*,*N*-dimethylaniline) gave a second PC whose value is high compared to the values found for neutral compounds, for which this second PC is practically zero (Fig. 12b). When the PCA was repeated on the data set containing only the retention factors of the neutral compounds of this test mixture, the relative weight of the second PC decreased to 2%. We showed earlier (Fig. 2), that the porosity difference among the five Symmetry columns is very small (the largest difference was found to be 0.2%). Thus, the first PC describes this very small variation while another (still smaller) effect, related to specific, secondary interactions between the stationary phase and the basic compounds, creates the second PC. This can be seen on the loading plot that indicates that only the basic compounds have a significant second factor (Fig. 12b).

PCA was carried out on the data matrix of the retention factors measured on the eleven columns packed with Kromasil C_{18}. The score plot and the loading plot for the components of Test 1 are presented in Fig. 13, those for the components of test 2 in Fig. 14. Again, two factors completely characterize the variations of the retention factors among the different columns. The relative weights of the second factor are 2% and 5% for the compounds in tests 1 and 2, respectively. Both plots indicate that

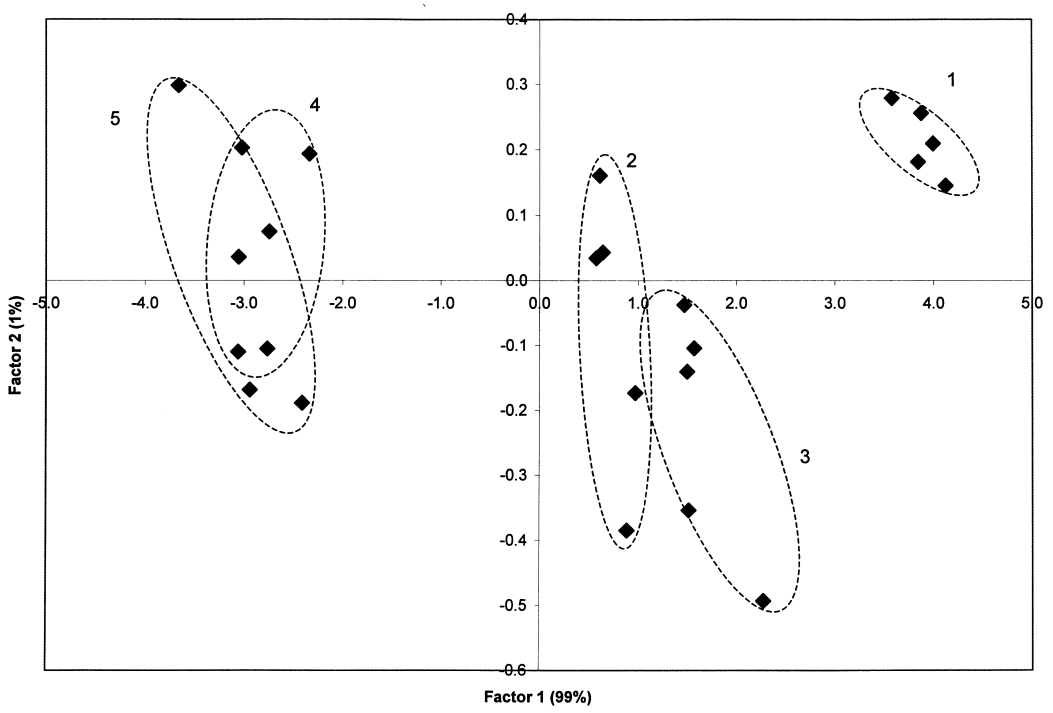


Fig. 10. Score plot of PCA performed on retention factors obtained by test 1 on five Kromasil C_{18} columns packed with packing material coming from the same batch.

the five columns packed with the same batch of packing material have nearly the same value of the second factor, the differences between them being described by the first PC. The loading plots indicate that the highest value of the second PC is obtained for triphenylene in test 1 and for two basic compounds, aniline and *N*,*N*-dimethylaniline in test 2. Note that, because of the partial separation of the toluidine isomers on Kromasil columns, the retention factors of these compounds were excluded from the data set for this phase. This demonstrates again that the second PC is related to secondary, specific retention mechanisms. This does not imply that the second PC has the same physical meaning in the two tests. The points representing the three batches based on the same batch of silica (marked with squares) have different positions along the ordinate (second PC) for tests 1 and 2. This shows that the second factor is different in the two tests. This is not surprising since it is likely that the retention of triphenylene is affected by steric effects while the retention of the basic compounds is affected by weak secondary interactions between the amino groups and the surface silanols. On the whole, however, some similarity can be observed in Figs. 13 and 14 regarding the relative position of the particular batches along the abscissa.

The position of the different batches on the score plots indicates that the first factor can again be related to the phase ratio or the hydrophobicity of the columns. This result is unexpected. One would expect to observe variations, even if minor, of the equilibrium constants from batch to batch, due to difference in the surface modification process. Changes of the alkyl chain density on the surface should cause changes of the equilibrium constant. The abscissa of the points representing the different batches on the score plots of the retention factors and the retention times should be different. This interpretation is supported by the fact that the second factor is mostly due to triphenylene (Fig. 13b) whose retention is greatly influenced by steric effects.

Therefore, this test shows that when three different batches of packing material are prepared by the

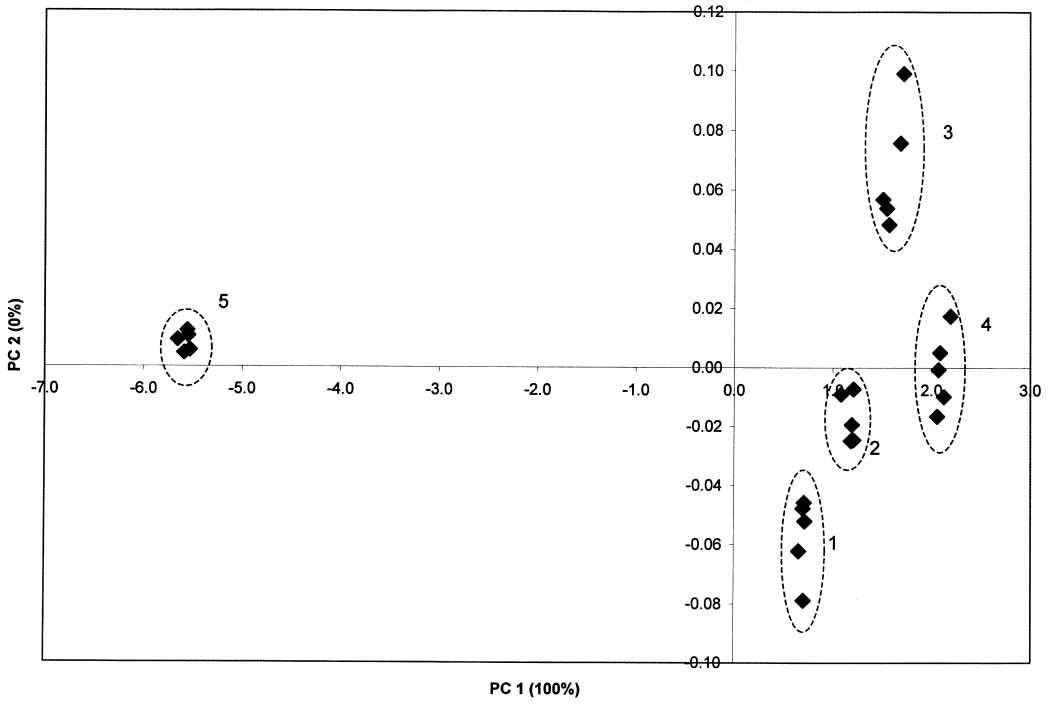


Fig. 11. Score plot of PCA performed on retention factors obtained by test 1 on five Luna C_{18} (2) columns packed with packing material coming from the same batch.

chemical modification of one batch of silica, the hydrophobicity of these three batches (first factor) is almost identical, whereas the selective retentivity (in this instance, the second factor is related to shape selectivity) is vastly affected by small fluctuations in the outcome of the chemical modification process. In test 2, the second factor is attributed to the secondary retention mechanism (selective retentivity) of the basic compounds. In this case, the effect of variations in chemical bonding process is somewhat less noticeable.

The results of tests 1 and 2 (Figs. 13 and 14) show that the dispersion of the points on the score plot along the abscissa is the same for the three batches of stationary phase made from the same batch of silica and for the five columns packed with material originating from the same batch. The corresponding variation can most probably be attributed to the packing procedure itself, since the columns numbered 1 to 5 were packed with the very same stationary phase. Only the packing procedure can alter the nonselective retentivity of these columns. Since columns 1–5 contain the same stationary phase, they show no variation along the ordinate at all. The effect of the secondary retention mechanism is significant only when columns are packed with stationary phases made from different batches of silica or chemical modification.

The results of PCA carried out on the data matrices for all the 15 Symmetry, 14 Luna and 11 Vydac columns indicate that the columns of each brand exhibit a similar behavior; the differences between columns of different batches are characterized by two factors. The first factor is related to the total column porosity, the second one to secondary interactions, i.e., to selective retention mechanisms. The stationary phases prepared as different bonding batches of the same lot of silica scatter along both axes in the score plot. For the Vydac columns this is illustrated in Fig. 15. For the Luna, Symmetry and Vydac columns, the chemical bonding alters both the nonselective and the selective retentivity. Hence, we cannot identify any consistent effect of the underlying silica on the variability of the retention factors.

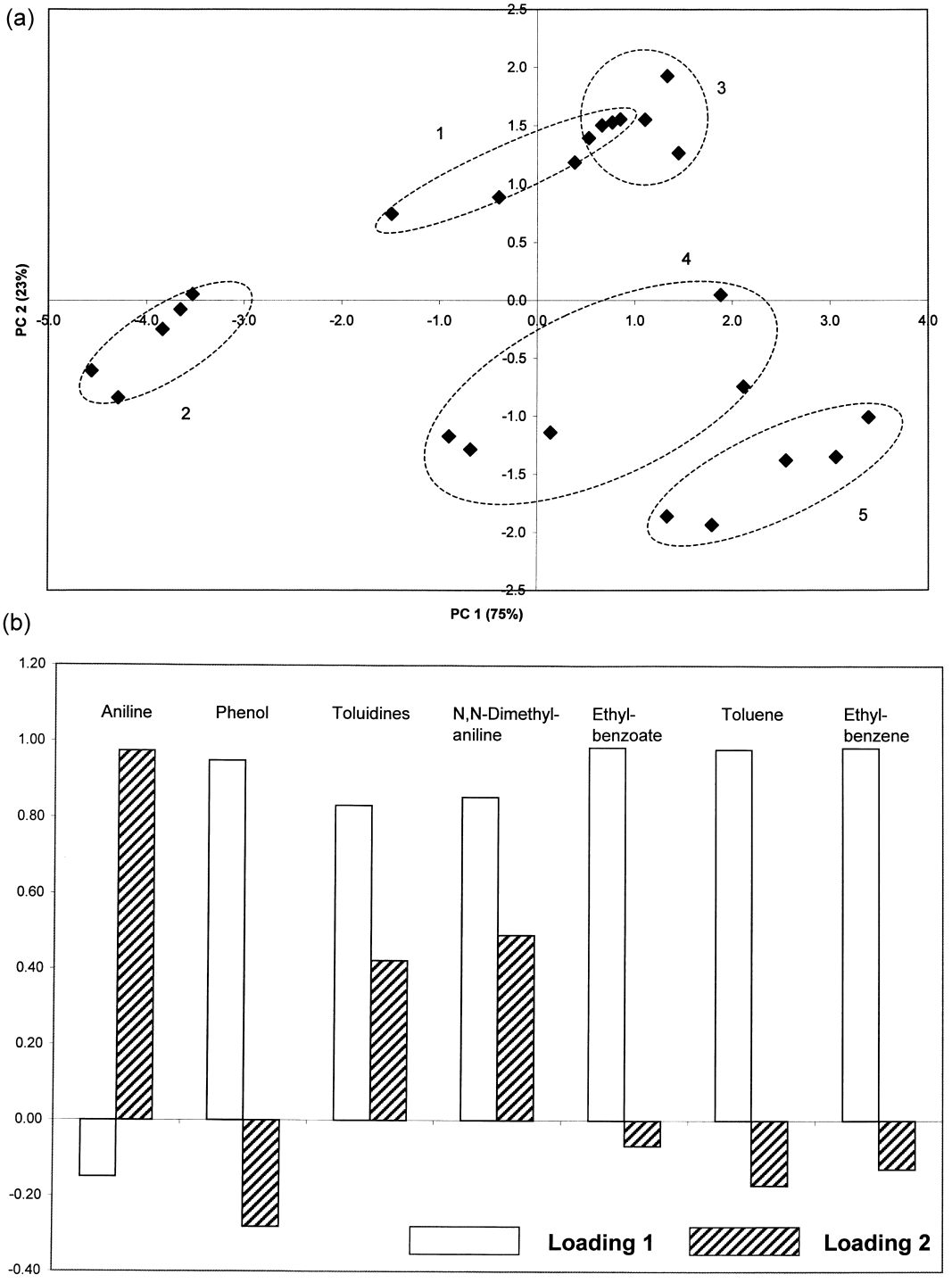


Fig. 12. (a) Score plot and (b) loading plot of PCA performed on retention factors obtained by test 2 on five Symmetry C_{18} columns packed with packing material coming from the same batch.

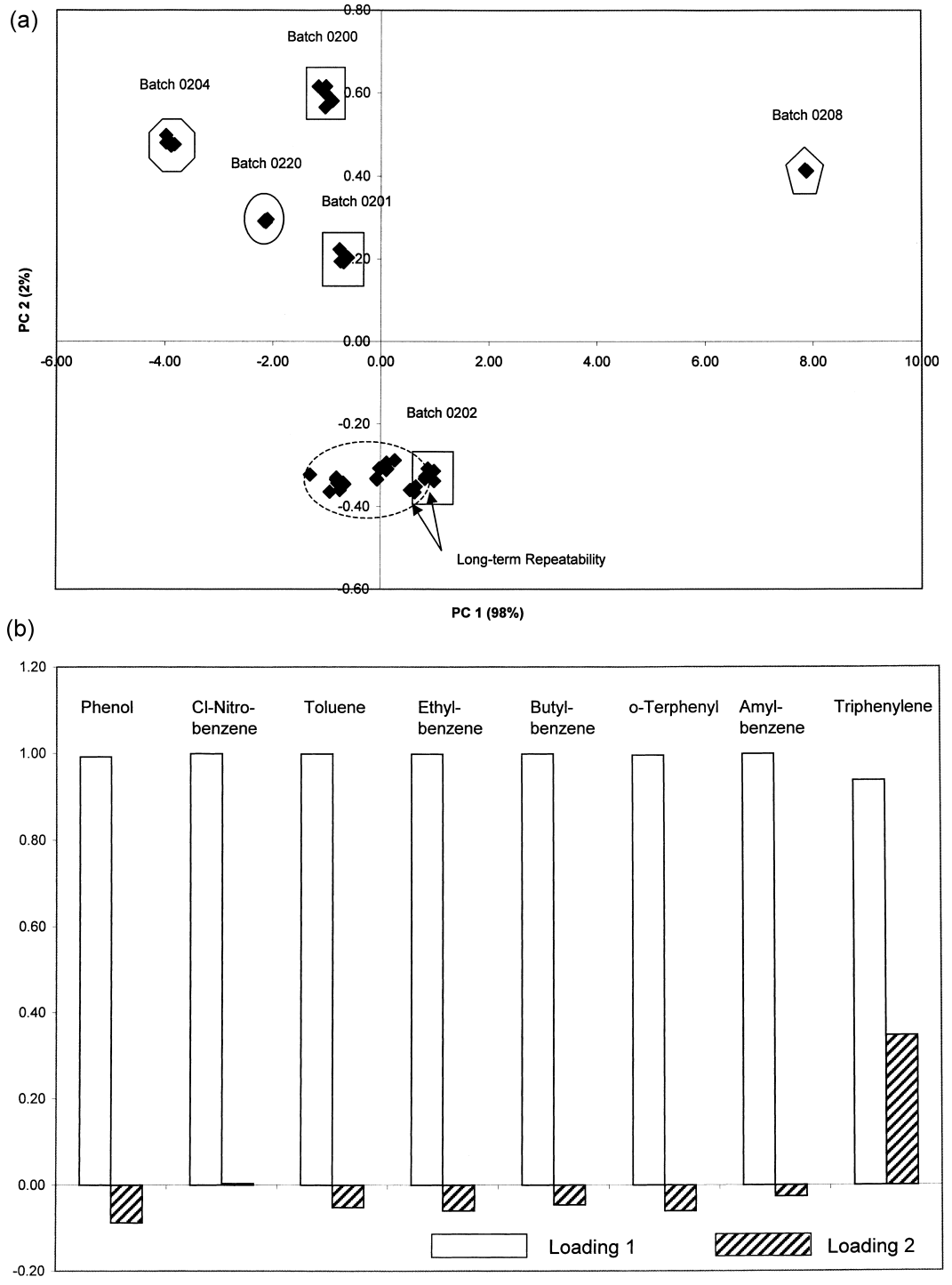


Fig. 13. (a) Score plot and (b) loading plot of PCA performed on retention factors obtained by test 1 on five Kromasil C_{18} columns packed with packing material coming from the same batch and on six columns from six different batches.

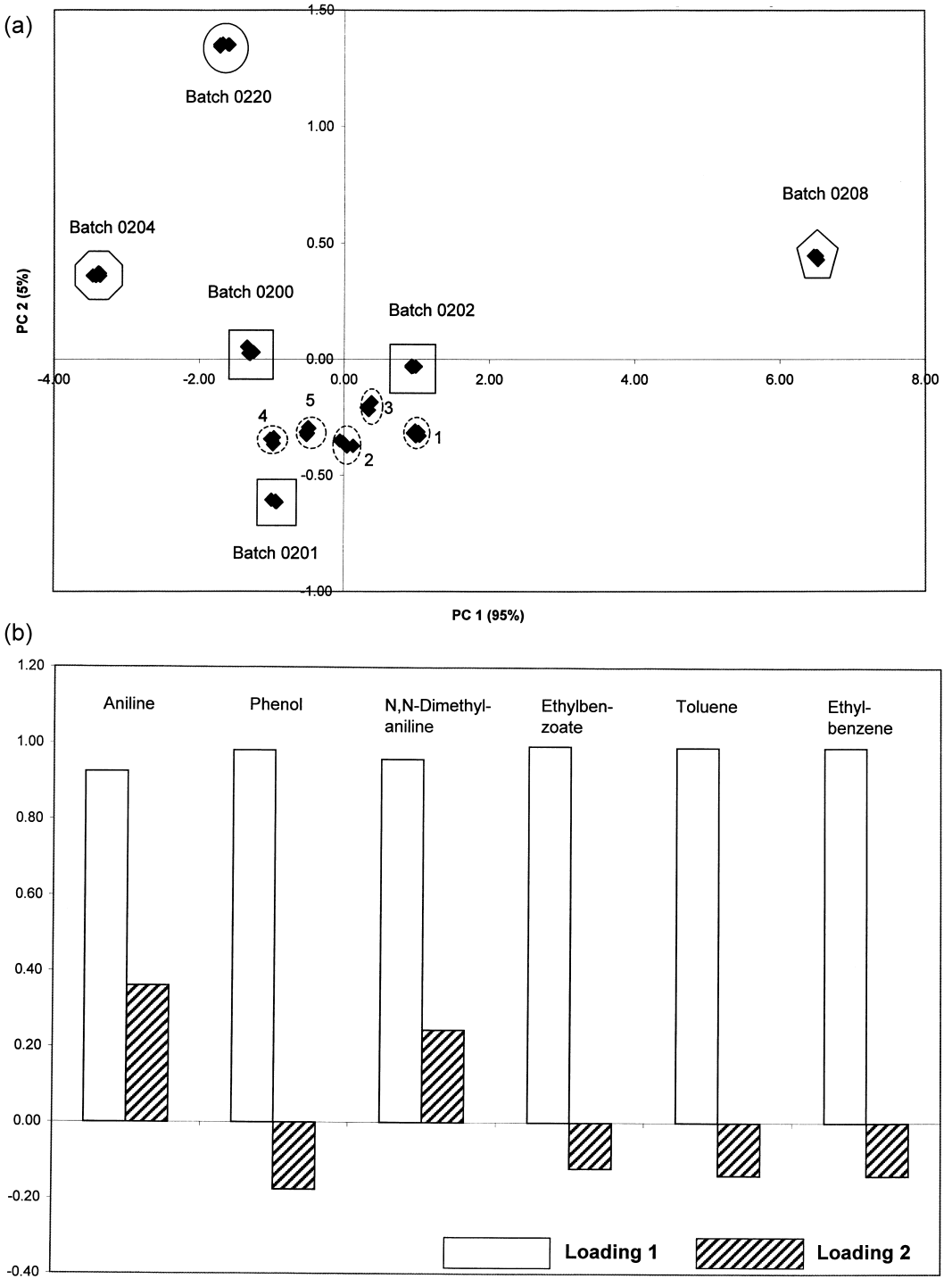


Fig. 14. (a) Score plot and (b) loading plot of PCA performed on retention factors obtained by test 2 on five Kromasil C_{18} columns packed with packing material coming from the same batch and on six columns from six different batches.

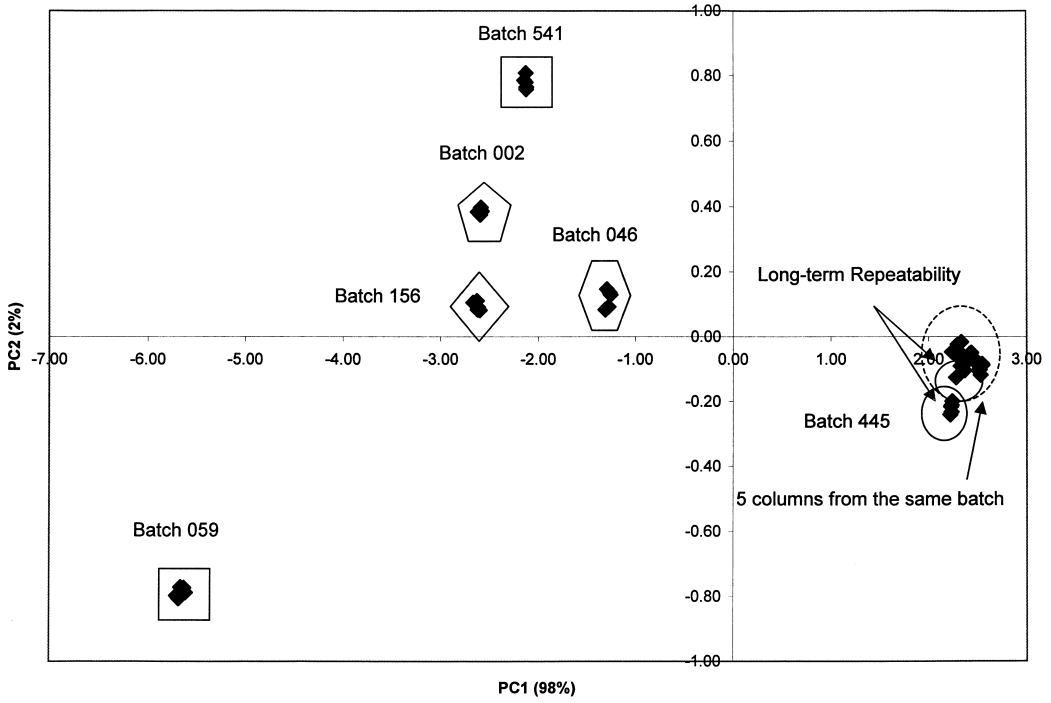


Fig. 15. Score plot of PCA performed on retention factors obtained by test 1 on five Vydac C_{18} (218TP54) columns packed with packing material coming from the same batch and on six columns from six different batches.

For the whole set of columns, the largest relative differences of the retention factors for the components of test 1 are 4.06% for the Symmetry, 9.27% for the Kromasil, 8.85% for the Luna and 14.75 for Vydac columns [2–4]. These values correspond to a range of relative porosity differences of 1.65% for the Symmetry, 3.75% for the Kromasil, 3.16% for the Luna and 4.98% for Vydac columns. These values are, respectively 11-fold, eightfold, 2.5-fold and 33-fold larger than those observed for columns of the same batch on these four brands (Fig. 2). The ratio of the dispersion ranges along the abscissa for all the data points and for the points corresponding to the five column sets are 8.5-fold for Symmetry, fivefold for Kromasil (Figs. 13 and 14), 2.1-fold for Luna and 20-fold for Vydac (Fig. 15), values which are in good agreement with those above.

This agreement between the ratios of the batch-to-batch and column-to-column reproducibilities derived from the score plots and from the retention factors is significant. The slight differences between the numbers in the two series is most probably due to the centering and scaling steps in the PCA. After completion of these operations, all the retention factors are normalized. It was observed that the reproducibility of a retention factor depends on the corresponding retention time. The peaks are not evenly distributed along the chromatogram. When the fluctuations of the retention factors are averaged, the segment of the chromatogram in which the peak density is higher weighs more heavily on the result which explains why the two methods lead to somewhat different results. In conclusion, the score plot is a valuable means to evaluate column-to-column or batch-to-batch reproducibility of the data.

4.5. Principal component analysis of the retention factors measured on the different brands

The physico-chemical properties of three brands of stationary phase (Symmetry, Kromasil and Luna) are very similar. The total carbon content, the surface coverage of the bonded phase, the average pore size and the specific surface area of the underlying silica

are all very close for these three brands [1–4]. For the Vydac material, the average pore size of the bare silica is large, 268 Å (as opposed to 90.8, 111.4, and 103.3 Å for Symmetry, Kromasil, and Luna, respectively). The specific surface area of the silica is much lower for Vydac RP C_{18}, 70.5 m^2/g than for Symmetry, Kromasil, and Luna, with 339.4, 320.5, and 401.0 m^2/g, respectively).

PCA was performed on the data matrix consisting of the retention factors for all the components of test 1 on the three sets (one per brand) of five columns (all from the same batch) with similar physico-chemical properties. The results show that two PCs explain the variations of these retention factors. As illustrated in Fig. 16a, the weight of these two factors are 88% and 12%, respectively. In Fig. 16, each brand is represented by five columns that are packed with stationary phases from the same batch. After excluding triphenylene (a planar polyaromatic hydrocarbon) from the data set, the relative importance of the second factor decreased to 6% (Fig. 16b). After excluding triphenylene and phenol (a weakly acidic compound) the weight of the second factor decreased to 1% (see Fig. 16c). After further excluding the values obtained for 1-Cl-nitrobenzene (a neutral but polar compound), which leaves only three alkylbenzene homologues (toluene, ethylbenzene and butylbenzene) only one factor completely describes the retention factor differences between the columns (Fig. 16d). This confirms that the second factor is related to secondary retention mechanisms (i.e., to differences between the equilibrium constants) on the different brands of columns. Note that in some instances the sign of the ordinate is changed in Fig. 16. This is a feature often observed in PCA. The relative position of the objects, however, is unchanged. Thus, the information obtained is not affected by this minor detail.

Finally, PCA was carried out on the data matrix consisting of all retention factors of the components of the first test mixture measured on all batches of the three different brands (Symmetry, Luna, Kromasil). The score plot is shown in Fig. 17a. Again, two factors are sufficient to explain the variations observed. The weights of these two factors are 88% and 12%, the same weights as found in the study discussed above, of the combination of the three sets (three brands) of five columns (one batch). Again, the weight of the second PC drops to zero when the initial data matrix is limited to the relative retentions of the alkylbenzenes (Fig. 17b). This further confirms that the second factor is related to specific retention mechanisms. After the addition of all the batches, the relative position of the three brands on the score plot (Fig. 17a) is unchanged compared to their position on the score plot for the five columns of the three brands (Fig. 16a). Therefore, we may assume that the PCs have the same physical meaning as when the retention properties measured on the five columns of the same batch were compared.

The physico-chemical properties of Vydac are different from those of the other three phases studied. Accordingly, one could expect that when the retention factors of all four brands are analyzed, at least a new PC would appear. As shown by the results plotted in Fig. 18, however, two factors are still sufficient. The relative position of the three brands along the abscissa is the same in Figs. 17 and 18 but the weight of the first factor increases from 88 to 97%. This change takes place because the nonselective retention is different on the Vydac material and on the other ones due to the different pore size and specific surface area. PCA is not able to distinguish between these effects but the results suggest that the change in the nonselective retention mechanisms can be explained with only one factor. We cannot observe the brand-to-brand change of the equilibrium constant, nor separate this effect from a change in the total bed porosity. This might be because the equilibrium constants are similar on all the reversed-phase liquid chromatography (RPLC) stationary phases studied, due to their high carbon content.

The second factor is again related to selective retention mechanisms because, after the exclusion of all the polar compounds, leaving the three alkylbenzene homologues, the second factor disappears, leaving only the first factor discussed above.

5. Conclusions

PCA demonstrates that the column-to-column reproducibility of retention time data depends on the column-to-column fluctuations of both the column

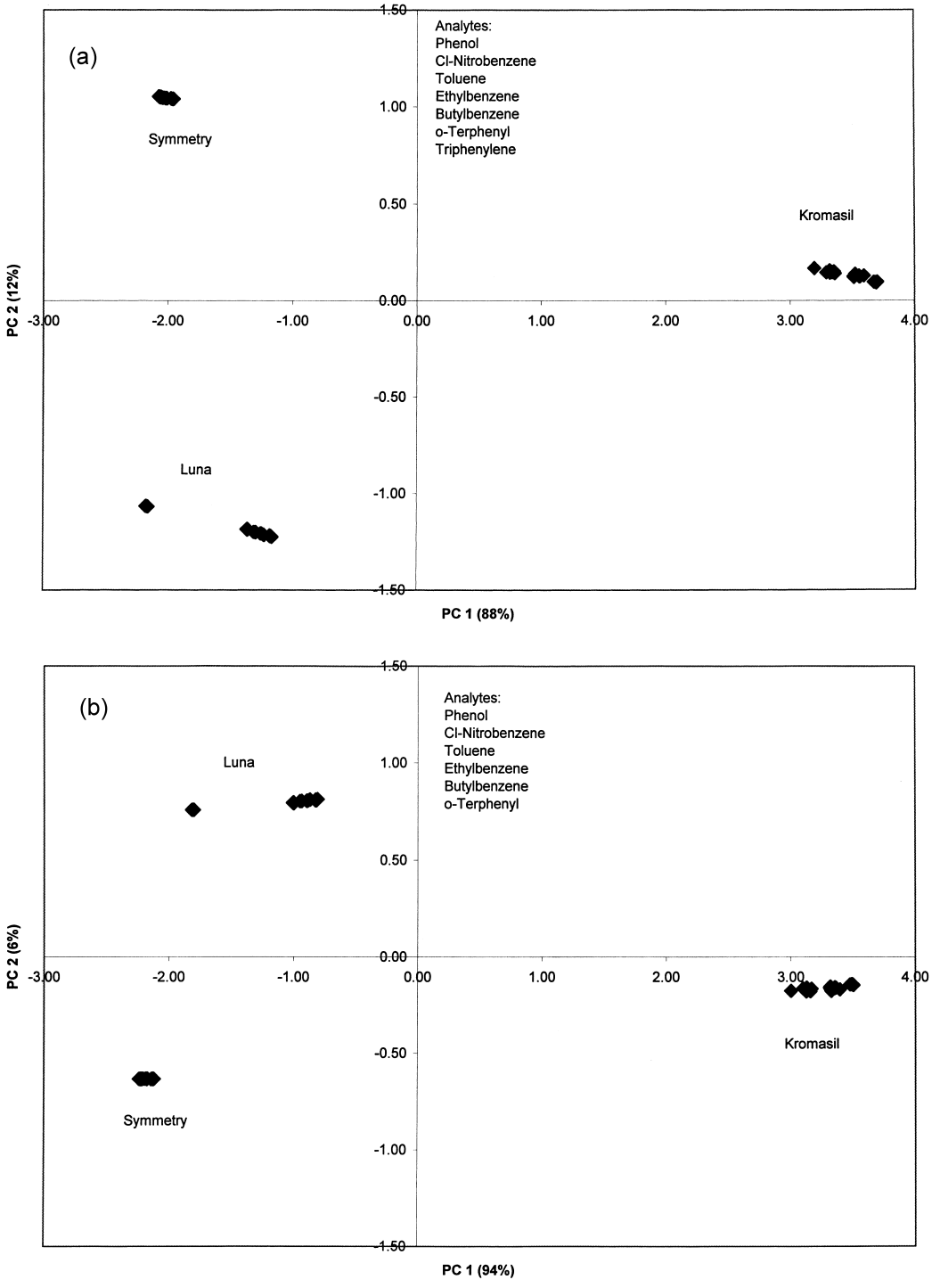
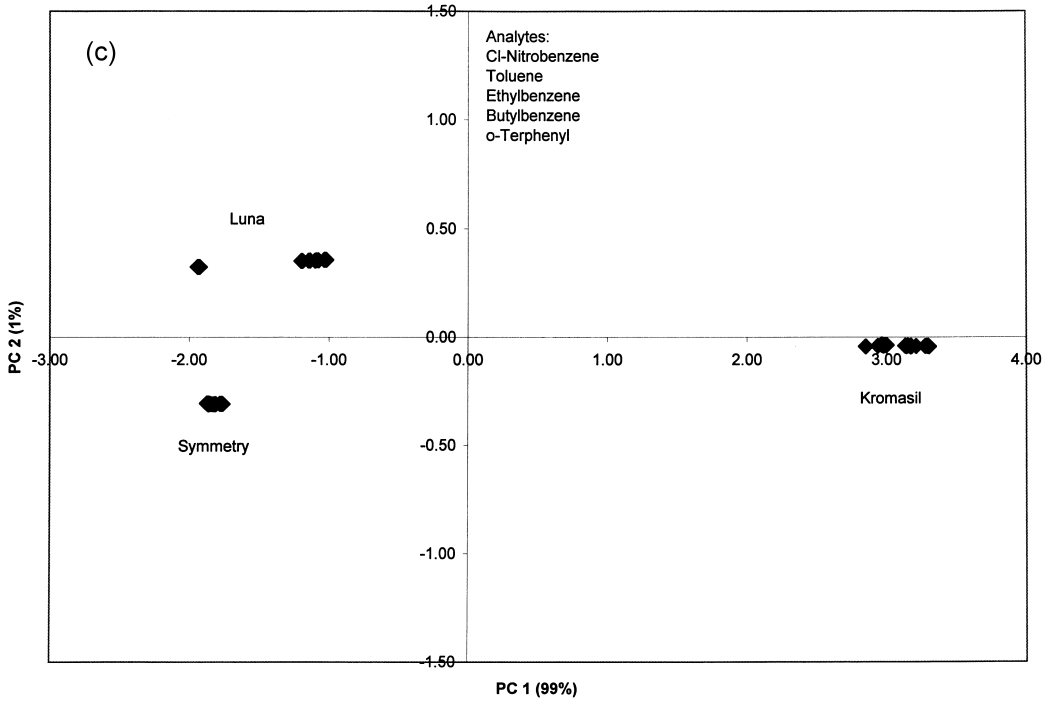


Fig. 16. Score plots of PCA performed on retention factors obtained by test 1 on five Symmetry $C_{18,}$ Kromasil C_{18}, and Luna C_{18} (2) columns packed with packing material coming from the same batch.



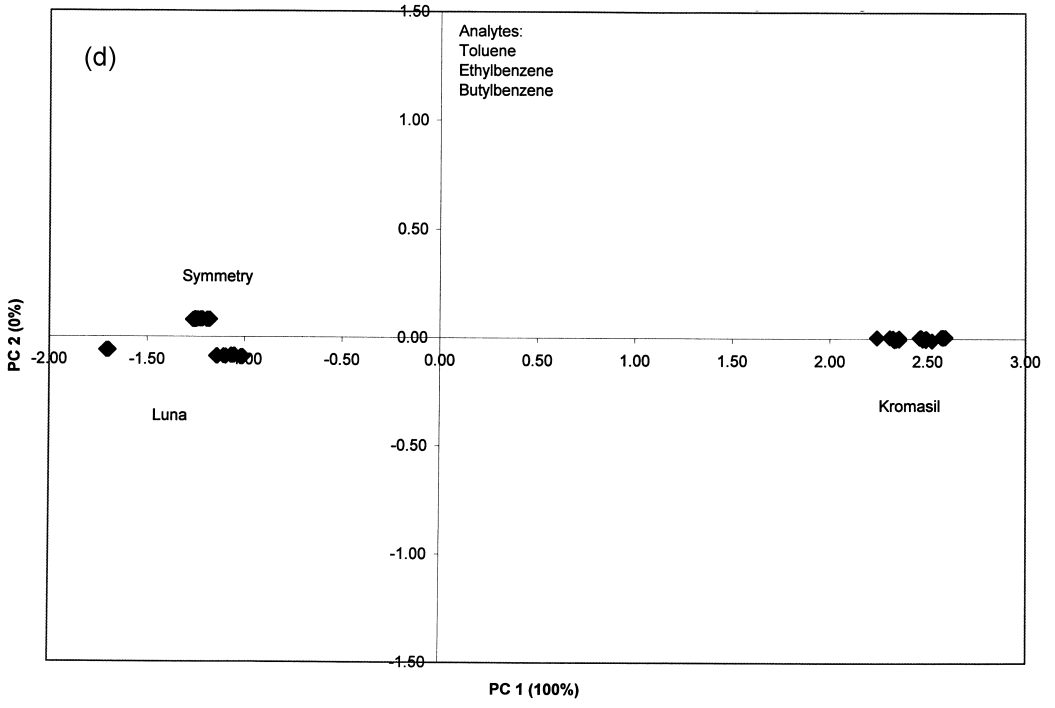


Fig. 16. (*continued*)

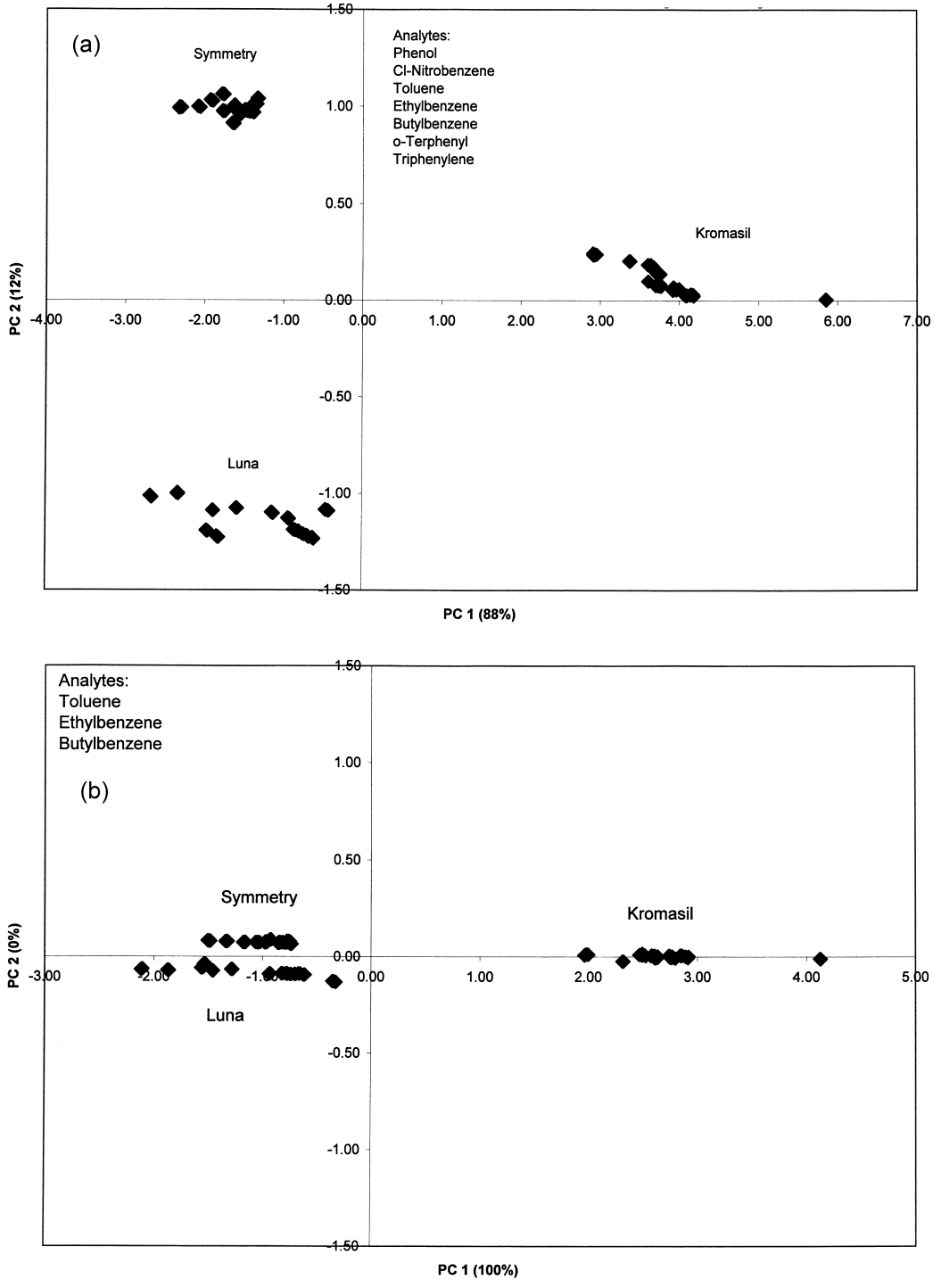


Fig. 17. Score plots of PCA performed on retention factors obtained by test 1 on five Symmetry $C_{18,}$ Kromasil C_{18}, and Luna C_{18} (2) columns packed with packing material coming from the same batch and on columns from different batches.

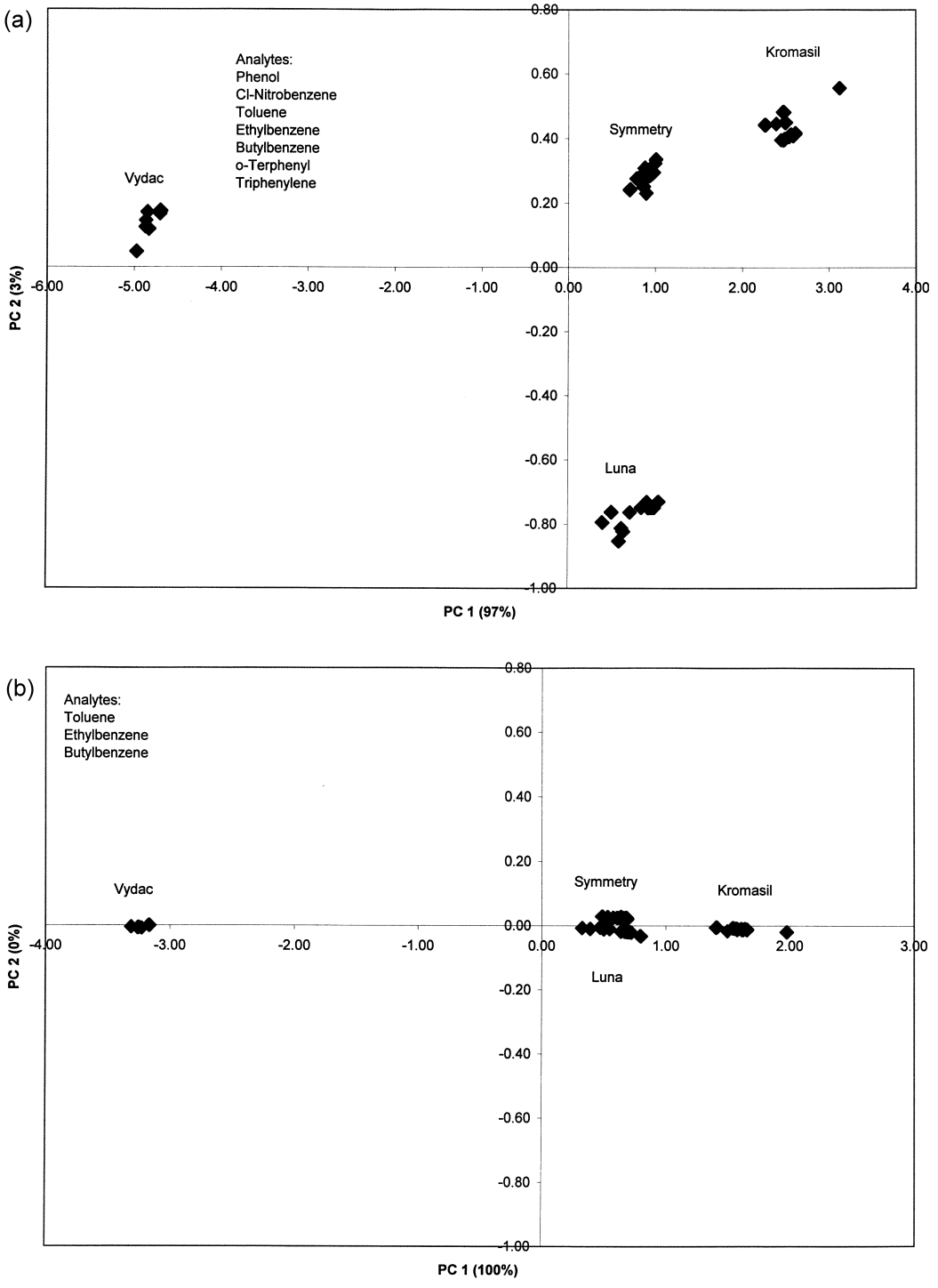


Fig. 18. Score plots of PCA performed on retention factors obtained by test 1 on five Symmetry $C_{18,}$ Kromasil C_{18}, Luna C_{18} (2), and Vydac RP C_{18} columns packed with packing material coming from the same batch and on columns from different batches.

volume and the packing density. Only two factors suffice to describe the variability within the data sets. Although in most cases the relative fluctuations of the column volume and the packing density are comparable, principal component analysis shows that, in general, the column volume fluctuations have a larger influence on the fluctuations of the retention times than the packing density variations. The score plots evidence that the short-term repeatability of the experiments is far better than the column-to-column reproducibility. It is rare that two columns have the same volume and the same bed porosity. Therefore, the data from nearly every column forms a separate cluster in the score plot when columns originating from the same batch are studied. When columns from different batches are subjected to PCA, the same-batch columns form a subset or cluster on the score plot, while the points corresponding to columns of different batches are scattered, demonstrating that there are significant differences between the batch-to-batch and the column-to-column reproducibility of retention times.

When the retention factors are analyzed, the effect of the column-volume fluctuations is eliminated and the packing density or the total porosity of the columns becomes the major factor that governs the variation of the retention data. In this case, again, at most two factors were needed to describe the variations. The first factor is the nonselective retention mechanism on the column and the second, a usually rather minor factor, characterizes the specific retention on the column.

Although our study included columns packed with materials of diverse pore sizes and specific surface areas, we found that, within the data set studied, there is only one factor that accounts for the variations of the hydrophobicity of the stationary phases. It is most probable that the packing density is this parameter; the equilibrium constants being very similar on all the stationary phases studied, due to their high carbon content and the similar properties of the underlying silica.

Acknowledgements

This work was supported in part by the cooperative agreement between the University of Tennessee and the Oak Ridge National Laboratory, by NATO Linkage Grant OUTR.LG971480, by grant T 025458 from the Hungarian National Science Foundation (OTKA), and by grant FKFP 0609/1997 from the Hungarian Ministry of Education. A.F. acknowledges the financial support of the Hungarian National Eötvös Foundation.

References

[1] M. Kele, G. Guiochon, J. Chromatogr. A 830 (1999) 41.
[2] M. Kele, G. Guiochon, J. Chromatogr. A 830 (1999) 55.
[3] M. Kele, G. Guiochon, J. Chromatogr. A 855 (1999) 423.
[4] M. Kele, G. Guiochon, J. Chromatogr. A 869 (2000) 181.
[5] M.F. Delaney, A.N. Papas, M.J. Walters, J. Chromatogr. 410 (1987) 31.
[6] S.J. Schmitz, H. Zwanziger, H. Engelhardt, J. Chromatogr. 544 (1991) 381.
[7] B.A. Olsen, G.R. Sullivan, J. Chromatogr. A 692 (1995) 147.
[8] R.J.M. Vervoort, M.W.J. Derksen, A.J.J. Debets, J. Chromatogr. A 765 (1997) 157.
[9] E. Cruz, M.R. Euerby, C.M. Johnson, C.A. Hackett, Chromatographia 44 (1997) 151.
[10] K. Kimata, K. Iwaguchi, S. Onishi, K. Jinno, R. Eksteen, K. Hosaya, M. Araki, N. Tanaka, J. Chromatogr. Sci. 7 (1989) 721.
[11] Á. Sándi, L. Szepesy, J. Chromatogr. A 818 (1998) 1.
[12] R.G. Brereton, D.V. McCalley, Analyst 123 (1998) 1175.
[13] R.J.M. Vervoort, M.W.J. Derksen, F.A. Maris, J. Chromatogr. A 678 (1994) 1.
[14] W.J. Welsh, W. Lin, S.H. Tersigni, E. Collantes, R. Duta, M.S. Carey, W.L. Zielinski, J. Brower, J.A. Spencer, T.P. Layloff, Anal. Chem. 68 (1996) 3473.
[15] E.R. Malinowski, Factor Analysis in Chemistry, Wiley, New York, 1991.
[16] S. Wold, K. Esbensen, P. Geladi, Chemometr. Int. Lab. Syst. 2 (1987) 37.
[17] A. Felinger, Data Analysis and Signal Processing in Chromatography, Elsevier, Amsterdam, 1998.
[18] H.F. Kaiser, Psychometrika 23 (1958) 187.