Principal component analysis of Kováts indices for carbonyl compounds in capillary gas chromatography

K. Héberger[a],*, M. Görgényi[b]

[a]*Institute of Chemistry, Chemical Research Center, Hungarian Academy of Sciences P.O. Box 17, H-1525 Budapest, Hungary*
[b]*Institute of Physical Chemistry, University of Szeged, P.O. Box 105, H-6701 Szeged, Hungary*

Abstract

Principal component analysis was performed on a data matrix consisting of Kováts indices of 35 aliphatic ketones and aldehydes. The calculations were carried out on the correlation matrices of Kováts indices. The Kováts indices were determined on capillary columns with four different stationary phases, namely bonded methyl- (HP-1), methylphenyl- (HP-50), and trifluoropropylmethylsiloxane (DB-210), as well as polyethylene glycol (HP-Innowax) at four different temperatures. It was found that one principal component accounts for more than 94% of the total variance in the data, indicating that the temperature does not change the dominant pattern in the data. The physical meaning attributable to the principal components, thus the most influential ones are as follows: the first principal component accounts for the boiling point (and/or the molecular mass) of carbonyl compounds whereas the second is responsible for the temperature dependence. The plots of component loadings showed a characteristic pattern (counterclockwise increasing temperature) whereas that of component scores showed a triangular structure and some groupings of oxo compounds. Abstract retention data (free from influence of temperature and column polarity) are non-linear functions of boiling points of solutes. Similarity among the solutes from the point of view of retention is represented by characteristic plots.   © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Principal component analysis; Retention indices; Chemometrics; Stationary phases, GC; Quantitative structure–retention relationships; Ketones; Aldehydes

1. Introduction

Application of multivariate data analyses [principal component analysis (PCA), factor analysis, cluster analysis and related techniques] to chromatography has become a popular method in the last decades [1–18], mainly because it can provide information not otherwise accessible. This informa-

tion includes e.g., classification, searching similarities, finding relationships, outlier detection, modeling, data reduction, finding physical significance to principal components, etc.

Multivariate techniques were devoted to analyze either gas or liquid chromatography data [1–5,7–11,14,15,18] or to evaluate reversed-phase high-performance liquid chromatography (RP-HPLC) data [6,12,13,16,17]. The main goal of these investigations was to classify and select proper stationary phases, to recommend preferred solvents to a given separation, to identify selectivity measures, to determine factors important in characterization of solvent properties, etc. It should be mentioned that

---

*Corresponding author. Tel.: +36-1-325-7900; fax: +36-1-325-7554.
E-mail addresses: heberger@cric.chemres.hu (K. Héberger), gorgenyi@chem.u-szeged.hu (M. Görgényi)

the results and/or conclusions of investigations mentioned are often contradictory.

Many examinations were related to the data set of McReynolds, i.e., retention indices of 10 solutes and 226 stationary phases [19], or were frequently limited to a special compound class. The reproducibility of Kováts retention indices surpasses the same of any other way of giving retention data. This is the reason why retention indices are so popular in multivariate analyses. Despite the different methods used to characterize the liquid phases, all the techniques found a single dominant factor in the data which was assigned to "polarity".

Despite their importance not much is known about the retention properties of ketones and aldehydes. A specific intramolecular interaction was observed between the neighboring CH_3 and $C=O$ groups [20,21]. The retention indices of ketones and aldehydes are linearly correlated to their boiling points [22,23] and other structure descriptors [24]. The retention index measurements are made either on a limited number of compounds or the influence of column polarity and column temperature has not been studied in detail [21,25–28]. The carbonyl compounds are chosen as model compounds less frequently [29–33], although some thermodynamic characteristics (such as solution heat values) of oxo compounds have already been calculated [34].

To the best of our knowledge, no PCA results have been published on temperature dependence of retention data so far.

The activity outlined in this paper can be summarized as measurement of retention indices of ketones and aldehydes in the function of temperature and column polarity, as well as mathematical description of retention indices on the base of PCA.

Our aim was to characterize the retention behavior of ketones and aldehydes on stationary phases of different polarities and to obtain information on whether the branching in the carbon chain, the position of $C=O$ group (ketone or aldehyde), unsaturation, etc., has a significant effect on the retention. Another aim was to examine whether it is possible to extract useful information in the case of highly correlated data. We tried to order (arrange) the functions influencing the retention (temperature, column polarity). Besides these, our aims were (i) to extract temperature- and polarity-independent reten-

tion data; (ii) to determine which molecular structural descriptors correlate with these temperature- and polarity-independent retention data and, if possible, (iii) to assess physical significance to principal components. Last but not least, we would like to classify ketones and aldehydes from the point of view of chromatographic retention.

2. Principles of PCA

In the course of defining principal components the original variables are transformed into new ones. Original variables are presented as linear combinations of principal components. The values of principal components are component scores. The linear coefficients are called the component loadings, i.e., the correlation coefficients between the original variables and the principal components. The algorithms for PCA can be found in standard chemometrics books and tutorials [10,35]. The principal components are orthogonal (independent) and they are scaled so that their variances are equal to unity. Further on, they are ordered in such a way that the variance explained by the first principal component is the greatest, the variance explained by the second one is smaller, and so on, whereas that of the last one is the smallest.

The columns of data matrices under study are intercorrelated, i.e., the data are redundant. PCA [10,35] makes use of the intercorrelations by starting from the correlation matrix of the variables, and it eliminates the redundancy from the data, i.e., it reduces their dimensionality by revealing several underlying components. The underlying components are represented by new variables called principal components. In other words, PCA is able to decompose the original input matrix into a sum of multiplication of loading and score vectors.

A basic assumption in the use of PCA is that the score and loading vectors corresponding to the largest eigenvalues contain the most useful information relating to a specific problem and that the remaining ones comprise mainly the noise, i.e., for a practical problem it is sufficient to retain only a few components accounting for a large percentage of the total variance [35].

3. Experimental

In this study selected aliphatic ketones and aldehydes were chosen as model compounds. They were purchased from Fluka and Aldrich and were used without further purification. The retention measurements were done by a 5890 Hewlett-Packard Series II gas chromatograph with HP3365 ChemStation. A flame ionization detector and split/splitless injector in split mode were used. The carrier gas was nitrogen. A constant inlet pressure was used and the carrier flow-rate was set to the minimum range of the height equivalent to a theoretical plate (HETP).

The characteristics of capillary columns are summarized in Table 1.

The retention indices were measured at temperatures of 50, 70, 90 and 110°C making three injections at each temperature. The experimental Kováts indices were evaluated by the well known method from retention data of *n*-alkanes with dead time corrections.

The input data set was arranged in columns with increasing polarity (HP-1<HP-50<DB-210<HP-Innowax) and, within one individual stationary phase, with increasing temperature. Similarly, the row arrangement corresponds the Kováts retention indices of 35 carbonyl compounds (ketones and aldehydes) with their boiling points (i.e., the solutes were the mathematical–statistical cases or objects). Hence, every matrix element corresponds to a Kováts index of a given solute on a given stationary phase and at a given temperature.

Each data set to be analyzed was essentially a data matrix containing the Kováts indices. The entire input matrix is given in Table 2.

The columns contained the Kováts indices at different temperatures (323 K, 343 K, 363 K, 383 K) and on different stationary phases (HP-1, HP-50, DB-210, HP-Innowax). That is, the columns correspond to variables (measured quantities), i.e., Kováts indices on a given stationary phase and at a given temperature; 16 columns in total.

First, the correlation matrix of the original variables was computed (this matrix contained the correlation coefficients of each original variable with the others). Then, the component loadings were calculated from the eigenvectors and eigenvalues of the correlation matrix. Finally, the component scores were obtained from the component loadings and the original variables. The computer program used [36] furnishes the variances of the principal components, the loading patterns, the score coefficients for the standardized variables at mean zero and at standard deviation one, and the component scores.

The following physical constants (structural descriptors), boiling point (T_B), molar volume (V_m), molecular mass (M_r), molar refraction (R_m), octanol–water partition coefficient (log P) considered as independent variables were used to search correlation with the retention data: the first four variables stem from the Handbook of Chemistry and Physics [37], whereas the last one is from Hansch and Leo's book [38]. The log P values were calculated using the additivity in cases when no measured log P values were available.

4. Results and discussion

Repeated PCAs were made in the following order:

4.1. Elimination of temperature dependence

Firstly, the temperature dependence was eliminated from the Kováts indices. The original input matrix was divided into four submatrices of equal

Table 1
Characteristics of capillary columns used

Notation	Stationary phase[a]	Inner diameter (mm)	Film thickness (μm)	Length (m)	Polarity[b]
A=HP-1	Dimethylpolysiloxane	0.32	1.05	50	222
B=HP-50	50% Phenylmethylpolysiloxane	0.32	0.5	30	884
C=J&W DB-210	50% Trifluoropropylmethylpolysiloxane	0.32	0.5	30	1520
D=HP-Innowax	Polyethylene glycol	0.32	0.5	30	2308

[a] All phases are bonded.
[b] Sum of the first five McReynolds constants [19].

Table 2
Kováts indices of ketones (1–19) and aldehydes (20–35)[a]

Compound No.[b]	A1	A2	A3	A4	B1	B2	B3	B4	C1	C2	C3	C4	D1	D2	D3	D4	TBP
1	469.7	469.3	469.4	470.2	606.3	603.7	601.9	600.7	792.9	799.6	807.7	816.6	835.0	837.5	840.8	843.5	329.4
2	574.7	575.0	574.8	575.8	711.6	710.9	709.7	710.4	882.1	889.8	898.5	906.8	919.8	923.7	927.8	932.3	352.8
3	639.9	640.5	641.5	642.8	767.1	767.2	767.3	767.8	943.3	951.8	960.9	970.7	949.4	954.4	959.6	965.0	367.7
4	675.4	675.5	675.8	676.9	809.8	809.6	809.6	810.4	960.8	968.1	976.1	984.1	996.9	1001.5	1006.3	1011.3	374.9
5	665.4	665.7	666.1	666.9	799.3	798.6	798.6	799.3	973.9	981.8	991.1	1000.7	996.2	1000.7	1005.6	1010.7	375.2
6	691.8	693.3	695.5	697.4	808.0	808.5	809.7	811.1	992.0	1000.9	1010.9	1021.2	968.5	974.3	980.2	986.3	379.2
7	720.1	720.7	721.7	722.9	841.8	841.6	842.0	842.5	1027.1	1035.3	1045.1	1054.7	1025.2	1029.0	1033.9	1039.5	390.0
8	733.5	734.9	736.8	739.0	859.6	860.5	862.2	864.2	1036.1	1045.7	1056.0	1066.8	1033.9	1040.8	1047.6	1054.3	391.2
9	777.7	779.4	781.1	783.1	881.0	882.2	883.4	885.3	1038.5	1046.3	1054.5	1064.4	1014.8	1020.8	1026.9	1033.2	397.7
10	764.1	764.4	765.0	766.0	893.6	893.8	894.5	895.2	1048.4	1055.7	1064.2	1073.0	1068.0	1073.3	1078.9	1084.5	398.2
11	767.0	767.3	768.0	769.0	901.3	901.6	902.1	903.0	1081.5	1090.2	1100.2	1110.4	1097.2	1102.2	1107.6	1113.0	401.2
12	851.8	852.8	853.7	855.1	976.1	976.6	977.2	978.4	1134.5	1142.6	1151.1	1160.3	1139.4	1145.0	1151.1	1157.1	417.2
13	835.4	836.1	837.2	838.4	964.3	964.7	965.1	966.1	1161.3	1170.0	1181.1	1192.2	1156.1	1160.9	1166.3	1172.1	417.2
14	864.9	865.4	866.6	867.3	994.4	994.8	995.4	996.9	1153.6	1161.2	1170.6	1180.0	1167.2	1173.2	1179.9	1185.0	420.2
15	867.5	867.9	868.7	869.9	1000.1	1002.8	1003.3	1004.2	1184.3	1193.4	1203.9	1214.6	1195.8	1201.9	1207.6	1213.8	424.6
16	917.5	918.7	920.1	921.6	1031.9	1032.9	1034.4	1035.6	1194.2	1202.0	1210.4	1220.1	1178.7	1185.0	1191.0	1197.5	431.2
17	921.7	922.9	924.5	926.1	1041.5	1042.3	1043.7	1045.4	1206.9	1214.2	1224.9	1234.8	1200.1	1205.5	1213.0	1219.5	432.5
18	964.8	965.4	966.0	967.3	1094.9	1095.4	1096.3	1097.4	1255.5	1264.0	1272.9	1282.6	1265.5	1271.4	1277.9	1284.3	440.2
19	1051.4	1052.4	1053.5	1055.1	1175.4	1176.1	1177.3	1178.8	1342.6	1351.2	1360.4	1370.3	1334.1	1340.6	1347.3	1353.9	461.6
20	360.4	360.6	360.4	360.9	487.9	485.3	484.3	484.9	630.4	636.3	641.5	649.2	715.8	716.6	717.8	718.8	294.0
21	472.7	472.7	473.0	473.6	604.5	603.1	601.1	601.3	739.4	746.2	753.0	761.6	808.8	810.4	812.9	816.5	322.0
22	462.8	462.6	462.9	463.5	603.3	603.9	603.8	605.6	743.7	751.4	760.3	768.4	867.0	869.4	871.8	876.1	326.2
23	540.3	540.9	541.6	543.0	660.5	659.6	658.3	659.4	803.7	810.8	819.5	828.4	830.4	836.8	838.6	842.7	336.2
24	571.1	571.9	572.5	573.3	702.5	702.9	702.6	703.3	843.1	851.1	859.9	869.0	894.8	897.8	901.8	907.2	348.9
25	581.7	583.3	584.5	586.1	681.0	680.9	681.6	682.1	841.6	849.7	858.2	867.1	822.6	826.7	831.1	835.6	350.7
26	635.0	636.4	637.5	639.2	757.6	758.0	758.7	760.1	912.8	922.2	932.4	943.0	936.0	940.3	945.6	951.0	365.7
27	645.3	646.9	648.2	650.0	767.5	767.9	768.6	770.5	913.3	922.0	932.0	941.9	931.2	936.7	942.8	949.3	365.7
28	674.4	675.2	676.2	677.3	807.4	807.7	808.5	809.5	953.8	963.0	972.9	983.3	998.1	1002.8	1007.4	1013.0	376.2
29	623.4	624.5	625.8	627.3	787.7	789.6	791.6	794.3	967.2	980.2	993.0	1006.4	1061.5	1069.3	1077.1	1086.2	377.7
30	689.1	691.4	694.0	697.0	803.7	805.5	807.3	809.3	978.4	989.2	1000.7	1012.3	968.6	974.8	981.4	988.3	378.1
31	742.1	744.0	746.3	748.9	862.7	864.8	866.8	869.1	1009.6	1019.3	1030.6	1041.5	1018.0	1025.3	1032.7	1040.3	391.2
32	776.5	777.2	778.5	780.0	909.9	910.5	911.4	912.6	1059.3	1068.6	1079.4	1090.1	1098.3	1104.0	1110.0	1114.6	401.2
33	877.2	878.7	880.0	881.7	1009.8	1011.4	1013.3	1013.8	1162.7	1173.2	1183.9	1194.8	1199.6	1205.3	1211.5	1217.8	426.0
34	933.2	935.4	937.5	940.1	1049.3	1051.4	1053.2	1055.4	1205.4	1217.4	1228.6	1240.8	1197.8	1205.9	1213.1	1220.6	434.1
35	977.8	979.7	981.2	983.0	1110.9	1111.5	1112.7	1114.4	1265.5	1275.1	1287.6	1299.8	1298.8	1306.1	1313.0	1319.5	444.2

[a] Notes: A1=HP-1, 50°C; A2=HP-1, 70°C; A3=HP-1, 90°C; A4=HP-1, 110°C; B1=HP-50, 50°C; B2=HP-50, 70°C; B3=HP-50, 90°C; B4=HP-50, 110°C; C1=DB-210, 50°C; C2=DB-210, 70°C; C3=DB-210, 90°C; C4=DB-210, 110°C; D1=HP-Innowax, 50°C; D2=HP-Innowax, 70°C; D3=HP-Innowax, 90°C; D4=HP-Innowax, 110°C. TBP=boiling points in K.

[b] Compounds: 1=acetone, 2=2-butanone, 3=3-methyl-2-butanone, 4=3-pentanone, 5=2-pentanone, 6=2,2-dimethyl-3-butanone, 7=4-methyl-2-pentanone, 8=3-methyl-2-pentanone, 9=2,4-dimethyl-3-pentanone, 10=3-hexanone, 11=2-hexanone, 12=4-heptanone, 13=5-methyl-2-hexanone, 14=3-heptanone, 15=2-heptanone, 16=2-methyl-3-heptanone, 17=5-methyl-3-heptanone, 18=3-octanone, 19=5-nonanone, 20=acetaldehyde, 21=propanal, 22=acrolein, 23=isobutanal, 24=butanal, 25=trimethylacetaldehyde, 26=isovaleraldehyde, 27=2-methylbutanal, 28=valeraldehyde, 29=2-butenal, 30=3,3-dimethylbutanal, 31=2-ethylbutanal, 32=hexanal, 33=heptanal, 34=2-ethylhexanal, 35=octanal.

size. Four PCAs were made, each of (4×35) at a given polarity of stationary phase. In these cases the four temperatures were ordered as columns in the input matrix. Only two factors were retained in the model. The first factors for each stationary phase were of overwhelming importance whereas the second factors were of subordinate importance. One principal component explained 99.9938–99.9985% of variance in the data. The loading plot for HP-1 stationary phase can be seen in Fig. 1 as an illustra-
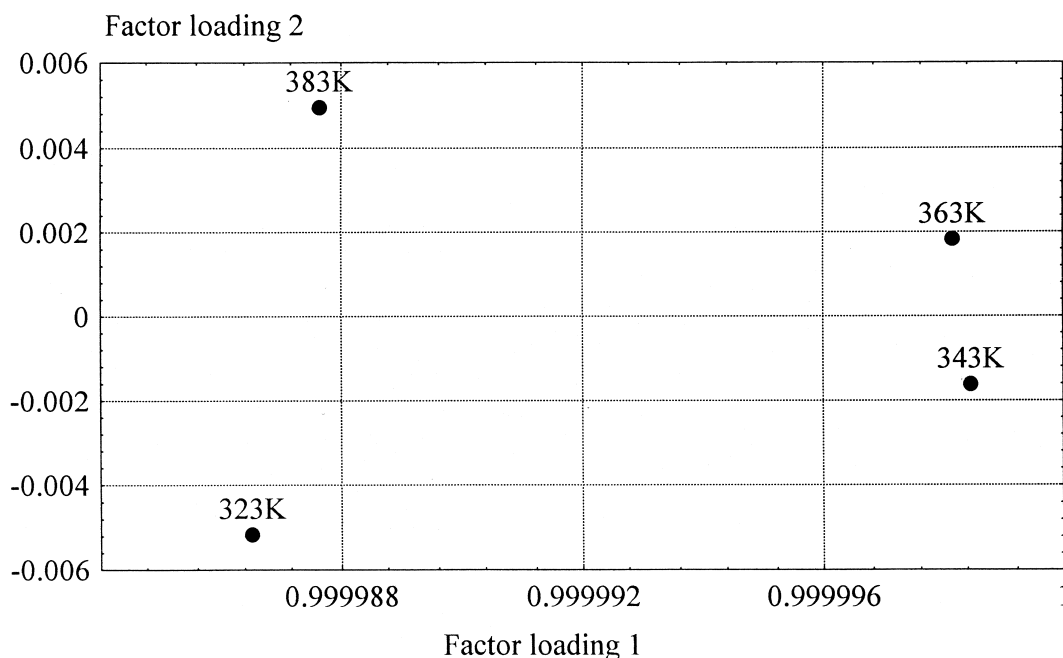
Fig. 1. Plot of factor loadings for temperature dependence of HP-1 stationary phase.

tion. The pattern observed is the same for all of the stationary phases investigated. The temperature increases counterclockwise and the loadings for the two medium temperatures (343 K and 363 K) are closer to each other than the loadings for the limit temperatures (323 K and 383 K). The physical reason for this pattern is not well understood. PCA tries to arrange the loadings on the contour of an ellipse if two principal components are retained. Anyway, it is an empirical, experimental observation. The loading values are, in fact, correlation coefficients. How closely related are the Kováts indices of carbonyl compounds at different temperatures? This can be seen by the closeness of loading values (factor 1) to unity.

How closely related are the retention data of carbonyl compounds at different polarities of the stationary phase? This can be seen by the correlations between score values to each other. As the polarity of phases increases (A–D) the correlation diminishes. The high correlation coefficients, however, indicate the close similarity between phases (Table 3).

This way useful information was extracted from highly correlated data.

4.2. Elimination of polarity dependence

Secondly, the effect of polarity (various stationary phases) were extracted by PCA of principal components of the first PCA. The input matrix of the second PCA either (i) consisted of the first PCs of the four temperature dependent PCA or (ii) consisted of the first *two* PCs of the four temperature dependent PCA, i.e., either (4×35) or (8×35) matrices were analyzed.

Table 3
Correlations between score values of the first PCA, $N = 35$[a]

	A score 1	B score 1	C score 1	D score 1
A score 1	1.00000	0.99756	0.99224	0.97076
B score 1	0.99756	1.00000	0.99554	0.98455
C score 1	0.99224	0.99554	1.00000	0.98392
D score 1	0.97076	0.98455	0.98392	1.00000

[a] A, B, C and D are the notations for column polarity, see Table 1.

The result in case (i) is that the first principal component explained 99.06%, whereas the second 0.785% and the third 0.1514% of the total variance in the data.

The loading plot can be seen in Fig. 2.

The proximity of points shows the similarity of stationary phases. The silicon phases form a subgroup whereas the Innowax is an outlier. The closeness of HP-50 and DB-210 is ostensible only, their points are far away in a three-dimensional space.

The result in case (ii) is that the first principal component explained 49.58%, whereas the second 40.13% and the third 5.05% of the total variance in the data. In the latter case three principal components had to be retained. Ninety-eight percent description is required and generally accepted in the field of chromatography [10] when analyzing Kováts retention indices, Table 4.

It can well be seen in the Table 4 that in case (i) one PC is overwhelming whereas in case (ii) the first two PCs, which are approximately of the same importance, are necessary for the description. A comparison of variance explained by the first and repeated PCAs suggest that the polarity of columns exerts more effect than the temperature of columns.

4.3. Physical significance of principal components

Similarity among the solutes from the point of view of retention is represented by characteristic plots. These score plots are shown in Figs. 3 and 4.

As it can be seen from the figures the aldehydes and ketones cannot be separated by the PCAs applied completely. Fig. 4, however, speaks well for an almost complete discrimination. The points for aldehydes take place on the outer part of the triangular structure, at left and right top. The exceptions are the 2,4-dimethyl-3-pentanone, 3-hexanone and 2-heptanone. Fig. 3 shows that an information loss takes place if omitting the information of second PC that explains 0.79% of the total variance in the data only. The score 1 explaining the largest variance can be considered as a descriptor of retention which is independent of temperature and independent of stationary phase polarity. The proximity of points shows similarity of compounds. No doubt, that butenal is an outlier. Similarity of some point pairs is
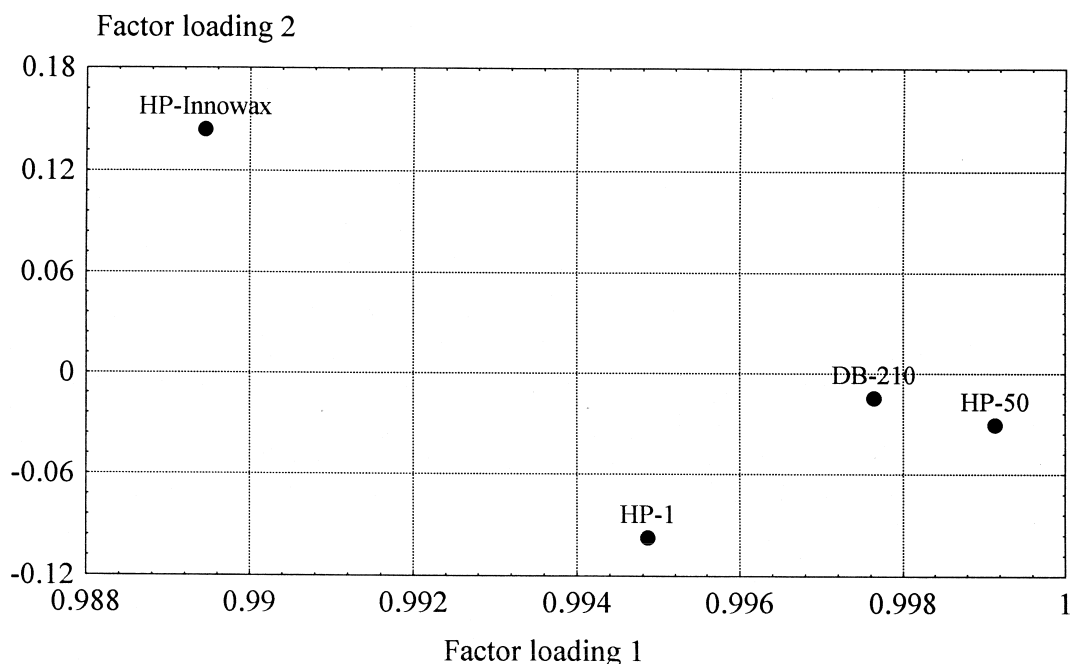

Fig. 2. Plot of factor loadings for polarity (stationary phase) dependence, case (i), further analysis of first principal components.

Table 4
Comparison of results of the second PCA

|	Loading 1	Loading 2	Loading 3
Case (i) PCA of the first principal components, an analysis of (4×35) matrix			
F1 (HP-1)	0.994863	−0.097549	−0.025714
F1 (HP-50)	0.999141	−0.030679	−0.025482
F1 (DB-210)	0.997643	−0.014785	0.067003
F1 (HP-Innowax)	0.989448	0.143970	−0.015972
Explained variance	3.962333	0.031403	0.006055
Proportional of total variance (%)	99.06	0.785	0.151
*Case (ii) PCA of the first **two** principal components, an analysis of (8×35) matrix*			
F1 (HP-1)	0.994248	0.028430	−0.086792
F2 (HP-1)	−0.062747	0.898971	−0.293934
F1 (HP-50)	0.998864	0.021241	−0.029568
F2 (HP-50)	−0.020162	0.914719	−0.011927
F1 (DB-210)	0.997289	0.029656	−0.005340
F2 (DB-210)	−0.026381	0.843887	0.515383
F1 (HP-Innowax)	0.989886	−0.003723	0.121279
F2 (HP-Innowax)	0.023566	0.922471	−0.169188
Explained variance	3.966314	3.210111	0.403928
Proportional of total variance (%)	49.58	40.13	5.05
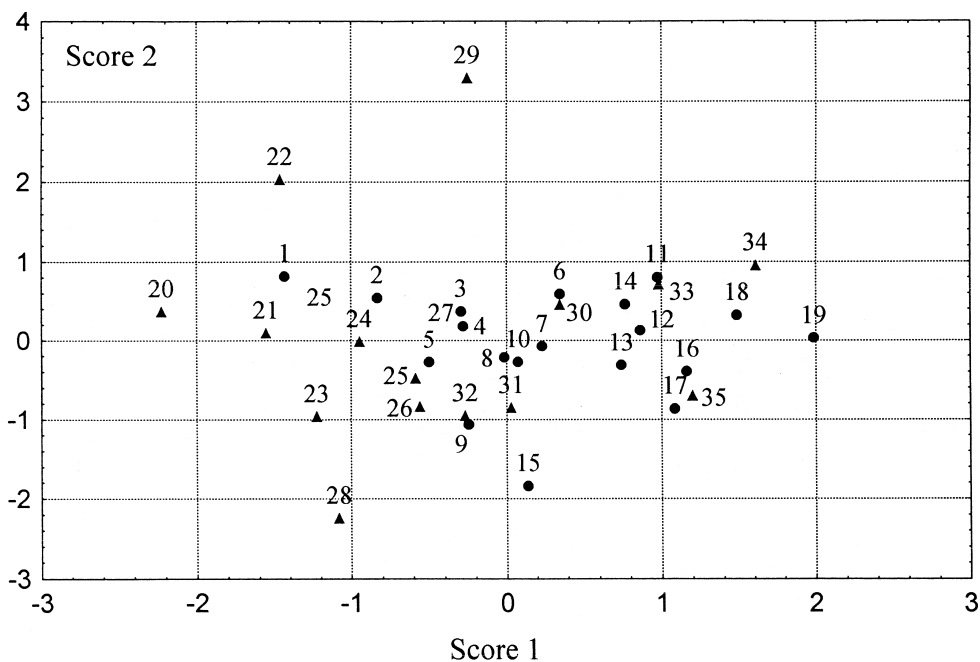
Fig. 3. Plot of factor scores, case (i), further analysis of first principal components. Notation is given in Table 1, (●) ketones, (▲) aldehydes.
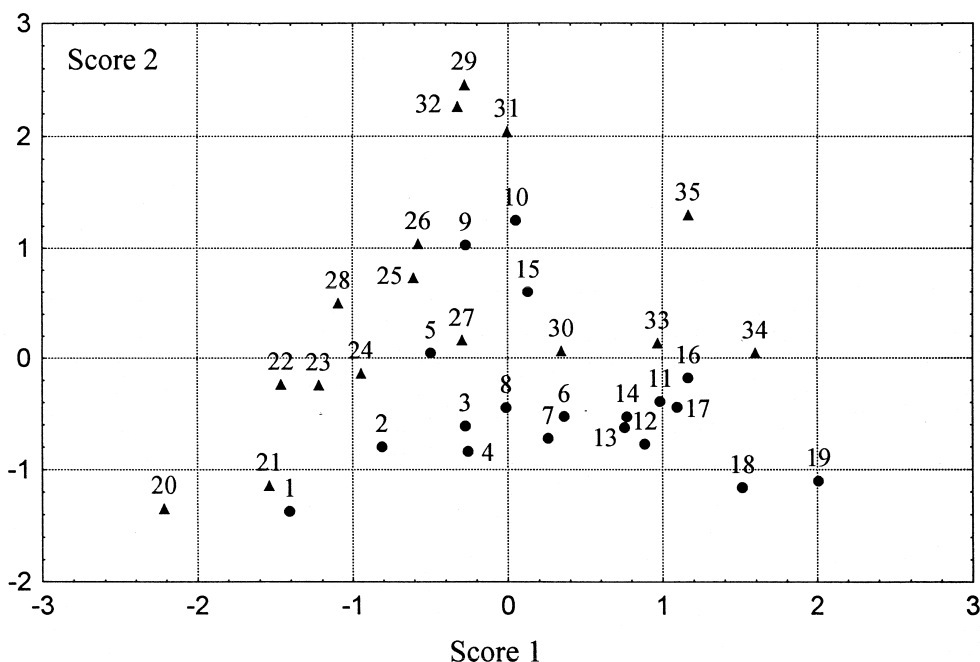
Fig. 4. Plot of factor scores, case (ii), further analysis of first *two* principal components. Notation is given in Table 1, (●) ketones, (▲) aldehydes.

trivial, e.g., 2,2-dimethyl-3-butanone and 3,3-dimethylbutanal as well as 2-hexanone and heptanal. The coincidence of 2,4-dimethyl-3-pentanone with hexanal can be accidental or consequence of experimental error.

These score values can be correlated with molecular structural descriptors. The best correlation was achieved by the boiling points:

$$S1_{(j)} = -9.795 + 0.02532 T_{B(j)};$$

$$R = 0.99759; F(1,33) = 6827.4;$$

$$p < 0.00000; s = 0.07040 \qquad (1)$$

where $S1$ is the first principal component (score 1) in case of (i), subscript "j" means the j-th alkylbenzene, R is the correlation coefficient, F is the Fisher statistic, p is the significance for the entire equation, and s is the standard error of estimate. The correlation is highly significant. However, a residual plot shows a curvature present in the retention data vs. boiling point relation as shown in Fig. 5. [Residuals are the difference between scores calculated

from measured Kováts indices and the predicted scores by Eq. (1)].

The Durbin–Watson test is suitable to detect trends or curvature in residuals. The value of test statistic is 1.089 whereas the critical value from the table is $d_L = 1.28$ ($N = 33$, $\alpha = 0.05$ and the number of independent variables is 1). It means that the null hypothesis of "no curvature in the residuals" should be rejected. The inherent non-linearity, which is not unique, was also observed for alkylbenzenes [39,40].

A similar correlation and figure pattern can be obtained in the case of (ii). The best correlation was achieved by the boiling points here, as well.

$$S1_{(j)} = -9.787 + 0.02530 T_{B(j)};$$

$$R = 0.99680; F(1,33) = 5124.6;$$

$$p < 0.00000; s = 0.08119 \qquad (2)$$

where $S1$ is the first principal component in case of (ii). Eq. (2) is identical to Eq. (1) within two decimal digits suggesting that the two approaches do not differ significantly. It is interesting to note that if
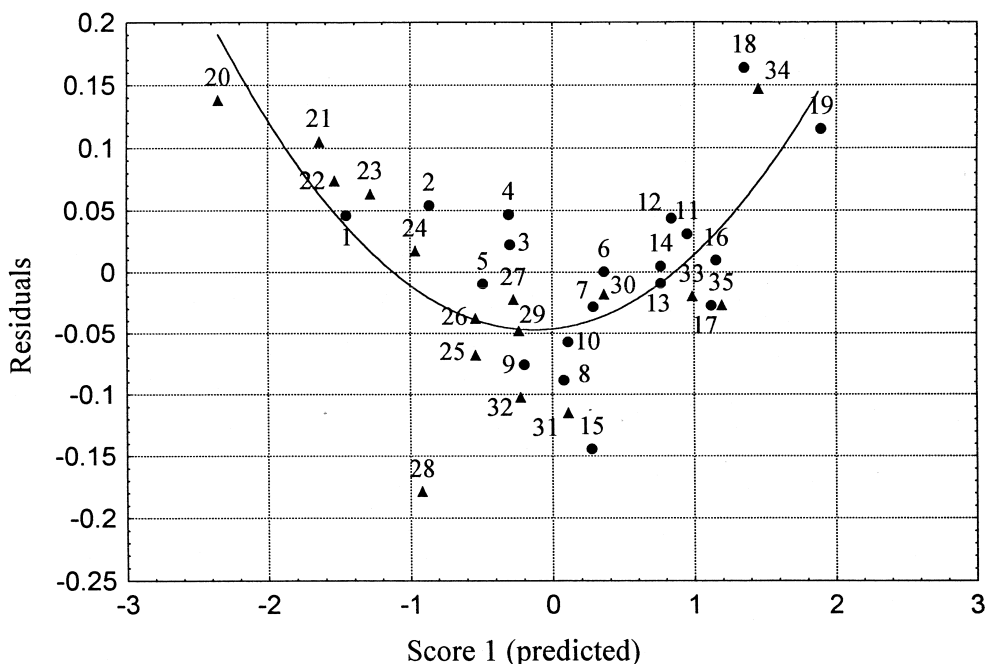
Fig. 5. Residual plot for linear boiling point dependence, case (i) and (ii) are identical, (●) ketones, (▲) aldehydes.

all variables are included into the model at once it results in a significant predictive equation.

$$S1_{(j)} = -5.208 + 0.01968T_{B(j)} - 0.04651M_{r(j)}$$
$$- 0.03046V_{m(j)} + 0.1043R_{m(j)}$$
$$+ 1.644 \log P_{(j)} + 0.9582 \, \text{Ind}_{(j)};$$

$R = 0.99901; F(6,28) = 2350.9;$

$p < 0.00000; s = 0.04905$ (3)

where Ind is an indicator variable, its value is 1.0 for ketones and 0.0 for aldehydes. The equation is highly significant again. To include any of the descriptor variables into the model is significant at the 0.05% or even less significance level. A somewhat smaller value of Fisher statistics is caused by the smaller degrees of freedom. It indicates the superiority of Eqs. (1) or (2) over Eq. (3). However, the residual error is much smaller in the latter case.

A similar equation can be derived for the second principal components:

$$S2_{(j)} = -24.30 + 0.4867M_{r(j)} - 0.03046V_{m(j)}$$
$$- 13.64 \log P_{(j)} - 8.500 \, \text{Ind}_{(j)};$$

$R = 0.82587; F(3,31) = 22.167;$

$p < 0.00000; s = 0.5905$ (4)

where $S2$ is the second principal component in case of (ii). The equation is highly significant although the description is much worse than for Eq. (3), as expected.

The physical significance of these equations is that knowing physical properties of an oxo compound the location of its score values can easily be calculated. Moreover, the temperature and polarity dependence of a "hypothetical" Kováts index can also be calculated from the score values estimated by Eqs. (3) and (4) without tedious experiments and even without having the incriminated compound as standard.

Principal component 1 (score 1) can be attributed to boiling point (or size of molecules) whereas score 2 contains information on the structural features of molecules, i.e., structure of apolar side chains,

position of branching to the carbonyl function. A triangular structure can be observed (Fig. 4). The outlying structures are positioned on the top, i.e., butenal and mostly aldehydes are located there. The aldehyde retention differs from that of ketones of the same size and similar structure.

5. Conclusions

The retention behavior of ketones and aldehydes is not distinguishable from each other completely although some characteristics of solute structure can be observed using PCA.

The polarity of columns is a more influential variable than their temperature.

Abstract retention data (free from influence of temperature and column polarity) are non-linear functions of boiling points of solutes.

Similarity among the solutes from the point of view of retention is represented by characteristic plots (Figs. 3 and 4).

Acknowledgements

References

[1] S. Wold, K. Andersson, J. Chromatogr. 80 (1973) 43–59.
[2] D.L. Massart, P. Lenders, M. Lauwereys, J. Chromatogr. Sci. 12 (1974) 617–625.
[3] S. Wold, J. Chromatogr. Sci. 13 (1975) 525–532.
[4] R. Fellous, L. Lizzani-Cuvelier, R. Luft, D. Lafaye de Micheaux, Anal. Chim. Acta 154 (1983) 191–201.
[5] J.F.K. Huber, G. Reich, J. Chromatogr. 294 (1984) 15–29.
[6] B. Walczak, M. Dreux, J.R. Chrétien, R. Szymoniak, M. Lafosse, L. Morin-Allory, J.P. Doucet, J. Chromatogr. 353 (1986) 109–121.
[7] J.A. Gracia-Dominguez, J. Garcia-Munoz, V. Menéndez, M.J. Molera, J.M. Santiuste, J. Chromatogr. 393 (1987) 209–219.
[8] M.H. Abraham, G.S. Whiting, R.M. Doherty, W.J. Shuley, J. Chromatogr. 518 (1990) 329–348.
[9] T.O. Kollie, C.F. Poole, J. Chromatogr. 550 (1990) 213–237.
[10] E.R. Malinowski, Factor Analysis in Chemistry, 2nd ed, Wiley-Interscience, New York, 1991, Ch. 9, pp. 266–291, and references cited therein.
[11] Z. Juvancz, T. Cserháti, K.E. Markides, J.S. Bradshaw, M.L. Lee, Chromatographia 38 (1994) 227–231.
[12] T. Cserháti, Anal. Lett. 27 (1994) 2615–2637.
[13] E. Forgács, Anal. Chim. Acta 296 (1994) 235–241.
[14] S.K. Poole, C.F. Poole, J. Chromatogr. A 697 (1995) 415–427.
[15] S.K. Poole, C.F. Poole, J. Chromatogr. A 697 (1995) 429–440.
[16] C.E. Reese, L. Huang, S.-H. Hsu, S. Tripathy, C.H. Lochmüller, J. Chromatogr. Sci. 34 (1996) 101–110.
[17] M. Turowski, R. Kaliszan, C. Lüllmann, H.G. Genieser, B. Jastorff, J. Chromatogr. A 728 (1996) 201–211.
[18] K. Héberger, Chemometrics Intell. Lab. Systems 47 (1998) 41–49.
[19] W.O. McReynolds, J. Chromatogr. Sci. 8 (1970) 685–691.
[20] R.V. Golovnya, D.N. Grigoryeva, A.V. Vasilyev, J. High Resolut. Chromatogr. 13 (1990) 47–51.
[21] J. Raymer, D. Wiesler, M. Novotny, J. Chromatogr. 325 (1982) 13–22.
[22] F. Saura-Calixto, A. Garcia-Raso, P.M. Déya, J. Chromatogr. Sci. 20 (1982) 7–10.
[23] F. Saura-Calixto, A. Garcia-Raso, J. Cañellas, J. Garcia-Raso, J. Chromatogr. Sci. 21 (1983) 267–271.
[24] L.S. Anker, P.C. Jurs, P.A. Edwards, Anal. Chem. 62 (1990) 2676–2684.
[25] V.P. Uralets, R.V. Golovnya, Zh. Anal. Khim. 33 (1978) 782–787.
[26] J.R. Donnelly, M.S. Abdel-Hamid, J.L. Jeter, D.F. Gurka, J. Chromatogr. 642 (1993) 409–415.
[27] M. Görgényi, H. Van Langenhove, Z. Király, J. Chromatogr. A 693 (1995) 181–185.
[28] M. Chastrette, P. Coillault, G. Casteignau, M. Mazet, Bull. Soc. Chim. France, (1973) 1605–1607.
[29] M. Gassiot-Matas, G. Firpo-Pamies, J. Chromatogr. 187 (1980) 1–19.
[30] F. Patte, M. Etcheto, P. Laffort, Anal. Chem. 54 (1982) 2239–2247.
[31] F.S. Calixto, A. Garcia Raso, Chromatographia 12 (1982) 771–775.
[32] H. Tamura, Sh. Kihara, H. Sugisawa, Agric. Biol. Chem. 54 (1990) 3171–3176.
[33] J.P. Vidal, S. Estreguil, R. Cantagrel, Chromatographia 36 (1993) 183–186.
[34] M. Görgényi, K. Héberger, J. Chromatogr. Sci. 37 (1999) 11–17.
[35] S. Wold, K. Esbensen, P. Geladi, Chemometrics Intell. Lab. Systems 2 (1987) 37–52.
[36] A. Lopata, A. Bencze DRUGIDEA version 3, release June 1 (1993), CheMicro Ltd., Budapest

[37] R.G. Weast (Ed.), Handbook of Chemistry and Physics, CRC Press, Boca Raton, FL, 62nd ed., 1984.

[38] C. Hansch, C. Leo, Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley, New York, 1979.

[39] K. Héberger, T. Kowalska, Chromatographia 48 (1997) 89–94.

[40] K. Héberger, T. Kowalska, Chemometrics Intell. Lab. Systems 47 (1998) 201–213.