

# Principal Component Analysis of Polarity and Interaction Parameters in Inverse Gas Chromatography\*

Károly Héberger<sup>1,†</sup>, Kasylda Milczewska<sup>2</sup>, and Adam Voelkel<sup>2,†</sup>

<sup>1</sup>Institute of Chemistry, Chemical Research Center, Hungarian Academy of Sciences H-1525 Budapest, P.O. Box 17, Hungary and <sup>2</sup>Poznań University of Technology, Institute of Chemical Technology and Engineering, pl. M. Skłodowskiej-Curie 2, 60-965 Poznań, Poland

## Abstract

Inverse gas chromatography is used in the characterization of aliphatic-aromatic and aromatic ketones, their oximes, and ketone-oxime or oxime-oxime mixtures. All these organic materials are used as liquid stationary phases in gas chromatographic columns. A series of polarity and Flory-Huggins interaction parameters are determined and used to describe the physicochemical properties of examined materials, metal extractants, and products of their degradation. Principal component analysis (PCA) is performed on a data matrix consisting of polarity and interaction parameters for ketones, their oximes, and mixtures. The calculations are carried out on the correlation matrix. It is found that seven principal components account for more than 95% of the total variance in the data, indicating that the polarity (interaction) parameters are not correlating well. Physical meanings are attributed to the principal components, the most influential ones being that the first and the second principal components account for several Flory-Huggins interaction parameters, whereas the fifth is correlated with criterion "A". The plots of component loadings show characteristic groupings of polarity indicators, whereas that of component scores show several groupings of stationary phases. Cluster analysis provides mainly the same groupings. PCA allows for the grouping of polarity and solubility parameters based on the information carried within those parameters. There is no need to use more than one parameter from each cluster. McReynolds polarity and the partial molar excess Gibbs free energy of solution per methylene group carry the same information. The groups of ketones, oximes, and their mixtures can be distinguished with the use of PCA on the basis of the measured polarity, solubility parameters, or both.

## Introduction

The term "polarity" is the most often used to characterize liquid stationary phases in gas chromatography. It may be described as the ability of the liquid phase to participate in all kinds of intermolecular interactions, including dispersive and specific (i.e., orientation, induction, and donor-acceptor) (1).

Inverse gas chromatography (IGC) has become a widely accepted, accurate, reliable, and relatively fast method for the representation of polymers and their blends, surface-active agents, metal extractants, silicas and modified silicas, and other minerals (2). IGC is an extension of conventional gas chromatography in which a nonvolatile material that is to be investigated is immobilized on a column. This stationary phase is then characterized by monitoring the passage of volatile probe molecules of known properties as they are carried through the column by an inert gas. The retention time and peak elution profile for standard solutes are influenced by the interactions between the solute and stationary phase. These retention parameters are used in the respective relations leading to the quantitative measures for physicochemical properties of examined material. The basic tools for IGC are inexpensive, widely available, and well-suited for routine laboratory applications. IGC data may be collected quite rapidly over extended temperature ranges. A variety of probes may be used in the mobile phase to elucidate the properties of the stationary phase, which otherwise can only be collected with a greater expenditure of time and money.

The application of principal component analysis (PCA) to chromatography has become a popular method in the last few decades (3-17), mainly because it can provide information not otherwise accessible. This information includes classification, determination of data structure, elimination of unimportant

\* Part XIII of the series: Quantitative Structure—Chromatographic Retention Relationship. For Part XII see T. Körtvélyesi, M. Görgényi, and K. Héberger. *Anal. Chim. Acta* **428**(1): 73-82 (2001).

† Authors to whom correspondence should be addressed: emails heberger@chemres.hu and Adam.Voelkel@put.poznan.pl.

variables, searching similarities, modeling, outlier detection, data reduction, and finding any physical significance of abstract factors.

PCA has been mainly applied to the classification and selection of proper stationary phases as well as the recommendation of preferred solvents to a given separation (11). The determination of factors important in the characterization of solvent properties has become the goal of recent investigations (17). Although the PCA results of different data sets have been frequently contradictory, there is no doubt that PCA is a proper tool for the representation of measures of polarity and solute-solvent interactions. Independent of the applied models, all investigations reported a dominant factor in the data, which was assigned the term "polarity".

To our knowledge, no PCA results have been published on the interaction parameters in IGC.

The work presented in this study can be summarized as the measurement of the polarity and interaction parameters of ketones, oximes, and their mixtures as well as the mathematical description of these parameters with the help of PCA. PCA was applied to examine the differences and similarities of both polarity parameters and examined stationary phases.

Our aim was to characterize the stationary phases (ketones, oximes, and their mixtures) and to find out which polarity parameters and stationary phases carry similar information (i.e., to determine which polarity measure is necessary and which one is superfluous for the proper characterization of stationary phases). Another aim was to examine whether it is possible to extract useful information for solute-solvent interactions. We have attempted to arrange in order the polarity measures according to their importance and, if possible, to attribute physical significance to principal components. Finally, we have attempted to classify ketones, oximes, and their mixtures from the point of view of polarity and solute-solvent interaction.

## Experimental

### Theory

#### *Polarity and interaction parameters determined*

The following parameters may be used for the characterization of the examined materials and stationary phases.

The most popular one is the Kováts retention index (RI), which describes the magnitude of solute-stationary phase interactions (for five or ten selected test solutes):

$$RI_x = 100 \left( z + \frac{\log t'_{R(x)} - \log t'_{R(z)}}{\log t'_{R(z+1)} - \log t'_{R(z)}} \right) \quad \text{Eq. 1}$$

where  $t'_{R(z+1)}$  and  $t'_{Rz}$  are the adjusted retention times of  $n$ -alkanes with  $z + 1$  and  $z$  carbon atoms, respectively.

The measure of stationary phase polarity is the difference:

$$\Delta RI = RI_p - RI_{np} \quad \text{Eq. 2}$$

where  $RI_p$  and  $RI_{np}$  denote the RI of the test solute on polar and nonpolar (squalane) stationary phases, respectively. The sum of

the RI differences for five test compounds is often called the McReynolds polarity (MP)—benzene (B),  $n$ -butanol, 2-pentanone, nitropropane, and pyridine (PY) are used as test compounds (18).

The polarity index (PI) is calculated from the equation:

$$PI = 100 \log(C - 4.7) + 60 \quad \text{Eq. 3}$$

where  $C$  is the number of carbon atoms in a hypothetical  $n$ -alkane having the same retention time as the test solute, either methanol (MEOL) or ethanol (ETOL) (19). The factor 4.7 was determined statistically and gave the least deviation from the regression line. The value of  $\log(C - 4.7)$  was multiplied by 100 in order to convert the PI to a whole number, and the value 60 was added to give the PI a positive value (19).

PI is an empirical parameter; in fact, it is strongly related to the RI of a polar test solute. It has often been used in the characterization of surface-active agents (2,19,20).

Coefficient  $\rho$  (the parameter very often used in the characterization of surfactants' polarity) (1,20) is calculated as the relative retention of a polar test solute (MEOL or ETOL) to a standard nonpolar solute, such as  $n$ -hexane.

$$\rho = \frac{t'_R(\text{EtOH})}{t'_R(n\text{-hexane})} \quad \text{Eq. 4}$$

Criterion "A" (CA) describes the magnitude of dispersive interactions between a methylene group and the stationary phase (21,22):

$$A = \frac{t'_{R(z+1)} - t'_{R(z)}}{t'_{R(z)} - t'_{R(z-1)}} \quad \text{Eq. 5}$$

where  $t'_{R(z+1)}$ ,  $t'_{Rz}$ , and  $t'_{R(z-1)}$  are the adjusted retention times of  $n$ -alkanes with  $z + 1$ ,  $z$ , and  $z - 1$  atoms of carbon, respectively, and coefficient  $A$  reflects the properties of the examined stationary phase. The values of CA are different, they are equal only if  $\Delta GCH_2$  is constant.

$\Delta G^E(CH_2)$  is the partial molar excess Gibbs free energy of solution per methylene group. It is a measure of the deviation of the methylene group solution in a given solvent from an ideal solution (23).  $\Delta G^E(CH_2)$  can be considered as a thermodynamic criterion for the polarity of liquid stationary phases.

A quantitative measure of the cohesive properties of a substance is the cohesive energy ( $\Delta E$ ), which when related to a unit of volume is called the  $\Delta E$  density (24). The square root of the  $\Delta E$  density is called the solubility parameter.

The solubility parameter is defined as:

$$\delta = \left( \frac{\Delta E}{V^\circ} \right)^{1/2} \quad \text{Eq. 6}$$

where  $V^\circ$  is the molar volume of the compound.

For volatile compounds the solubility parameter may be calculated from the following equation:

$$\delta = \left( \frac{\Delta H_v - RT}{V^\circ} \right)^{1/2} \quad \text{Eq. 7}$$

where  $\Delta H_v$  is the enthalpy of vaporization,  $R$  is the gas constant, and  $T$  is the absolute temperature.

For species with low volatility the use of equation 7 is not possible. Therefore, DiPaola-Baranyi and Guillet (25) proposed to calculate this parameter with the use of the solute-solvent interaction parameter ( $\chi_{1,2}^\infty$ ) given as:

$$\chi_{1,2}^\infty = \ln \left( \frac{273.15 R}{p_1^\circ V_g^\circ M_1} \right) - \frac{p_1^\circ}{RT} (B_{11} - V_1^\circ) + \ln \left( \frac{\rho_1}{\rho_2} \right) - \left( 1 - \frac{V_1^\circ}{V_2^\circ} \right) \quad \text{Eq. 8}$$

where  $M_1$  is the molecular mass;  $p_1^\circ$  is the saturated vapor pressure;  $B_{11}$  is the second virial coefficient;  $V_1^\circ$  is the molar volume;  $\rho_1$  is the density;  $V_g^\circ$  is the specific retention volume of the solute;  $\rho_2$  and  $V_2^\circ$  are the density and molar volume of the stationary phase, respectively; and  $T$  is the column temperature. Subscript "1" refers to the "test solute" used in the IGC experiment, whereas subscript "2" refers to the examined material.

Assuming that  $\chi_{1,2}^\infty$  has a free energy characteristic with additive enthalpy ( $\chi_H^\infty$ ) and entropy terms ( $\chi_S^\infty$ ):

$$\chi_{1,2}^\infty = \chi_H^\infty + \chi_S^\infty \quad \text{Eq. 9}$$

DiPaola-Baranyi and Guillet (25) expressed  $\chi_{1,2}^\infty$  in the form:

$$\chi_{1,2}^\infty = \frac{V_1^\circ (\delta_1 - \delta_2)^2}{RT} + \chi_S^\infty \quad \text{Eq. 10}$$

where  $\delta_1$  and  $\delta_2$  are solubility parameters of the test solute and the examined material, respectively. Equation 10 could be rearranged to form the following equation, which allows for the calculation of  $\delta_2$ :

$$\frac{\delta_1^2}{RT} - \frac{\chi_{1,2}^\infty}{V_1^\circ} = \frac{2\delta_2}{RT} \delta_1 - \left( \frac{\delta_2^2}{RT} + \frac{\chi_S^\infty}{V_1^\circ} \right) \quad \text{Eq. 11}$$

For the series of test solutes (having different  $\delta_1$  values), retention data are collected and the respective values of the solute-solvent interaction parameters are calculated (using equation 8). By having a set of  $\chi_{(1,2)i}^\infty$  and  $\delta_{1i}$  values for the respective test solutes, it is possible to calculate the slope of the following linear relationship:

$$\frac{\delta_{1i}^2}{RT} - \frac{\chi_{(1,2)i}^\infty}{V_1^\circ} = \frac{2\delta_2}{RT} \delta_{1i} - \left( \frac{\delta_2^2}{RT} + \frac{\chi_S^\infty}{V_1^\circ} \right) \quad \text{Eq. 12}$$

The slope is proportional to  $\delta_2$  of the examined material. Using different types of probe solutes, it is possible to obtain the dispersive and specific components of the solubility parameter (26,27). Hansen (26) and later Voelkel and Janas (27,28) separated the contributions of the solubility parameters attributed to dispersive ( $\delta_d$ ), inductive ( $\delta_{in}$ ), and hydrogen-bonding ( $\delta_h$ ) solute-solvent interactions.

The total (corrected) solubility parameter ( $\delta_T$ ) was then calculated from the equation:

$$\delta_T^2 = \delta_d^2 + \delta_{in}^2 + \delta_h^2 \quad \text{Eq. 13}$$

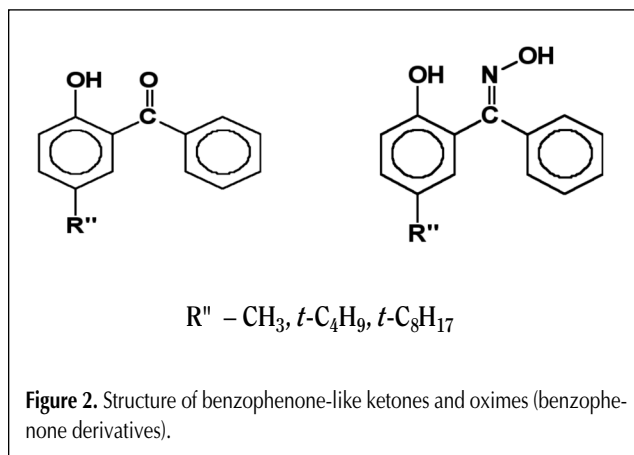
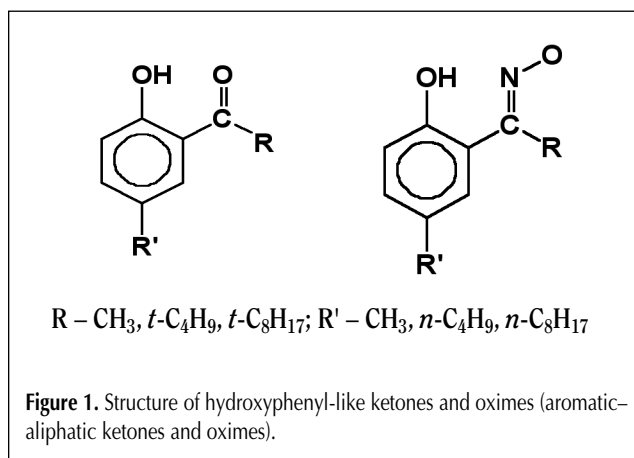
This series of polarity parameters were used to distinguish stationary phases (metal extractants) and their mixtures.

#### PCA

PCA can be considered as a projection method; a multidimensional problem can be represented in smaller dimensions. In the course of defining principal components, the original variables are transformed into new ones. Original variables (columns in the input matrix) are presented as linear combinations of principal components. The values of principal components are called component scores. The linear coefficients are called the component loadings (i.e., the correlation coefficients between the original variables and the principal components).

The columns of data matrices under study were intercorrelated (i.e., the data were redundant). The method of PCA eliminated the redundancy from the data (starting from the correlation matrix of the variables); in other words, it reduced their dimensionality by revealing several underlying components. These underlying components are called principal components. PCA was able to decompose the original input matrix into a sum of multiplication products of loading and score vectors.

The principal components were orthogonal (independent) and were scaled so that their variances were equal to unity. Later on in the study, they were ordered in such a way that the



variance explained by the first principal component was the greatest, the variance explained by the second one was smaller, and so on until that of the last one was the smallest.

The algorithms for PCA can be found in standard chemometric books and tutorials (11,30,31). A basic assumption used in PCA is that the score and loading vectors corresponding to the largest eigenvalues contain the most useful information relating to a specific problem. The remaining ones are comprised of mainly the noise; that is, for a practical problem it is sufficient to retain only a few components accounting for a large percentage of the total variance (30).

Euclidean distance was used in the case of cluster analysis using Ward's method. The classification by cluster analysis is always subordinate to PCA, because the former applies distance for classification, whereas the latter applies distance as well as direction.

### Materials

Copper extractants from aromatic and aliphatic hydroxy oximes and their photodegradation products (in this study ketones) as well as their mixtures were used as liquid stationary phases in the IGC experiments. Figures 1 and 2 present the base structure of the used stationary phases.

The abbreviations for the active substances (stationary phases) and supports are summarized in Table I.

All ketones and oximes were synthesized by E. Krzyżanowska at Poznań University of Technology according to the methods described previously (32,33).

### IGC experiments

The conditions of the IGC experiments were as follows. The column used had a 1-m × 3-mm i.d. and a column temperature of 130°C. The column loading was comprised of 14% (w/w) of the liquid phase on silanized DMCS Chromosorb W (w) or silanized DMCS Chromosorb P (p) 60-80 mesh. The carrier gas was helium at a flow rate of 20 mL/min. A Chrom 5 (Kovo, Czech Republic) gas chromatograph was used with a flame ionization detector. The column loading was verified by Soxhlet extraction. The loss of the liquid stationary phase (measured after the IGC experiment) was lower than 0.1% of the initial mass of liquid phase.

We used 16 different, carefully selected volatile test compounds. The injection volume was 0.5–1 µL.

### Calculation of polarity and interaction parameters

The void volume was calculated

according to the procedure proposed by Grobler-Bálizs (29). Polarity parameters were set by calculating PI and p with the use of the retention data of MEOL and ETOL as polar probes. Thermodynamic functions of solution (i.e.,  $\Delta G^E(\text{CH}_2)$ ) were calculated from the following equation (23) for two members of the homologous series having z and z + k methylene groups, respectively:

$$\Delta G^E(\text{CH}_2) = \frac{1}{k} RT \ln \frac{(V_g^{\circ} p^{\circ})_z}{(V_g^{\circ} p^{\circ})_{z+k}} \quad \text{Eq. 14}$$

The solubility parameter was calculated by the procedure given by DiPaola-Baranyi and Guillet using equation 11. Increments of the solubility parameters were calculated as described previously (27,28).

The adsorption of the test solutes during IGC experiments could influence the values of the determined parameters. However, in our case we have observed no significant adsorption effects. This was probably because of the relatively high content of the examined substance (in this study liquid stationary phase).

**Table I. Abbreviations Used for Surface-Active Agents and for Column Fillings (Support)**

Abbreviation	Surface-active agents (stationary phase)	Support Chromosorb	Structure
w1k	1-(2'-hydroxy-5'-methylphenyl)-butan-1-one	w	
w1o	1-(2'-hydroxy-5'-methylphenyl)-butan-1-one oxime	w	
w1k1o	1:1 mixture of w1k and w1o (w/w)	w	
p1o	1-(2'-hydroxy-5'-methylphenyl)-butan-1-one oxime	p	
w2k	(2'-hydroxy-5'-t-octyl)-benzophenone	w	Figure 2
w2o	(2'-hydroxy-5'-t-octyl)-benzophenone oxime	w	Figure 2
w2k2o	1:1 mixture of w2k and w2o (w/w)	w	Figure 2
w3k	(2'-hydroxy-5'-t-butyl)-benzophenone	w	Figure 2
w3o	(2'-hydroxy-5'-t-butyl)-benzophenone oxime	w	Figure 2
w3k3o	1:1 mixture of w3k and w3o (w/w)	w	Figure 2
w4k	1-(2'-hydroxy-5'-methylphenyl)-octan-1-one	w	
w4o	1-(2'-hydroxy-5'-methylphenyl)-octan-1-one oxime	w	
w4k4o	1:1 mixture of w4k and w4o (w/w)	w	
p4o	1-(2'-hydroxy-5'-methylphenyl)-octan-1-one	p	
p5o	(2'-hydroxy-5'-methyl)-acetophenone oxime	p	Figure 1
p6o	(2'-hydroxy-5'-t-butyl)-acetophenone oxime	p	Figure 1
p7o	(2'-hydroxy-5'-methyl)-benzophenone oxime	p	Figure 2
p7o1o	1:1 mixture of p7o and p1o (w/w)	p	Figure 2 and Figure 1
p7o2o	1:1 mixture of p7o and w2o (w/w)	p	Figure 2 and Figure 2
p7o5o	1:1 mixture of p7o and p5o (w/w)	p	Figure 2 and Figure 1
p5o6o	1:1 mixture of p5o and p6o (w/w)	p	Figure 1 and Figure 1

**Data pretreatment**

All polarity and interaction parameters corresponded to 403 K.

The stationary phases (ketones, oximes, and mixtures) were ordered in rows in the input matrix: (w1k, w1o, w1k1o, w2k, w2o, w2k2o, w3k, w3o, w3k3o, w4k, w4o, w4k4o, p5o, p6o, p5o6o, p1o, p7o, p7o1o, p7o5o, p7o2o, and p4o; altogether 21

stationary phases, or rows, in the input matrix).

The following polarity parameters were ordered in the columns of the input matrix: CA, coefficient  $\rho^{\text{MeOH}}$  (CRM), coefficient  $\rho^{\text{EtOH}}$  (CRE),  $\text{PI}^{\text{MeOH}}$  (PIM),  $\text{PI}^{\text{EtOH}}$  (PIE),  $\delta_2$ ,  $\delta_d$ ,  $\delta_{\text{in}}$ ,  $\delta_h$ ,  $\delta_T$ ,  $\Delta G^{\text{E}}(\text{CH}_2)$ , MP,  $\chi_{1,2}^{\infty}$  for *n*-hexane (HX),  $\chi_{1,2}^{\infty}$  for *n*-pentane (P),  $\chi_{1,2}^{\infty}$  for *n*-heptane (HP),  $\chi_{1,2}^{\infty}$  for *n*-octane (O),  $\chi_{1,2}^{\infty}$  for *n*-nonane (N),  $\chi_{1,2}^{\infty}$  for decane (D), B, MEOL, ETOL, butan-1-ol

**Table II. Polarity Parameters Determined**

	CA	CRM	CRE	PIM	PIE	$\delta_2$ ( $10^3$ (J/m <sup>3</sup> ) <sup>1/2</sup> )	$\delta_d$ ( $10^3$ (J/m <sup>3</sup> ) <sup>1/2</sup> )	$\delta_{\text{in}}$ ( $10^3$ (J/m <sup>3</sup> ) <sup>1/2</sup> )	$\delta_h$ ( $10^3$ (J/m <sup>3</sup> ) <sup>1/2</sup> )	$\delta_T$ ( $10^3$ (J/m <sup>3</sup> ) <sup>1/2</sup> )	$\Delta G^{\text{E}}(\text{CH}_2)$	MP
w1k	0.00	0.98	0.99	50.60	64.30	16.15	4.09	7.74	17.74	19.78	2403	2452
w1o	1.07	1.02	1.04	79.00	82.70	16.38	4.82	7.41	17.42	19.53	2164	2213
w1k1o	1.07	0.88	0.90	39.20	47.50	16.06	4.44	7.20	17.56	19.48	2456	2505
w2k	1.83	0.60	0.62	-181.70	-17.50	17.07	10.82	8.98	11.23	18.00	623	672
w2o	1.97	1.28	1.61	87.00	100.00	18.35	9.64	10.19	12.03	18.48	1247	1298
w2k2o	1.70	1.12	1.29	47.70	68.50	17.91	4.98	17.35	24.61	30.52	1623	1701
w3k	2.58	1.06	1.22	80.39	82.58	17.42	15.51	4.26	5.31	16.94	-454	-412
w3o	1.59	1.28	1.63	92.83	106.29	17.61	8.28	11.06	12.40	18.57	1697	1748
w3k3o	1.44	1.18	1.26	80.26	87.37	17.40	7.51	10.12	14.31	19.07	2029	2080
w4k	3.02	1.03	1.05	89.20	90.90	16.33	4.86	7.01	17.21	19.21	2217	2265
w4o	1.90	0.94	1.08	66.50	84.30	17.09	9.36	7.76	11.53	16.75	1942	1987
w4k4o	1.67	0.43	0.26	52.10	78.70	17.68	10.64	3.06	9.65	14.69	810	848
p5o	5.31	1.05	1.09	95.00	99.50	19.12	7.54	6.28	20.19	22.45	2390	2447
p6o	1.42	1.18	1.22	58.60	67.60	16.33	6.67	8.60	13.82	17.59	2247	2294
p5o6o	1.49	1.14	1.20	100.40	107.30	16.18	5.80	9.27	15.03	18.59	2111	2160
p1o	1.25	1.10	1.12	68.80	72.30	15.90	7.47	6.06	12.82	16.03	2117	2159
p7o	2.91	0.92	0.86	47.40	26.30	16.73	4.74	11.01	-8.54	14.72	2558	2579
p7o1o	0.85	1.27	1.22	180.50	147.90	17.53	3.72	14.00	4.58	15.19	2101	2139
p7o5o	0.54	1.16	1.14	117.80	99.10	15.76	3.95	10.28	4.75	11.99	2069	2100
p7o2o	2.20	1.00	1.09	92.50	126.20	17.82	5.16	11.52	4.36	13.35	2486	2520
p4o	2.84	0.75	1.34	93.00	107.00	17.96	13.54	7.78	4.28	16.19	3300	3342

**Table III. Flory–Huggins Interaction Parameters Determined**

	P	HX	HP	O	N	D	B	MEOL	ETOL	BUOL	BUON	PEON	NPR	PY	TOL	EB
w1k	0.82	1.46	2.11	2.79	3.45	4.12	1.56	1.23	1.34	2.36	1.53	1.97	2.85	2.32	2.25	2.79
w1o	0.70	1.29	1.87	2.49	3.07	3.70	1.41	1.07	1.16	2.16	1.33	1.72	2.50	2.13	1.95	2.54
w1k1o	0.67	0.99	1.52	2.03	2.55	3.11	1.16	1.36	1.33	2.05	1.20	1.49	2.26	1.95	1.73	2.16
w2k	0.60	-0.90	-0.82	-0.85	-0.79	5.91	4.14	0.15	-2.56	2.97	0.69	1.13	0.75	0.74	0.84	0.99
w2o	0.46	0.90	1.17	1.35	1.51	1.59	0.17	0.34	0.21	0.33	-0.17	-0.14	0.75	-1.07	0.36	0.49
w2k2o	0.65	0.88	1.16	1.38	1.52	1.59	0.37	0.74	0.63	0.77	0.26	0.26	0.93	-0.61	0.61	0.74
w3k	0.87	1.05	1.26	1.37	1.46	1.56	-0.25	1.81	1.62	1.46	0.27	0.23	0.67	0.27	0.27	0.34
w3o	0.77	1.19	1.60	2.00	2.25	2.50	0.94	0.78	0.65	0.91	0.41	0.49	1.40	-0.49	1.08	1.27
w3k3o	1.01	1.09	1.49	1.90	2.19	2.41	0.79	0.98	0.92	1.26	0.64	0.78	1.45	-0.02	1.14	1.36
w4k	0.56	1.11	1.69	2.35	2.88	3.42	1.24	1.04	1.11	2.04	1.18	1.57	2.49	2.07	1.88	2.40
w4o	0.78	1.02	1.51	1.73	1.94	2.08	0.72	0.86	0.83	1.16	0.56	0.67	1.45	0.12	1.00	1.20
w4k4o	0.22	0.26	0.30	0.33	0.12	0.22	-0.20	0.58	0.32	-0.11	-0.45	-0.61	0.16	1.19	-0.13	-0.15
p5o	1.53	1.30	1.98	2.52	3.14	3.54	1.36	0.12	0.40	2.03	2.17	0.70	1.61	3.42	2.49	2.07
p6o	1.34	1.86	2.46	2.98	3.46	3.93	1.46	1.39	1.47	2.18	1.44	1.77	2.37	1.44	1.98	1.48
p5o6o	1.85	1.48	1.89	2.40	2.83	3.40	1.31	1.40	1.53	2.29	1.23	1.68	2.45	1.80	1.79	1.48
p1o	0.45	1.87	2.43	2.96	3.39	3.75	1.75	1.52	1.61	2.37	1.68	3.11	2.63	2.00	2.25	2.57
p7o	-5.71	0.62	1.31	1.93	2.39	6.85	0.81	-0.07	0.23	1.47	0.67	0.22	1.57	3.83	1.20	1.44
p7o1o	-1.28	0.79	1.24	1.82	2.26	6.64	0.07	0.02	0.02	0.91	0.22	0.49	1.10	0.62	0.65	1.48
p7o5o	-0.91	1.25	1.68	2.18	2.59	7.22	1.25	1.15	1.13	1.82	1.25	1.67	2.07	1.59	1.55	2.01
p7o2o	-1.64	0.24	0.78	1.25	1.59	5.93	0.19	0.18	0.04	0.81	0.24	0.63	0.92	0.52	0.59	0.93
p4o	1.07	-0.38	-0.26	-0.18	-1.27	5.64	2.80	-0.68	-3.95	2.30	0.26	0.04	1.09	1.19	1.03	1.48

(BUOL), butan-2-one (BUON), pentan-2-one (PEON), 1-nitropropane (NPR), PY, toluene (TOL), and ethyl benzene (EB). Altogether there were 28 polarity and interaction parameters, or columns, in the input matrix.

First, the data were mean-centered (column means subtracted from each matrix element), then each matrix element was divided by the standard deviation of the respective column. In this way, the correlation matrix was established.

## Results and Discussion

The values of polarity and  $\chi_{1,2}^{\infty}$  are presented in Tables II and III.

The first problem in the examination of these data is the determination of factors required for the proper characterization of the so-called "polarity". It is worthwhile to consider how many factors should be retained in the model. The answer could be found in the determination of eigenvalues. The scree plot (Figure 3) suggests using seven principal components.

A criterion proposed by Kaiser (34) (i.e., an average eigen-

value) accepts all components having eigenvalues higher than the average and withdraws all with lower eigenvalues. The average eigenvalue equals one in the PCA practice if analyzing correlation matrices. Therefore, only components having eigenvalues larger than one should be used for further analysis. In our case, seven components had their eigenvalues larger than one. This means that at least seven underlying components (factors) were required to represent the variability in the data. Seven principal components explained more than 95% of the total variance in the data. The 21-dimensional problem cannot be simplified correctly going below seven dimensions. This means that the polarity indicators were not highly correlated, they were "balanced" (i.e., they carried more or less independent information). The characteristic pattern can be seen in Table IV.

It is difficult to attribute physical meaning to the abstract factors; nevertheless, it can be established that (a) the first principal component was well-correlated (the highest correlation coefficients) with HX, HP, O, N, BUON, PEON, NPR, TOL, and EB; (b) the second principal component was well-correlated with  $\chi_{1,2}^{\infty}$  for two other test solutes (B and BUOL); and (c) the fifth principal component correlated at best with CA.

**Table IV. Factors Loadings: Seven Principal Components Retained in the Model\***

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
CA	-0.18208	0.13889	-0.08143	0.41591	0.74198	-0.15523	-0.37309
CPM	0.49952	-0.61892	-0.28209	0.18970	-0.25880	0.16739	-0.37732
CPE	0.20039	-0.51614	-0.29905	0.47514	-0.24193	0.39113	-0.32656
PIM	0.31553	-0.61990	-0.53533	-0.01910	0.27993	0.24586	0.18580
PIE	0.09268	-0.60322	-0.49432	0.16117	0.19038	0.39387	0.26110
$\delta_2$	-0.62162	-0.29120	-0.18285	0.52068	0.35464	-0.12092	-0.14662
$\delta_d$	-0.64582	0.00937	0.39820	0.11591	0.28406	0.48450	-0.18162
$\delta_{in}$	-0.04451	-0.23915	-0.52558	0.25263	-0.65227	-0.34692	-0.16628
$\delta_h$	0.36710	-0.17937	0.54219	0.60504	-0.10152	-0.22786	0.28382
$\delta_T$	0.11649	-0.19962	0.33206	0.69007	-0.14409	-0.51885	-0.04262
$\Delta G^{\ddagger}(\text{CH}_2)$	0.48108	0.30512	-0.65720	0.27618	0.06000	-0.03149	0.32125
MP	0.48287	0.30210	-0.65216	0.28681	0.05707	-0.03544	0.32271
P	0.14016	-0.14874	0.61883	0.57286	0.00308	0.36692	0.27507
HX	<b>0.83338</b>	-0.48421	0.12872	-0.06590	0.10880	0.03219	-0.06677
HP	<b>0.88617</b>	-0.39839	0.03164	-0.06813	0.13856	-0.03271	-0.09324
O	<b>0.91843</b>	-0.32966	-0.05681	-0.08349	0.12850	-0.06758	-0.08121
N	<b>0.90724</b>	-0.29829	-0.02381	-0.12958	0.08673	-0.16452	-0.12768
D	0.14498	0.57372	-0.64714	-0.18166	-0.22059	0.12095	-0.23713
B	0.07772	<b>0.85550</b>	0.15856	0.32370	-0.23323	0.20958	-0.09235
MEOL	0.57303	-0.32296	0.62943	-0.30545	-0.12682	0.12294	0.00516
ETOL	0.65552	-0.55551	0.28186	-0.35775	0.06627	-0.18992	-0.03386
BUOL	0.49509	<b>0.72675</b>	0.21253	0.17642	-0.08504	0.24391	-0.20098
BUON	<b>0.85052</b>	0.35720	0.13783	0.16895	0.17901	-0.01236	-0.16523
PEON	<b>0.83522</b>	0.30454	0.21698	-0.08018	-0.23224	0.18014	0.00391
NPR	<b>0.95160</b>	0.21898	0.03852	-0.00904	-0.05868	0.06823	0.08102
PY	0.50359	0.57490	-0.13480	-0.20984	0.50347	-0.22429	-0.05133
TOL	<b>0.89058</b>	0.35081	0.03775	0.22285	0.15192	0.01757	-0.04907
EB	<b>0.84846</b>	0.37984	-0.12267	0.12310	-0.00123	0.05688	0.06821
Explained variance	10.139	5.2965	3.9075	2.7052	2.0123	1.5425	1.1244
Proportion of total variance	0.3621	0.1892	0.1396	0.09662	0.07187	0.05509	0.04016

\* Loadings (correlation coefficients between the old variables and new principle components) that are higher than 0.70000 are indicated in bold.

The plot of loadings allowed for the establishing of several clusters. These were as follows: (a)  $\Delta G^E(\text{CH}_2)$  and MP (cluster A in Figures 4A, 4B, and 4C); (b)  $\chi_{1,2}^\infty$  for *n*-alkanes (cluster B in Figures 4A, 4B, and 4C) and polar test solutes, particularly BUON, PEON, NPR, TOL, and EB (cluster C in Figures 4A, 4B, and 4C); (c) the "broad" cluster containing PIs and  $\rho$  values for both MEOL and ETOL (cluster D in Figure 4A). Similar to this, the close resemblance between  $\delta_T$  and P could only be seen in Figure 4A (cluster E).

The close agreement of some polarity parameters means that there is no need to use all of them. One can choose one parameter from each cluster. MP and  $\Delta G^E(\text{CH}_2)$  carry the same information, thus one of them is superfluous. Similarly, some unimportant polarity (or interaction) parameters can be selected out. For example, it is satisfactory to determine  $\chi_{1,2}^\infty$  for one *n*-alkane (cluster B) and PEON (cluster C) only. Furthermore, the first principal component was well-correlated with all the parameters from clusters B and C. Because of this, it is no surprise that the two clusters merged into each other in higher dimensions. Probably, any parameter from one cluster would properly represent the properties of the examined materials.

The loading plots show clearly that all of the solubility parameters were scattered despite the correlation of factors. The information carried by  $\delta_2$ ,  $\delta_T$ , and its increments corresponding to  $\delta_d$ ,  $\delta_{in}$ , and  $\delta_h$  interactions was different. Therefore, the determination of these parameters and their use in the characterization of liquid materials is highly recommended. This is specially justified when only solubility parameters are used for characterization. Furthermore, none of the solubility parameters belonged to one of the previously indicated clusters. This means that even when more polarity (or interaction) parameters are applied, the use of the solubility parameters should also be taken into account. The same formulation is valid for CA.

It should be emphasized that the close similarity between  $\Delta G^E(\text{CH}_2)$  and MP was an unexpected result. Only *n*-alkanes (or two other compounds from the homologous series, particularly 1-alkanols or 2-alkanones) are needed to determine  $\Delta G^E(\text{CH}_2)$ , whereas five different solutes of various polarities are necessary to determine MP. Interestingly, the points belonging to D, BUOL, and B were outside the corresponding cluster. Even if the physical reason for clustering is not understood completely, it is an experimental empirical fact.

The grouping of the examined stationary phases may be carried out by plotting the factor scores (Figures 5A and 5B).

Several clusters can be observed grouping stationary phases of similar properties. Upon further notice of Figure 5A, it can be seen that aromatic-aliphatic ketones, their oximes, and mixtures of these ketones and oximes are grouped mainly in cluster I. These compounds contain rather short alkyl chains (R,

R' = CH<sub>3</sub>, *n*-C<sub>4</sub>H<sub>9</sub>, or both). Therefore, it may be described as the methyl-butyl cluster.

The only exception was the point corresponding to the mixture of benzophenone (7) and acetophenone (5) derivatives (p7o5o). However, this point was separated from the center of cluster I in Figure 5A. All benzophenone derivatives belonged to cluster II in both cases (i.e., in Figures 5A and 5B). These compounds have methyl, *t*-butyl, and *t*-octyl substituents as well as the p7o1o mixture.

However, the presence of the separated cluster III (Figure 5B) formed by the points corresponding to stationary phase 7o and mixtures containing this oxime suggests several different properties of p7o from other benzophenone derivatives.

The outlying points are w4k4o, w2k, p4o, and p7o in Figure 5A and w4k4o, w2k, and w3k in Figure 5B.

The presence of the point corresponding to the w4o stationary phase in cluster II was unexpected. All other points representing w4k and the mixtures w4k4o and p4o (this oxime on another support) were outside this cluster. The point for w4k was always in cluster I, and points w4k4o and p4o were randomly distributed outside any cluster. The possible reason could be the large R, which significantly increases the hydrophobicity of this compound. Other similarly separated points in Figure 5A were located closely to each other in Figure 5B (e.g., the points corresponding with w2k and w3k stationary phases such as ketones with larger R'' substituents, particularly *t*-octyl and *t*-butyl).

Cluster analysis allowed for the separation of aliphatic-aromatic ketones and their oximes from aromatic ones (i.e., from benzophenone derivatives). The position of the points corresponding with the mixtures of aromatic and aliphatic-aromatic oximes (p7o5o and p7o1o) depended on the structure of the aliphatic-aromatic ketone. If the acetophenone derivative (5o) was present in the mixture, then the corresponding point was placed in cluster I. The increase of R causes an increase of the hydrophobicity of the compound, thus the corresponding point moved to cluster II (Figure 5A) or to the new cluster III in Figure 5B.

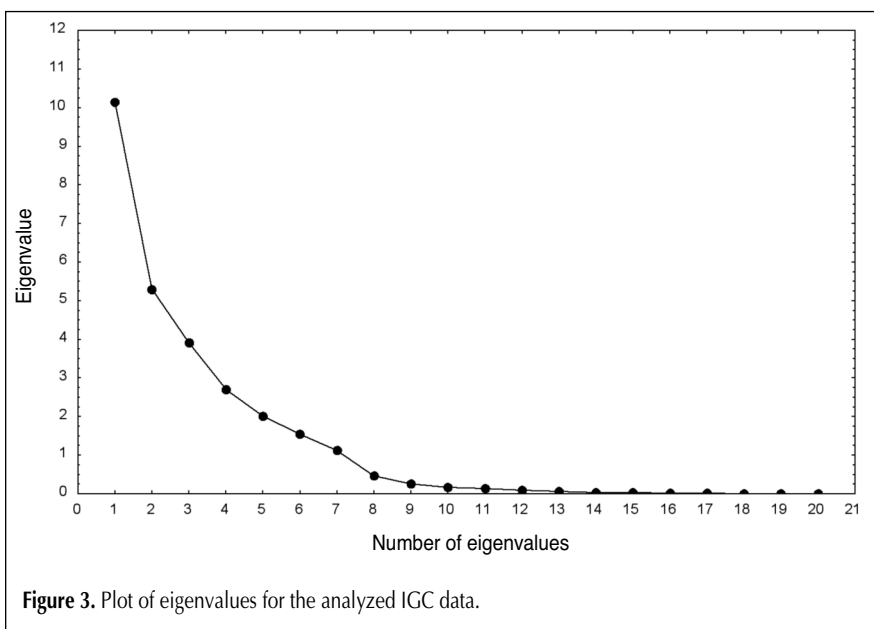
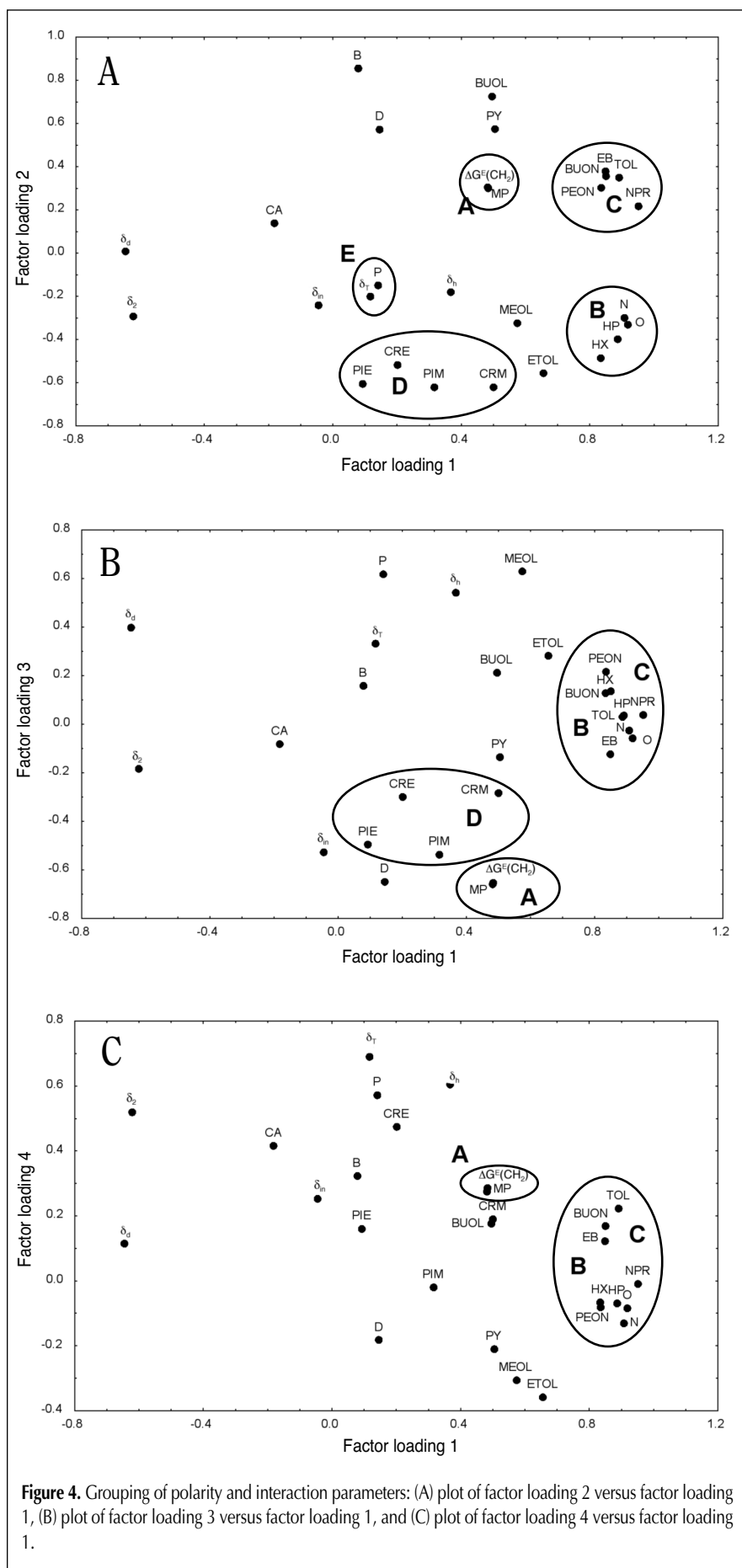


Figure 3. Plot of eigenvalues for the analyzed IGC data.



Only a slightly different classification could be derived from the dendrogram presented in Figure 6.

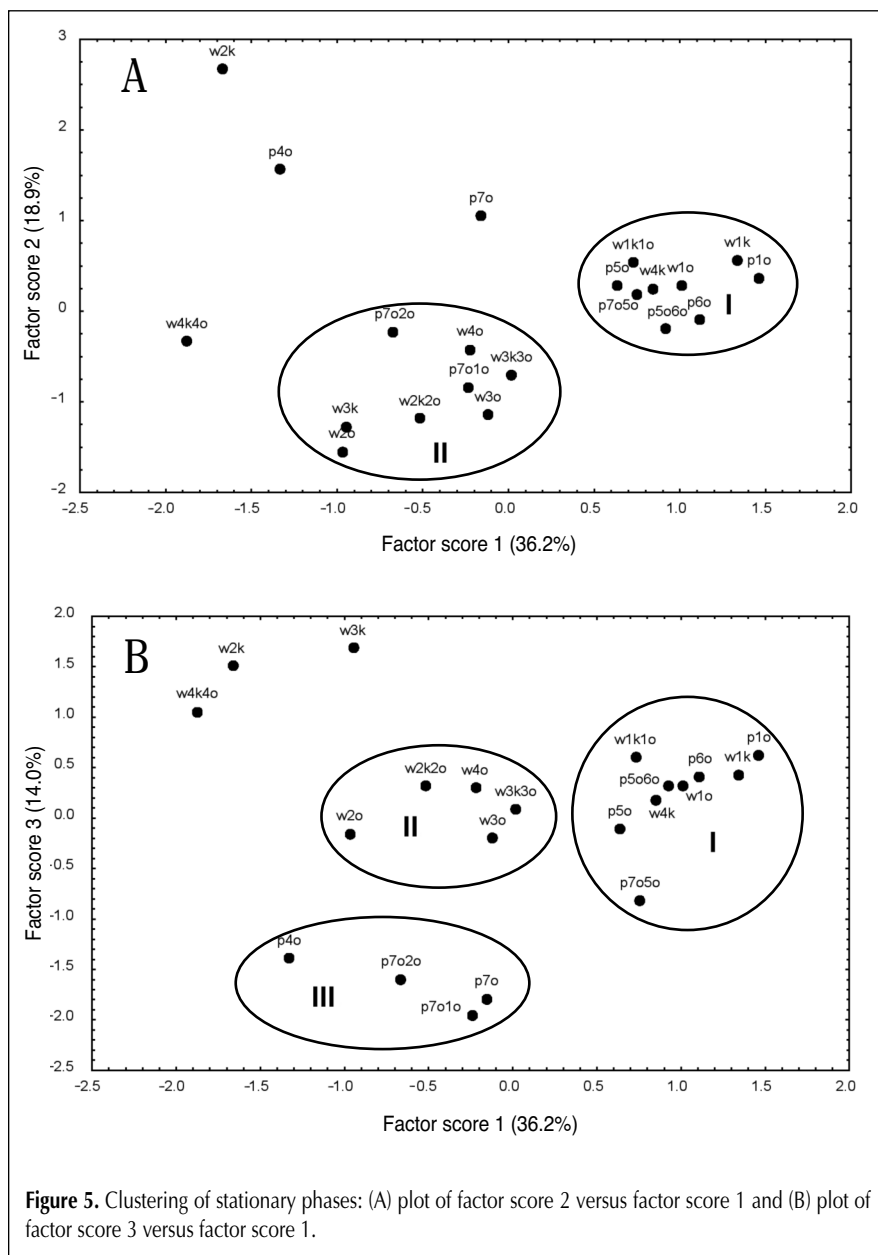
A careful inspection of the dendrogram (Ward method) in Figure 6 suggests the presence of the following three clusters: (a) w1k, w1o, w1k1o, w4k, p6o, p5o6o, p1o, p7o5o, and p5o (aliphatic–aromatic ketones, oximes, and mixtures having longer Rs such as *t*-butyl and *t*-octyl, oximes of acetophenone derivatives, mixtures of oximes of benzophenone and acetophenone derivatives, and the p1o phase all belong to this group); (b) w2k and p4o (these can be considered as outliers of groups 1 and 3); and (c) w2o, w3o, w3k3o, w4o, w2k2o, p7o, p7o1o, p7o2o, w3k, and w4k4o. This last group contained oximes of benzophenone derivatives having different R's, p7o2o, and p7o1o as well as benzophenone derivatives (ketones) having a long R' alkyl group (*t*-octyl or *t*-butyl), w2k2o and w4k4o, and the two oximes 7o and 5o.

## Conclusion

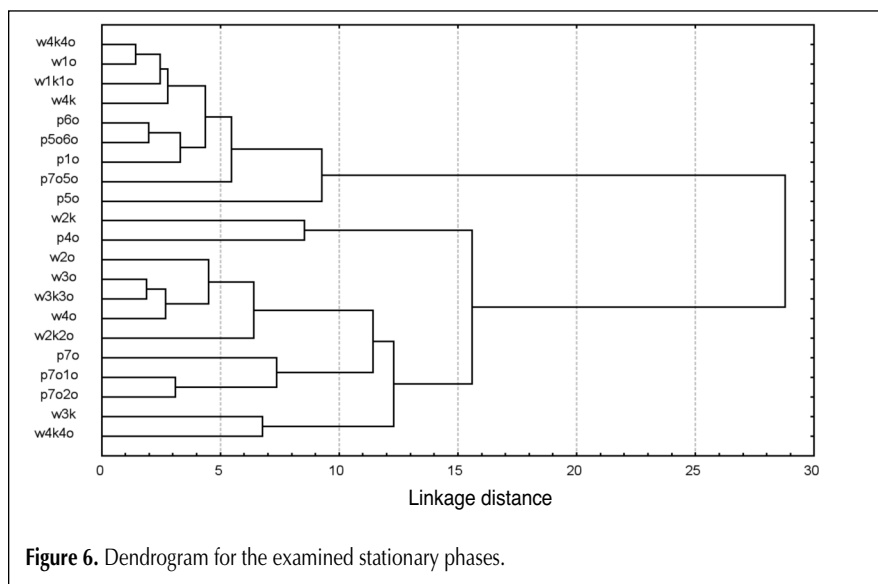
PCA allows for the grouping of polarity and solubility parameters as well as stationary phases. There is no need to use more than one parameter from each cluster. This finding diminishes the number of polarity and interaction parameters considerably. MP and  $\Delta G^E(\text{CH}_2)$  carry the same information. Only one of them is needed for the proper characterization of stationary phases; however, the latter is preferred because it is thermodynamically based. Solubility parameters reveal different aspects of the solution than other interaction parameters. Therefore, the determination of these parameters and their use in the characterization of liquid materials is highly recommended. Because none of the solubility parameters belong to any of the previously indicated clusters, their use should be taken into account even when other polarity (or interaction) parameters are applied.

The groups of ketones, oximes, and their mixtures could be distinguished (different structures influencing their properties) with the use of PCA on the basis of the measured polarity, solubility parameters, or both.





**Figure 5.** Clustering of stationary phases: (A) plot of factor score 2 versus factor score 1 and (B) plot of factor score 3 versus factor score 1.



**Figure 6.** Dendrogram for the examined stationary phases.

## Acknowledgments

This work was supported by the Hungarian Research Foundation (OTKA T 029748) and by the PUT DS 32/045/2001 Grant.

## References

- C.F. Poole and S.A. Schuette. *Contemporary Practice of Chromatography*. Elsevier, Amsterdam, The Netherlands, 1984, p 31.
- A. Voelkel. Inverse gas chromatography: characterization of polymers, fibers, modified silicas and surfactants. *Crit. Rev. Anal. Chem.* **22**: 411–39 (1991).
- D.L. Massart, P. Lenders, and M. Lauwereys. The selection of preferred liquid stationary phases after classification by numerical taxonomy techniques. *J. Chromatogr. Sci.* **12**: 617–25 (1974).
- S. Wold. Analysis of similarities and dissimilarities between chromatographic liquid phases by means of pattern cognition. *J. Chromatogr. Sci.* **13**: 525–32 (1975).
- R. Fellous, L. Lizzani-Cuvelier, R. Luft, and D. Lafaye de Micheaux. Data Analysis in gas-liquid chromatography of benzene derivatives. *Anal. Chim. Acta* **154**: 191–201 (1983).
- J.F.K. Huber and G. Reich. Characterization and selection of stationary phases for gas liquid chromatography by pattern recognition methods. *J. Chromatogr.* **294**: 15–29 (1984).
- B. Walczak, M. Dreux, J.R. Chrétien, R. Szymoniak, M. Lafosse, L. Morin-Allory, and J.P. Doucet. Factor analysis and experimental design in high-performance liquid chromatography I. Trends in selectivity of 53 chalcones in reversed phase high-performance liquid chromatography on alkyl- or phenyl-bonded stationary phases. *J. Chromatogr.* **353**: 109–21 (1986).
- J.A. Garcia-Dominguez, J. Garcia-Munoz, V. Menéndez, M.J. Molera, and J.M. Santiuste. Method for the classification and selection of stationary phases in gas chromatography. *J. Chromatogr.* **393**: 209–19 (1987).
- M.H. Abraham, G.S. Whiting, R.M. Doherty, and W.J. Shuley. Hydrogen bonding XV. A new characterization of the McReynolds 77-stationary phase set. *J. Chromatogr.* **518**: 329–48 (1990).
- T.O. Kollie and C.F. Poole. Application of multivariate analysis to the selection of test solutes for characterizing stationary phase selectivity in gas chromatography. *J. Chromatogr.* **550**: 213–37 (1990).
- E.R. Malinowski. *Factor Analysis in Chemistry*, 2nd ed. Wiley Interscience, New York, NY, 1991, Chapter 9, pp. 266–91.
- S.K. Poole and C.F. Poole. Chemometric

- classification of the solvent properties (selectivity) of commonly used gas chromatographic stationary phases. *J. Chromatogr. A* **697**: 415–27 (1995).
13. S.K. Poole and C.F. Poole. Application of principal component factor analysis to the cavity model of solvation to identify factors important in characterizing the solvent properties of gas chromatographic stationary phases. *J. Chromatogr. A* **697**: 429–40 (1995).
  14. C.E. Reese, L. Huang, S.-H. Hsu, S. Tripathy, and C.H. Lochmüller. Universal retention indices and their prediction in reversed-phase liquid chromatography based on principal component analysis and target testing. *J. Chromatogr. Sci.* **34**: 101–10 (1996).
  15. M. Turowski, R. Kalisz, C. Lüllmann, H.G. Genieser, and B. Jastorff. New stationary phases for the high-performance liquid chromatographic separation of nucleosides and cyclic nucleotides. Synthetic and chemometric analysis of retention data. *J. Chromatogr. A* **728**: 201–11 (1996).
  16. K. Héberger and M. Görgényi. Principal component analysis of Kováts indices for carbonyl compounds in capillary gas chromatography. *J. Chromatogr. A* **845**: 21–31 (1999).
  17. K. Héberger. Evaluation of polarity indicators and stationary phases by principal component analysis in gas–liquid chromatography. *Chemometrics Intell. Lab. Systems* **47**: 41–49 (1999).
  18. W.O. McReynolds. Characterization of some liquid phases. *J. Chromatogr. Sci.* **8**: 685–91 (1970).
  19. V.H. Huebner. Determination of the relative polarity of surface active agents by gas–liquid chromatography. *Anal. Chem.* **34**: 488 (1962).
  20. J. Szymanowski. Polarity of nonionic surfactants as determined by gas chromatography. *Crit. Rev. Anal. Chem.* **21**: 407–51 (1990).
  21. J. Ševčík and M.S.H. Löwentap. New criterion for polarity of stationary phases in gas chromatography. *J. Chromatogr.* **217**: 139 (1981).
  22. A. Voelkel. Polarity of liquid phase in gas chromatography. *Wiad. Chem.* **41**: 77–97 (1987).
  23. M. Roth and J. Novak. Correlation of two criteria of polarity for stationary phases in gas–liquid chromatography. *J. Chromatogr.* **234**: 337–45 (1982).
  24. J.H. Hildebrand and R.L. Scott. *The Solubility of Nonelectrolytes*. Van Nostrand, Princeton, NY, 1950.
  25. G. DiPaola-Baranyi and J.E. Guillet. Estimation of polymer solubility parameters by gas chromatography. *Macromolecules* **11**: 228–35 (1978).
  26. C.M. Hansen. The three-dimensional solubility parameter—key to paint component affinities: I. Solvents, plasticizers, polymers and resins. *J. Paint Technol.* **39**: 104–17 (1967).
  27. A. Voelkel and J. Janas. Solubility parameters of broad and narrow distributed oxyethylates of fatty alcohols. *J. Chromatogr. A* **645**: 141–51 (1993).
  28. A. Voelkel and J. Janas. The solubility parameters of fluorine containing oxyethylates. *J. Fluorine Chem.* **67**: 75–81 (1994).
  29. A. Grobler and G. Bálizs. Computer program for a statistical calculation of retention indices. *J. Chromatogr. Sci.* **12**: 57–61 (1974).
  30. S.Wold, K. Esbensen, and P. Geladi. Principal component analysis. *Chemometrics Intell. Lab. Systems* **2**: 37–52 (1987).
  31. M. Otto. *Chemometrics. Statistics and Computer Application in Analytical Chemistry*. Wiley-VCH, Weinheim, 1999, pp. 125–32 and 137–41.
  32. M. Łożyński and E. Krzyżanowska. Mass spectra of 1-(2'-hydroxy-5'-alkylphenyl)-1-alkanone-(E)-oximes. *Org. Mass Spectrom.* **21**: 33–39 (1986).
  33. E. Krzyżanowska, A. Olszanowski, and M. Juskowiak. Synthesis, Spectral data and extraction of copper by 1-(2'-hydroxy-5'-alkylphenyl)-1-alkanone oximes. *J. Prakt. Chemie* **331**: 617–30 (1989).
  34. H.F. Kaiser. The application of electronic computers to factor analysis. *Educ. Psychol. Meas.* **20**: 141–51 (1960).

Manuscript accepted June 7, 2001.