JOURNAL OF PHYSICAL ORGANIC CHEMISTRY, VOL. 2, 232–242 (1989)

NON-HIERARCHICAL CLASSIFICATION OF ORGANIC SOLVENTS USING CHARACTERISTIC VECTOR ANALYSIS OF PHYSICAL AND EMPIRICAL SOLVENT PARAMETERS

ROMUALD I. ZALEWSKI* AND HENRYKA KOKOCIŃSKA

Division of General and Analytical Chemistry, Academy of Economics, Ul. Marchlewskiego 146/150, PL-60-967 Poznań, Poland

AND

CHRISTIAN REICHARDT

Department of Chemistry, Philipps-University, Hans-Meerwein-Strasse, D-3550 Marburg, Fed. Rep. Germany

ABSTRACT

Characteristic vector analysis of a set of six physical and empirical parameters of 103 commonly used organic solvents (bp, ε_r, μ, n_D, E_T^N, and δ) gives four vectors describing 95% of the total data variability. Non-hierarchical cluster analysis, applied to our results, leads to ten separate classes of organic solvents.

INTRODUCTION

Solvent effects on chemical and physical processes, i.e. reaction rates and chemical equilibria as well as spectral absorptions, have been intensively investigated during the last decades, but there are still no reliable and precise methods for their quantitative description and prediction. Organic chemists have usually attempted to explain solvent effects on chemical reactivity and spectral absorption in terms of the so-called 'solvent polarity', using macroscopic physical solvent parameters (e.g. dielectric constant, dipole moment, index of refraction) as well as molecular–microscopic empirical parameters derived from carefully selected solvent-dependent reference processes (e.g. Y, Z, E_T (30), π^*) as solvent polarity measures.[1a,b] A great number of various empirical solvent polarity scales are known. However, due to the complexity of the solute/solvent interactions the search for a unique and general scale of solvent polarity was not entirely successful until now.

Solute/solvent interactions can be roughly divided into specific and nonspecific ones.[1a] The latter comprise directional, inductive, and dispersion forces; specific solute/solvent interactions include hydrogen-bonding, electron-pair donor/acceptor, and solvophobic interactions. In this respect, it is rather surprising that, despite the complexity of solute/solvent interactions, very frequently linear relationships are obtained between solvent-dependent

---

*Author for correspondence

0894–3230/89/030232–11$05.50

Received 20 January 1988
Revised 30 May 1988

processes and single empirical parameters of solvent polarity.[1–3] The chance that solvent-dependent processes will follow such a linear relationship depends strongly on the proper choice of the solvent polarity parameter and thus on the underlying reference process.

Therefore, multiparameter approaches were suggested which describe solvent effects on chemical and physical processes by more than one solvent parameter in the corresponding regression equation. The first of such multiparameter equations has been tried out by Katritzky and colleagues,[4] who were followed by Koppel and Palm[5] and many other authors; see References 1–3, 6 and 7 for reviews.

A rational application of multiparameter equations of the type

$$D = D_0 + a_1X_1 + a_2X_2 + \ldots + a_nX_n \tag{1}$$

requires an *a priori* knowledge of the complexity of the studied solvent-dedependent phenomena, and sets of appropriate explanatory solvent parameters X_1, $X_2, \ldots X_n$ for the various types of solute/solvent interactions. The number of the solvent-dependent data D (D_0 for the gas-phase or a reference solvent) used in the regression analysis has to exceed greatly the number of the explanatory solvent parameters. To reach a mathematical solution which reflects the chemical complexity of the investigated phenomenon, one has to be sure that multicolinearity of the explanatory solvent parameters is not important. Otherwise, the regression coefficients a_1, $a_2, \ldots a_n$ are meaningless. To fulfill all these precautions at once is rather difficult and rarely possible. Various empirical solvent polarity scales are mutually linearly related,[1,8] and this fact limits the value of multiparameter regression analysis considerably.

The awareness of these limitations turned investigators to different statistical approaches such as characteristic vector analysis (CVA)[9] or principal component analysis (PCA).[10] These mathematical analyses start with a given matrix of experimental data (e.g. various scales of solvent parameters) and lead finally to the minimum number of orthogonal principal component scores and principal component loadings which can reproduce the original data set. Applications of this mathematical tool in solving the problem of quantitative evaluation of solvent effects on chemical and physical processes are not numerous. We can refer to the publications by Chastrette *et al.*,[8] Krygowski and Fawcett,[6] Sjöström and Wold,[11] Cramer,[12] Bohle *et al.*,[13] Svoboda *et al.*[14] and Carlson *et al.*.[15] In some of these papers, the data matrices used were incomplete. In all papers, between two and five principal components were generated mathematically and more or less well explained chemically.

The aim of this paper is to perform characteristic vector analysis with a set of six physical and empirical solvent parameters of 103 commonly used organic solvents, and to classify these solvents in a multidimensional space of principal axes using non-hierarchical cluster analysis.[16]

DATA COLLECTING AND PROCESSING

First, we selected more than 120 frequently used organic solvents, covering a wide range of physical and chemical properties, and tried to collect all relevant and important physical and empirical solvent parameters. Unfortunately, this data matrix contained many empty places. Eventually, we decided to choose only a full data matrix without any empty places.[17] In this way, we ended up with a data matrix of size 103 × 6 including 103 solvents and six solvent parameters (cf. Table 1).

In the calculation, we used the following solvent parameters: boiling point (bp), dielectric constant (ε_r), dipole moment (μ), refractive index (n_D), normalized E_T^N-values (derived from a

Table 1. Solvents and their physical and empirical parameters on which characteristic vector analysis was performed

Solvents	bp/°C	ε_r	μ a)	n_D	δ_H b)	E_T^N
(1) Water	100.00	78.30	5.90	1.33	48.13	1.00
(2) Formamide	210.50	111.00	11.20	1.45	39.30	0.80
(3) 1,2-Ethanediol	198.00	37.70	7.60	1.43	29.10	0.79
(4) Methanol	64.70	32.70	5.70	1.33	29.70	0.76
(5) *N*-Methylformamide	182.50	182.40	12.90	1.43	20.30	0.72
(6) Diethylene glycol	244,80	30.90	7.70	1.45	29.13	0.71
(7) Triethylene glycol	278.30	23.40	10.00	1.45	21.90	0.70
(8) 2-Methoxyethanol	124.60	16.93	6.80	1.40	22.10	0.67
(9) Tetraethyleneglycol	327.30	19.70	10.80	1.46	20.30*)	0.66
(10) *N*-Methylacetamide	206.70	179.00	14.70	1.43	29.90	0.66
(11) Ethanol	78.50	24.55	5.80	1.36	26.14	0.65
(12) 2-Aminoethanol	171.00	37.72	7.60	1.45	31.80	0.65
(13) Acetic acid	117.90	6.15	5.60	1.37	20.70	0.65
(14) Benzyl alcohol	205.30	13.10	5.50	1.54	24.80	0.62
(15) 1-Propanol	97.40	20.33	5.50	1.38	24.91	0.62
(16) 1-Butanol	117.20	17.51	5.80	1.40	23.73	0.60
(17) 1-Pentanol	137.80	13.90	5.70	1.41	22.30	0.57
(18) 2-Methyl-1-propanol	108.10	17.93	6.00	1.40	21.50	0.57
(19) 3-Methyl-1-butanol	130.50	14.70	6.10	1.41	22.50	0.57
(20) 2-Propanol	82.40	19.92	5.50	1.38	23.50	0.55
(21) 2-Butanol	99.50	16.56	5.50	1.40	22.10	0.51
(22) Cyclohexanol	161.10	15.00	6.20	1.46	23.30	0.50
(23) Propylene carbonate	241.70	64.92	16.50	1.42	27.30	0.49
(24) 2-Pentanol	118.90	14.72	5.50	1.41	22.00	0.49
(25) Nitromethane	101.20	35.87	11.90	1.38	26.00	0.48
(26) Acetonitrile	81.60	36.00	13.70	1.34	24.77	0.46
(27) 3-Pentanol	116.10	14.02	5.50	1.41	20.80	0.46
(28) Dimethylsulfoxide	189.00	46.70	13.70	1.48	24.50	0.44
(29) Aniline	184.40	6.89	5.00	1.59	21.10	0.42
(30) Sulfolane	285.00	43.30	15.70	1.48	25.45**)	0.41
(31) Acetic anhydride	139.50	20.70	9.40	1.39	21.10	0.41
(32) 2-Methyl-2-propanol	82.30	12.47	5.50	1.39	21.70	0.39
(33) Dimethylformamide	152.30	37.00	13.00	1.43	24.80	0.40
(34) Dimethylacetamide	166.10	37.78	12.40	1.44	22.10	0.40
(35) Propane nitrile	97.30	27.20	11.90	1.37	22.10	0.40
(36) 1-Methylpyrrolidin-2-one	202.00	32.00	13.90	1.47	23.10	0.35
(37) Acetone	56.20	20.70	9.50	1.36	20.50	0.35
(38) Nitrobenzene	210.80	34.82	13.40	1.56	20.50	0.35
(39) Cyanobenzene	190.70	25.20	13.50	1.53	17.20	0.35
(40) 1,2-Diaminoethane	116.50	12.90	6.30	1.46	25.20	0.35
(41) 1,2-Dichloroethane	83.50	10.36	6.20	1.44	20.00	0.35
(42) 2-Butanone	79.60	18.51	9.20	1.38	19.00	0.33
(43) Acetophenone	202.60	17.39	9.90	1.54	21.65	0.33
(44) 2-Methyl-2-pentanol	102.00	5.82	5.70	1.40	19.60	0.32
(45) 2-Pentanone	102.40	15.45	9.00	1.39	18.20	0.32
(46) Dichloromethane	39.80	8.93	5.20	1.42	20.21	0.31
(47) Morpholine	128.30	7.42	5.00	1.45	22.10	0.32
(48) HMPT	235.00	29.60	18.50	1.46	21.50	0.31
(49) 3-Methyl-2-butanone	94.20	15.87	9.20	1.39	17.20*)	0.31
(50) Pyridine	115.50	12.30	7.90	1.51	21.70	0.30
(51) Methyl acetate	56.30	6.68	5.40	1.36	19.60	0.29
(52) Cyclohexanone	155.70	18.30	10.00	1.45	20.30	0.28
(53) 4-Methyl-2-pentanone	116.80	13.11	9.20	1.40	17.20	0.27
(54) 1,1-Dichloroethane	57.30	10.00	6.60	1.42	18.30	0.27
(55) Quinoline	238.00	9.00	7.30	1.63	22.10	0.27
(56) 3-Pentanone	101.70	17.00	9.40	1.39	18.00	0.26
(57) Chloroform	61.70	4.81	3.80	1.45	19.42 **)	0.26
(58) 3,3-Dimethyl-2-butanone	106.30	13.10	9.30	1.40	16.95	0.26
(59) 2,4-Dimethyl-3-pentanone	124.50	17.20	9.20	1.40	16.40	0.25
(60) Diethylene glycol dimethyl ether	162.50	5.80	6.60	1.41	19.30	0.24

(Table 1 continued)

Solvents	bp/°C	ε_r	μ [a)]	n_D	δ_H [b)]	E_T^N
(61) Diethyl carbonate	126.80	2.82	3.00	1.38	18.00	0.24
(62) 1,2-Dimethoxyethane	85.00	7.20	5.70	1.38	17.60	0.23
(63) 1,2-Dichlorobenzene	180.50	9.93	7.60	1.55	20.56	0.23
(64) Ethyl acetate	77.10	6.02	6.30	1.37	18.60	0.23
(65) Fluorobenzene	85.10	5.42	4.90	1.47	18.56	0.19
(66) 2,6-Dimethyl-4-heptanone	169.40	9.91	8.90	1.41	16.00	0.23
(67) Iodobenzene	188.30	4.63	4.60	1.62	20.70	0.17
(68) Chlorobenzene	131.70	5.62	5.10	1.52	19.81	0.19
(69) Bromobenzene	155.90	5.40	5.20	1.56	20.19	0.18
(70) Tetrahydrofuran	66.00	7.58	5.80	1.40	20.30	0.21
(71) Methoxybenzene	155.00	4.33	4.20	1.52	19.50	0.20
(72) 1,1,1-Trichloroethane	74.10	7.53	5.20	1.44	17.40 *)	0.17
(73) 1,4-Dioxane	101.30	2.21	1.50	1.42	20.72	0.16
(74) Trichloroethene	87.20	3.42	2.70	1.48	19.00	0.16
(75) Piperidine	106.40	5.80	4.00	1.45	17.80	0.15
(76) Diphenyl ether	257.90	3.69	3.90	1.58	20.66	0.14
(77) Benzene	80.10	2.28	0.00	1.50	18.74	0.11
(78) 1,2-Dimethylbenzene	144.40	2.57	1.50	1.51	18.00	0.07
(79) 1,4-Dimethylbenzene	138.30	2.27	0.00	1.50	17.94	0.07
(80) 1,3-Dimethylbenzene	139.10	2.37	1.00	1.50	18.00	0.07
(81) Diethyl ether	34.50	4.34	4.30	1.35	15.10	0.12
(82) Di-i-propyl ether	68.30	3.88	4.20	1.37	14.40	0.08
(83) Di-n-butyl ether	142.20	3.08	3.90	1.40	15.90	0.07
(84) Carbon disulfide	46.20	2.64	0.00	1.63	20.29	0.06
(85) Methylcyclohexane	100.90	2.02	0.00	1.42	15.99	0.01
(86) Tetrachloroethene	121.20	2.30	0.00	1.51	19.00	0.10
(87) Toluene	110.60	2.38	1.40	1.50	18.23	0.10
(88) Triethylamine	89.30	2.42	2.90	1.40	15.18	0.04
(89) Mesitylene	164.70	2.28	0.00	1.50	18.00	0.07
(90) Tetrachloromethane	76.50	2.24	0.00	1.46	17.49	0.05
(91) Cyclohexane	80.70	2.02	0.00	1.43	16.78	0.01
(92) n-Hexane	68.70	1.88	0.00	1.37	14.87	0.01
(93) n-Butyronitrile	118.00	20.30	11.90	1.38	21.50	0.38
(94) 1,3-Dichlorobenzene	173.00	5.04	4.60	1.55	20.00	0.18
(95) Ethyl benzoate	213.00	6.02	6.60	1.50	19.95	0.23
(96) 1-Octanol	194.40	10.34	5.90	1.43	21.10 **)	0.54
(97) 2,2,2-Trifluoro-ethanol	74.00	26.14	8.40	1.29	25.25 **)	0.90
(98) Trifluoroacetic acid	72.40	8.55	7.60	1.29	21.58	1.08
(99) Diethylamine	56.30	3.58	3.70	1.39	16.40	0.15
(100) n-Pentane	36.10	1.84	0.00	1.36	14.36	0.01
(101) Cyclopentane	49.20	1.97	0.00	1.41	16.57	0.01
(102) n-Heptane	98.40	1.92	0.00	1.39	15.20	0.01
(103) cis-Decalin	194.60	2.15	0.00	1.48	18.80	0.01

[a)] In units of 10^{-30} C m. [b)] In units of MPa$^{1/2}$. Data for δ_H are from reference (18), except those marked with *) and **) which are taken from references (19) and (8c), respectively.

solvatochromic pyridinium-N-phenolate betaine dye),[1c] and Hildebrand's cohesion parameter (δ_H).[18] All physical data were collected from recently published sources[18,19] and are presented in Table 1.

The E_T^N-values can be considered as solvent Lewis acidity parameters.[1] Unfortunately, it was not possible to include a solvent Lewis basicity parameter into the calculations because of a lack of data. Commonly used Lewis basicity scales such as the donor numbers DN (Gutmann), and $-\Delta H^0_{BF_3}$-values (Gal, Maria), the B_{MeOD}-values (Koppel, Palm),[5] and the β-values (Kamlet, Taft)[2] are known for a limited number of solvents only (see Reference 1a for a review). In particular, they are not known for all solvents of Table 1. Provided that only a full data matrix without empty places should be used, the inclusion of a solvent Lewis basicity parameter would considerably reduce the number of solvents, and the solvent classification derived therefrom would be less valid.

The method of characteristic vector or principal component analysis has been described previously,[9] and only a brief description will be given here. Solvent parameters P for $i = 1,2,\dots n$ scales and $m = 1,2,\dots k$ solvents form a data matrix of format $i{\cdot}m$. The experimental data matrix can be approximated by the CVA, equation (2),

$$P_{i,m} = \bar{P}_i + V_{1,i}L_{m,1} + \dots + V_{l,i}L_{m,l} + \varepsilon_{i,m} \tag{2}$$

where the characteristic vectors V are specific to the variation of scales i, and L, the characteristic vector loadings, describe their magnitudes which must be added to the mean data vector $\bar{P}_i$ in order to reconstitute the experimental solvent parameter set. Characteristic vectors V are contained in the matrix $V\ (i,l)$, and their loadings in the matrix $L\ (m,l)$. The number of characteristic vectors, l, existing in space is equal to, or less than the number of solvent scales in the data set, n. The first characteristic vector is defined as the linear combination of variables which has the maximum variance of all linear functions derivable from the given variables. The second characteristic vector is the linear combination of variables having the maximum variance of all linear functions of the given variables that are orthogonal to the first characteristic vector, and so on.

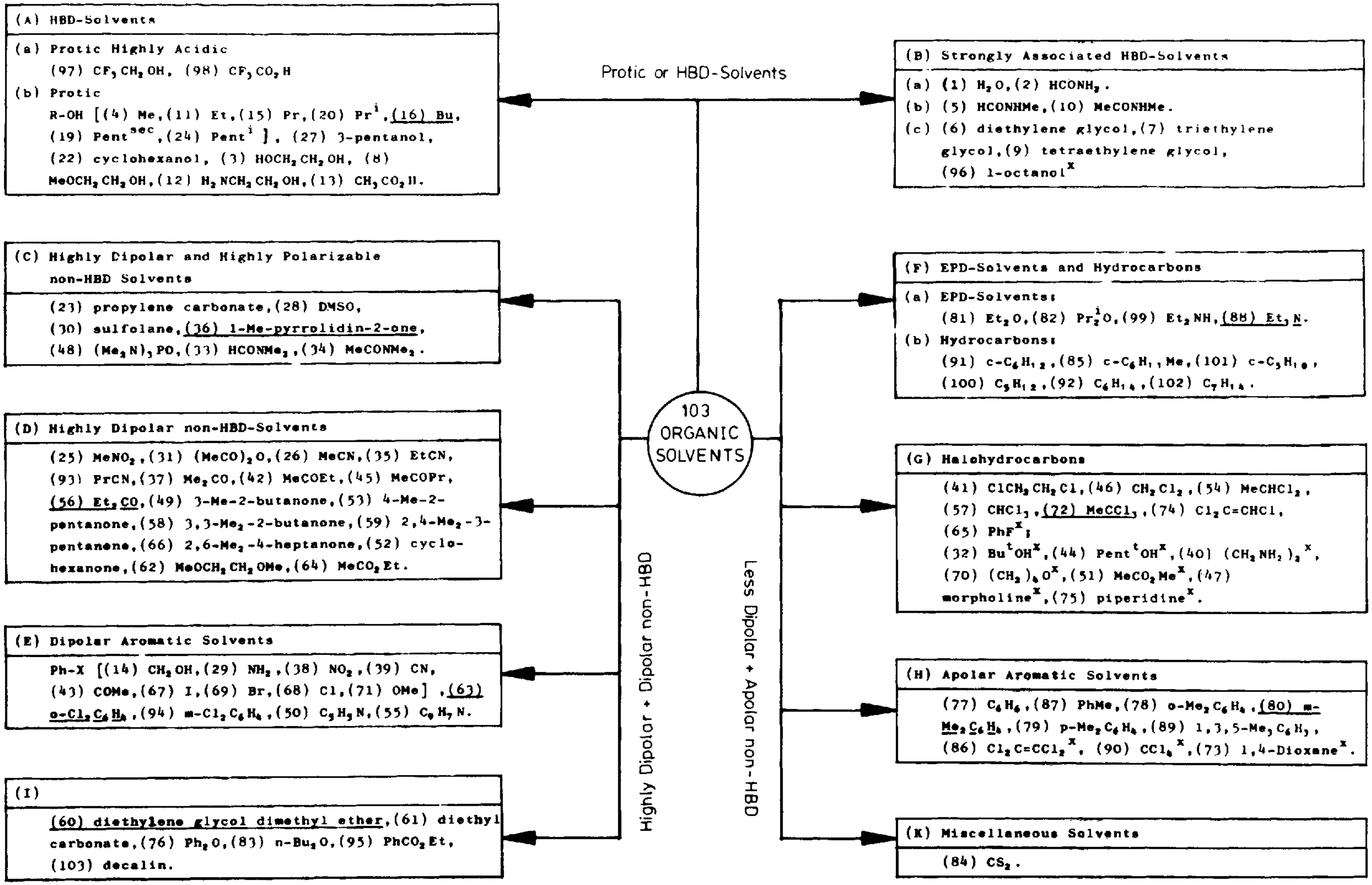
The importance of the particular characteristic vector is measured by the percentage of the total variability (TV) explained by this vector, % TV, and by the variance of the difference $\varepsilon_{i,m}$ between the experimental and the reconstituted data matrix.

An appropriate program has been executed with the computer ODRA 1305. With the resulting matrix of characteristic vector loadings a standard non-hierarchical cluster analysis has been carried out using the ICL-1900 package.[16] The classification started initially with 15 solvent classes, the number of which was reduced by a mathematical analysis averaging the distance of points in a six-dimensional space. The advantage of this method is that, being non-hierarchical, it distributes all the individual solvents at a completely new classification level. The degree of fit in non-hierarchical cluster analyses can be measured by F statistics and is best if $F = \max$. Finally, ten clusters of solvents appeared in the multidimensional space as the best solution (cf. Scheme 1).

RESULTS AND DISCUSSION

Starting with the experimental data matrix of the size 103 × 6 which includes the numerical data of boiling point bp, dielectric constant ε_r, dipole moment μ, refractive index n_D,[18] the empirical solvent parameter E_T^N.[1c] and Hildebrand's cohesion parameter δ_H[18] for 103 organic solvents (cf. Table 1), the solvent classification given in Scheme 1 was finally obtained. As indicated by the correlation coefficients r given in Table 2, the mutual linear correlation between the six solvent parameters chosen is rather poor, as desired. The best, but still poor, correlation exists between δ_H and E_T^N ($r = 0{\cdot}7640$) as well as δ_H and ε_r ($r = 0{\cdot}6210$).

All the experimental solvent data, distributed over a six-dimensional space, can be described by only six characteristic vectors. The first characteristic vector accounts for 46·2%



(A) HBD-Solvents

(a) Protic Highly Acidic
(97) CF_3CH_2OH, (98) CF_3CO_2H
(b) Protic
R-OH [(4) Me, (11) Et, (15) Pr, (20) Pr^i, (16) Bu, (19) $Pent^{sec}$, (24) $Pent^i$], (27) 3-pentanol, (22) cyclohexanol, (3) $HOCH_2CH_2OH$, (8) $MeOCH_2CH_2OH$, (12) $H_2NCH_2CH_2OH$, (13) CH_3CO_2H.

(B) Strongly Associated HBD-Solvents

(a) (1) H_2O, (2) $HCONH_2$.
(b) (5) HCONHMe, (10) MeCONHMe.
(c) (6) diethylene glycol, (7) triethylene glycol, (9) tetraethylene glycol, (96) 1-octanol^x

(C) Highly Dipolar and Highly Polarizable non-HBD Solvents

(23) propylene carbonate, (28) DMSO, (30) sulfolane, (36) 1-Me-pyrrolidin-2-one, (48) $(Me_2N)_3PO$, (33) $HCONMe_2$, (34) $MeCONMe_2$.

(D) Highly Dipolar non-HBD-Solvents

(25) $MeNO_2$, (31) $(MeCO)_2O$, (26) MeCN, (35) EtCN, (93) PrCN, (37) Me_2CO, (42) MeCOEt, (45) MeCOPr, (56) Et_2CO, (49) 3-Me-2-butanone, (53) 4-Me-2-pentanone, (58) 3,3-Me_2-2-butanone, (59) 2,4-Me_2-3-pentanone, (66) 2,6-Me_2-4-heptanone, (52) cyclohexanone, (62) $MeOCH_2CH_2OMe$, (64) $MeCO_2Et$.

(E) Dipolar Aromatic Solvents

Ph-X [(14) CH_2OH, (29) NH_2, (38) NO_2, (39) CN, (43) COMe, (67) I, (69) Br, (68) Cl, (71) OMe], (63) o-$Cl_2C_6H_4$, (94) m-$Cl_2C_6H_4$, (50) C_5H_5N, (55) C_9H_7N.

(I)

(60) diethylene glycol dimethyl ether, (61) diethyl carbonate, (76) Ph_2O, (83) n-Bu_2O, (95) $PhCO_2Et$, (103) decalin.

(F) EPD-Solvents and Hydrocarbons

(a) EPD-Solvents:
(81) Et_2O, (82) Pr^i_2O, (99) Et_2NH, (88) Et_3N.
(b) Hydrocarbons:
(91) c-C_6H_{12}, (85) c-$C_6H_{11}Me$, (101) c-C_5H_{10}, (100) C_5H_{12}, (92) C_6H_{14}, (102) C_7H_{16}.

(G) Halohydrocarbons

(41) $ClCH_2CH_2Cl$, (46) CH_2Cl_2, (54) $MeCHCl_2$, (57) $CHCl_3$, (72) $MeCCl_3$, (74) $Cl_2C{=}CHCl$, (65) PhF^x;
(32) Bu^tOH^x, (44) $Pent^tOH^x$, (40) $(CH_2NH_2)_2{}^x$, (70) $(CH_2)_4O^x$, (51) $MeCO_2Me^x$, (47) morpholine^x, (75) piperidine^x.

(H) Apolar Aromatic Solvents

(77) C_6H_6, (87) PhMe, (78) o-$Me_2C_6H_4$, (80) m-$Me_2C_6H_4$, (79) p-$Me_2C_6H_4$, (89) 1,3,5-$Me_3C_6H_3$, (86) $Cl_2C{=}CCl_2{}^x$, (90) $CCl_4{}^x$, (73) 1,4-Dioxane^x.

(K) Miscellaneous Solvents

(84) CS_2.

Scheme 1. Classification of 103 organic solvents by characteristic vector analysis

Table 2. Correlation coefficients r for the mutual linear correlation between the six solvent parameters used in the characteristic vector analysis (cf. Table 1)

	bp	ε_r	μ	n_D	δ_H	E_T^N
bp	1·0000	0·3166	0·4634	0·5208	0·2844	0·2463
ε_r		1·0000	0·5732	−0·1053	0·5690	0·5321
μ			1·0000	−0·1006	0·3789	0·5104
n_D				1·0000	−0·0737	−0·3357
δ_H					1·0000	0·7444
E_T^N						1·0000

of the total data variability, the second for more than 28·2%, and the third for about 13%. All three vectors together describe 88·0% of the total data variability. Inclusion of a fourth characteristic vector increases this amount to 95%. The unexplained residue is less than 10% of the bp and E_T^N-scale, less than 5% of the μ and n_D-scale, and less than 2% for the δ_H and ε_r-scale.

The resulting matrix of the characteristic vector loadings has then been used to conduct a non-hierarchical classification of the 103 organic solvents. To reduce the computer execution time, the classification started with the fifteen solvent classes and stopped with one. The goodness of classification was expressed by the function F, which reaches a maximum when the classification is best from the statistical point of view. A compromise between statistical demands and chemical sense forced us to accept finally only ten solvent classes (with F = 2·0771) and not twelve (with F = 2·3249).

This classification of solvents into ten classes (A)–(K) is presented in Scheme 1, together with the proposed group headings. Strange placements of solvents according to this classification are indicated by an asterisk. The solvent which is closest to the class center of each solvent class is underlined.

According to Scheme 1, hydrogen-bond donor solvents (i.e. HBD or protic solvents) form two different classes (A) and (B). One of them, called ‘strongly associated HBD-solvents’, reasonably includes water, *N*-methyl amides and amides, various ethylene glycols, and as a strange member 1-octanol [no. (96) in Table 1]. The grouping of (B) into three subclasses (a)–(c) is due to the difference in the magnitude of the characteristic vector loading $L_{2,m}$. The other class of HBD solvents, (A), includes highly acidic and common carboxylic acids and alcohols. The high acidity of the HBD-solvents no. (97) and (98) is caused by the CF_3 group. Only two common alcohols, *t*-butanol (32) and *t*-pentanol (44) are surprisingly found in another solvent class (G), maybe due to steric hindrance of hydrogen-bonding.

There are two classes, (C) and (D), of dipolar aprotic, i.e. dipolar non-HBD solvents: one of them, (C), includes highly polarizable dipolar non-HBD solvents such as propylene carbonate, DMSO, HMPT, and DMF, centered around 1-methylpyrrolidin-2-one. The other class (D) consists mainly of ketones and nitriles, centered around 3-pentanone.

Satisfactorily, class (D) does not contain aromatic solvents, in contrast to the analogous class of ‘aprotic highly dipolar (AHD) solvents’ in Chastrette's classification.[8c]

Dipolar aromatic solvents flock together in class (E), which includes mono- and disubstituted benzenes with electron-donating and electron-attracting substituents, pyridine and quinoline, the high polarizability of which may be similar to that of aromatic solvents.

This class (E) is well-separated from class (H) containing less dipolar and apolar aromatic solvents such as benzene and its methyl derivatives. The two perhalohydrocarbons no. (86) and (90) found in class (H) are also apolar ($\mu = 0$) but their high polarizability seems to be similar to that of aromatic solvents.

Electro-pair donor (EPD) solvents such as ethers and amines, centered around triethylamine, form class (F) together with the aliphatic hydrocarbons. Presumably, this grouping stems from the fact that the six solvent parameters of Table 1 are not suitable to differentiate solvents according to their Lewis basicity. Introduction of an additional parameter of Lewis basicity should lead to a separation of the two solvent groups in class (F). In Chastrette's classification scheme the aliphatic hydrocarbons are also found in the group of EPD solvents.[8c]

The majority of halohydrocarbons constitutes class (G), centered around 1,1,1-trichloroethane, with solvents no. (32), (44), (40), (65), (70), (51), (47), and (75) as strange members. Fluorobenzene (65) should be located in class (E), and solvents no. (40), (70), (47), and (75) in class (F).

Class (I) consists of six solvents which appear to have nothing in common; finally CS_2 (84) forms a separate class (K).

Altogether, 91 of the 103 organic solvents listed in Table 1 (i.e. 89%) are ordered into nine classes (A)–(I) which correlate to some extent with the chemist's intuition, by using a purely statistical method without any premises. The intuitive solvent classification of Parker[1a] (i.e. protic/dipolar aprotic/apolar aprotic solvents) can be recognized in Scheme 1.

The relative position of solvents in the multidimensional space gives information on the similarities and dissimilarities between the solvents. Solvents with similar properties are located close to each other in the space and form solvent classes. However, solvents located in between or at the border of such classes can be sometimes statistically classified against chemical intuition. Therefore, the results of purely statistical methods such as characteristic vector analysis need testing and interpretation by a chemist.

The information given by characteristic vector analysis on the systematic variation in the solvent space can be used in different ways: (i) to classify solvents into classes of solvents with similar properties; (ii) to select a representative, typical solvent for each solvent class (underlined in Scheme 1); (iii) to select solvents with properties disimilar to other solvents as much as possible, using maximum distances in the solvent space; and (iv) to select the best solvent for a particular chemical reaction by application of the simplex strategy in a three-dimensional solvent space.[20]

The results given in Scheme 1, and in References 8, 12–15, should not be considered as final solutions of the solvent classification problem. It is very likely that the characteristic vector projection will change if other solvents or new physical solvent properties and empirical solvent scales are used in the calculations. It should be stressed, however, that our results, given in Scheme 1, are essentially similar to those obtained by Chastrette *et al.*,[8c] in spite of the fact that we started with a different number of solvents with different sets of solvent parameters.

An analysis of the plot of characteristic vector values V_1 versus V_2 is shown in Figure 1. This plot indicates that the V_1 axis is mainly explained by ε_r, μ, and δ_H, whereas the V_2 axis is strongly correlated with n_D. This means that solvent dipolarity and polarizability are the most important descriptors. E_T^N and bp seem to correlate well with both axes. A plot of the corresponding vector loadings L_2 and L_1 gives a very complicated figure due to the large number of points. In a first approximation, this planar plot reflects the results of the non-hierarchical cluster analysis quite well. More information can be gained by a detailed

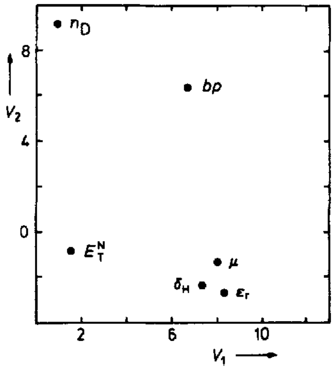


Figure 1. Characteristic vector values V_1 versus V_2

exploration of these data for a particular solvent class with respect to the chemical structure of the solvents. For example, class (A) of HBD solvents includes various alcohols and two carboxylic acids. L_1 varies very little in this solvent class, as compared to the large variation in L_2. The magnitude of L_2 depends mainly on the polarizability as measured by the molar refraction $R_M = (n^2 - 1)V_m/(n^2 + 2)$. This is shown in Figure 2a and reflects the influence of the structure of the alkyl group (*tert* < iso < *sec* < normal). Consequently, methanol has the lowest value of L_2 among the alcohols. Cyclohexanol (22) and the substituted alcohols (3), (8), and (12) do not follow the curve given in Figure 2a. On the same curve one can find points for the three nitriles (26), (35) and (93). A similar pattern can be seem for the dipolar non-HBD solvents of class (D). Again, the distribution of points along the L_1 axis is rather small, and the main variation occurs in a similar fashion as for alcohols and nitriles. Only cyclohexanone does not follow the curve for ketones.

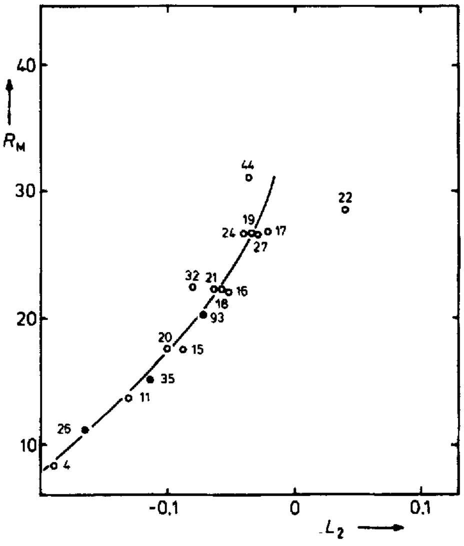


Figure 2a. Characteristic vector loadings L_2 versus molar refraction, R_M, for 17 solvents (○ alcohols, ● nitriles); cyclohexanol (22) excluded. Solvent numbering as in Table 1

A different picture can be seen for alkanes and cycloalkanes, for which L_1 and L_2 values vary in a similar way. A corresponding plot of molar refraction R_M versus L_2 for these solvents is given in Figure 2b.

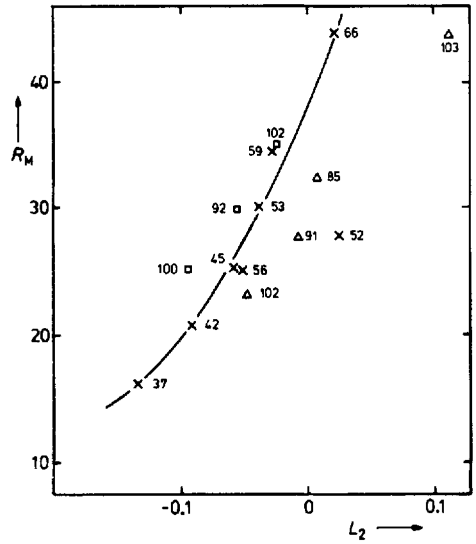


Figure 2b. Characteristic vector loadings L_2 versus molar refraction, R_M, for 14 solvents (× ketones, □ alkanes, △ cycloalkanes). The curve represents the ketones; cyclohexanone (52) excluded. Solvent numbering as in Table 1

CONCLUSION

In conclusion, the present work illustrates the possibility of getting a fairly reasonable classification of organic solvents by application of purely statistical methods, i.e. characteristic vector analysis and non-hierarchical cluster analysis, to a matrix of 103 solvents with six physical and empirical solvent parameters. No other premises are needed, but the results deserve interpretation by a chemist, particularly in the case of the borderline behavior of some solvents. Certain limitations are obvious (solvents in wrong places; difficult interpretation of characteristic vector loadings with respect to chemical structure), particularly due to the limited number of physical and empirical solvent parameters, which do not express all possible solute/solvent interactions. In spite of these limitations, the solvent classification obtained in this work is very similar to those obtained by other authors with other sets of solvents and solvent parameters.[8,12–15]

ACKNOWLEDGEMENT

One of us (R. I. Z.) thanks the Polish Academy of Science for a grant (No. CPBP 01.13.1.24).

REFERENCES

1. (a) C. Reichardt, *Solvents and Solvent Effects in Organic Chemistry*. Second edn, VCH Publishers, Weinheim 1988; (b) C. Reichardt, *Empirical Parameters of Solvent Polarity and Chemical Reactivity*, in H. Ratajczak and W. J. Orville-Thomas (eds.), *Molecular Interactions*. Vol. 3, Chapter 5, pp. 241–282, Wiley, Chichester 1982; (c) C. Reichardt and E. Harbusch-Görnert, *Liebigs Ann. Chem.*, 721 (1983); C. Laurence, P. Nicolet, M. Lucon and C. Reichardt, *Bull. Soc. Chim. Fr.*, **125**, 1001 (1987).

2. J.-L. M. Abboud, M. J. Kamlet and R. W. Taft, *Progr. Phys. Org. Chem.*, **13**, 485–630 (1981).
3. J. Shorter, *Correlation Analysis in Organic Reactivity*, Research Studies Press, Chichester 1982, Chapter 5, pp. 127–172.
4. F. W. Fowler, A. R. Katritzky and R. J. D. Rutherford, *J. Chem. Soc., Part B*, 460–469 (1971).
5. I. A. Koppel and V. A. Palm, *The Influence of the Solvent on Organic Reactivity*, in N. B. Chapman and J. Shorter (eds.), *Advances in Linear Free Energy Relationships*, Plenum Press, London 1972, Chapter 5, pp. 203–280.
6. T. M. Krygowski and W. R. Fawcett, *Can. J. Chem.*, **54**, 3283–3292 (1976).
7. C. Reichardt, *Angew. Chem.*, **91**, 119–131 (1979); *Angew. Chem., Int. Ed. Engl.*, **18**, 98–110 (1979).
8. (a) M. Chastrette and J. Carretto, *Tetrahedron*, **38**, 1615–1618 (1982); (b) M. Chastrette, *Tetrahedron*, **35**, 1441–1448 (1979); (c) M. Chastrette, M. Rajzmann, M. Chanon and K. F. Purcell, *J. Am. Chem. Soc.*, **107**, 1–11 (1985).
9. R. I. Zalewski, *J. Chem. Soc., Perkin Trans. II*, 495–499 (1986).
10. E. R. Malinowski and D. G. Howery, *Factor Analysis in Chemistry*, Wiley, New York, 1980.
11. M. Sjöström and S. Wold, *J. Chem. Soc., Perkin Trans. II*, 104–109 (1981).
12. R. D. Cramer, *J. Am. Chem. Soc.*, **102**, 1837–1849, 1849–1859 (1980).
13. M. Bohle, W. Kollecker and D. Martin, *Z. Chem.*, **17**, 161–168 (1977).
14. P. Svoboda, O. Pytela and M. Večeřa, *Collect. Czech. Chem. Commun.*, **48**, 3287–3306 (1983).
15. R. Carlson, T. Lundstedt and C. Albano, *Acta Chem. Scand., Part B*, **39**, 79–91 (1985).
16. D. N. Sparks, *Applied Statistics*, **22**, 127 ff. (1983).
17. P. F. Strong, C. G. Swain and M. S. Swain, *J. Chem. Inf. Comput. Sci.*, **19**, 13–18 (1979).
18. J. A. Riddick, W. B. Bunger and T. K. Sakano, *Organic Solvents. Physical Properties and Methods of Purification*. 4th edn, in A. Weissberger (ed.), *Techniques of Chemistry*. Vol. II, Wiley-Interscience, New York 1986.
19. A. F. M. Barton: *Handbook of Solubility Parameters and Other Cohesion Parameters*, CRC Press, Boca Raton/Florida 1983.
20. M. J. Kamlet, J.-L. M. Abboud and R. W. Taft, *J. Am. Chem. Soc.*, **99**, 6027–6038 (1977); M. J. Kamlet and R. W. Taft, *J. Chem. Soc., Perkin Trans. II*, 337–341 (1979).