

## A NEW CLASSIFICATION OF SOLVENTS BASED ON CHEMOMETRIC EMPIRICAL SCALE OF PARAMETERS

Oldřich PYTELA

*Department of Organic Chemistry,  
Institute of Chemical Technology, 532 10 Pardubice*

Received June 14, 1989

Accepted June 27, 1989

The paper presents a classification of 51 solvents based on clustering in three-dimensional space formed by the empirical scale of PAC, PBC, and PPC parameters designed for interpretation of solvent effect on a model with cross-terms. For the classification used are the clustering methods of the nearest neighbour, of the furthest neighbour, of average bond, and the centroid method. As a result, the solvents have been divided into 8 classes denoted as: I — nonpolar–inert solvents (aliphatic hydrocarbons), Iip — nonpolar–polarizable (aromatic hydrocarbons, tetrachloromethane, carbon disulphide), Iib — nonpolar–basic (ethers, triethylamine), IIIp — little polar–polarizable (aliphatic halogen derivatives, substituted benzenes with heteroatom-containing substituents), IIIb — little polar–basic (cyclic ethers, ketones, esters, pyridine), IVa — polar–aprotic (acetanhydride, dialkylamides, acetonitrile, nitromethane, dimethyl sulfoxide, sulfolane), IVp — polar–protic (alcohols, acetic acid), and V — exceptional solvents (water, formamide, glycol, hexamethylphosphoric triamide). The information content of the individual parameters used for the classification has been determined. The classification is based primarily on solvent polarity/acidity (PAC), less on polarity/basicity (PBC), and the least on polarity/polarizability (PPC). Causal relation between chemical structure of solvent and its effect on the process taking place therein has been established.

The classification of solvents is important both theoretically and practically, that used at present being based on various macroscopic and microscopic properties of solvents (for a review see ref.<sup>1</sup>). By their chemical constitution, solvents can be classified as molecular, ionic, and atomic liquids. Another classification is based on physical properties such as the boiling temperature, dipole moment, relative permittivity, refractive index, internal pressure, cohesion pressure, etc. An important and frequently used classification is based on acid–base behaviour of solvents. Brönsted<sup>2–4</sup> distinguishes 8 classes of solvents (amphiprotic with high relative permittivity, protogenic with low relative permittivity, protophilic with low relative permittivity, and aprotic with low relative permittivity). This classification became a basis for the well-known classification by Brönsted & Lowry<sup>1</sup>: amphiprotic solvents (neutral, protogenic, protophilic) and aprotic solvents (dipolar protophilic, dipolar protophobic, and inert). Specific solvent–solute interactions form the basis of a classification by Parker<sup>5–7</sup>, who distinguishes only three solvent types (apolar protic,

dipolar aprotic, and protic) with possible transition types. Janz and Danyluck<sup>8</sup> apply the conductivity of hydrogen chloride solution to distinguish the levelling solvents from the differentiating ones. Acid-base properties of solvents are described in a report by Bagno and Scorano<sup>9</sup>. Rice and Bobbitt<sup>10</sup> dealt with a classification of solvents for chromatographical purposes.

Mathematical-statistical methods are very useful for classifications of solvents, since they enable to quantitatively treat big data sets. The obtained general characteristics of solvents which are given in literature<sup>11-26</sup> can serve as a basis for solvent classification<sup>14,15,22,24</sup>. The classification by Chastrette et al.<sup>22</sup> can serve as an example: it divides solvents into four basic classes – nonpolar (with the subclasses of electron-pair donors, aromatic relatively polar, and aromatic nonpolar), aprotic (with the subclasses dipolar, highly dipolar, and highly dipolar–highly polarizable), protic (with the subclasses hydrogen-bond-forming and strong-hydrogen-bond-forming solvents), and miscellaneous. Since, however, the starting data included both physical properties of solvents and some empirical parameters for description of solvent effect, some solvents are classified illogically or even paradoxically, e.g. cyclohexane together with ethers and amines, tetrahydrofurane with alcohols and acetic acid, etc. The same approach was chosen by the authors of ref.<sup>24</sup>: they used the boiling temperatures, relative permittivities, dipole moments, refractive indexes, the Hildebrand  $\delta$  parameters, and the empirical  $E_T^N$  parameters as the basis for a classification of 103 solvents. Applying the method of principal components and subsequent nonhierarchical classification, they divided the solvents into 10 groups, the group denoted as halogen derivatives of hydrocarbons involving 1,2-dichloroethane and fluorobenzene together with tert.butyl alcohol, tetrahydrofurane, methyl acetate, and piperidine. Such results are obviously due to combination of incongruous properties of solvents as it is documented by the correlation matrix given in the cited paper<sup>24</sup> as well as by the amount of information isolated therefrom into the individual principal components from 8 columns (46.2%, 28.2%, and 13%). Perhaps more promising are the mathematical-statistical methods applied to homogeneous characteristics of solvents such as empirical and semi-empirical parameters.

The aim of the present communication is a classification of solvents based on statistical parameters obtained by the method of conjugated deviations<sup>25,26</sup> on a model with cross-terms<sup>26</sup> from a large set of experimental data (6 334 data in 367 series).

## CALCULATIONS

*Pretreatment of data and methods used.* For the calculation we have used the known<sup>26</sup> three-parameter empirical scale for description of solvent effect involving the cross-terms. The parameters of this scale standardized within the interval  $\langle 0, 1 \rangle$

were standardized to zero average and unit standard deviation before the classification calculation. The Minkowski metrics (1) was used for the calculation of the distance matrix

$$d_{ij} = \left[ \sum_{k=1}^p (x_{ik} - x_{jk})^\lambda \right]^{1/\lambda}, \quad (1)$$

where  $i$  and  $j$  are the objects considered,  $p$  is the space dimension (in the present paper  $p = 3$ ),  $x$  are the coordinates of the object in space, and  $\lambda$  is a coefficient. The values used in the present paper were  $\lambda = 2$  (the Euclidean metrics) and  $\lambda = 1$  (Hamming's metrics).

*Classification methods*<sup>27-31</sup>. For the classification we used the method of hierarchical clustering<sup>27-31</sup>, the method of the nearest neighbour<sup>32</sup>, that of the furthest neighbour<sup>33</sup>, the average bond method<sup>30</sup> (the average nonsimilarity<sup>31</sup>) and the centroid method<sup>34</sup>. The course of clustering was followed by the command variable of the method and by the average intra-class variability  $S_I$  defined as follows:

$$S_I = \sum_{m=1}^M \left[ \left( \sum_{i=1}^{N_m} \sum_{j>1}^{N_m} d_{ij} \right) / (0.5 (N_m - 1) N_m) \right] / M, \quad (2)$$

where  $M$  means number of classes of objects (solvents),  $N_m$  is number of objects in the  $m$ -th class, and  $d_{ij}$  is defined in Eq. (1).

The information measure of the individual columns (parameters for description of solvent effect) was determined as the quotient of interclass and intraclass variabilities defined analogously by Eq. (2), of course, only for one selected dimension (parameter).

The calculations were carried according to our own programs in the FORTRAN language using an EC 1 033 computer.

## RESULTS AND DISCUSSION

In the clustering of 51 objects (solvents) in the three-dimensional space of parameters PAC, PBC, and PPC from ref.<sup>26</sup>, attention was focused upon the phases in which there occurred no distinct change in the command variable of the method and in the average intraclass variability  $S_I$ , i.e. the phases of formation of stable clusters.

The clustering by the method of the nearest neighbour was not successful from the standpoint of classification, since it suffered from the well-known chaining effect. This effect, however, can be used to order the solvents into a series (one of the variants of solution of the problem of the shortest nonclosed path). In the Minkowski metrics ( $\lambda = 1, 2$ ) this method gave the solvent series given in Table II (with application of the numbers of Table I). From the results it is obvious that the expected order from water (35) to hexane (1) is obtained in the Euclidean metrics, which corresponds

to the protic properties of alcohols (the beginning) and to the solvent polarity. However, the connection of hexamethylphosphoric triamide (HMPA) (32) and carbon disulfide (51), i.e. in a broader context the combination of polar aprotic

TABLE I  
Numbers of the solvents used in the classification

No.	Solvent	No.	Solvent	No.	Solvent
1	Hexane	18	Diisopropyl ether	35	Water
2	Heptane	19	Anisole	36	Methanol
3	Cyclohexane	20	Phenetole	37	Ethanol
4	Benzene	21	Tetrahydrofurane	38	1-Butanol
5	Toluene	22	Dioxane	39	2-Propanol
6	<i>m</i> -Xylene	23	Acetone	40	tert. Butyl alcohol
7	<i>p</i> -Xylene	24	Butanone	41	Benzyl alcohol
8	Mesitylene	25	Cyclohexanone	42	1,2-Ethanediol
9	Tetrachloromethane	26	Methyl acetate	43	2-Methoxyethanol
10	Chloroform	27	Ethyl acetate	44	Acetic acid
11	Dichloromethane	28	Acetanhydride	45	Triethylamine
12	1,2-Dichloroethane	29	Formamide	46	Pyridine
13	Chlorobenzene	30	N,N-Dimethylformamide	47	Nitromethane
14	Bromobenzene	31	N,N-Dimethylacetamide	48	Nitrobenzene
15	Fluorobenzene	32	HMPA <sup>a</sup>	49	Dimethyl sulphoxide
16	Diethyl ether	33	Acetonitrile	50	Sulfolane
17	Dibutyl ether	34	Benzonitrile	51	Carbon disulfide

<sup>a</sup> Hexamethylphosphoric triamide.

TABLE II  
The continuous solvent sequences obtained by the method of the nearest neighbour in the Minowski metrics with  $\lambda = 2$  and  $\lambda = 1$  (for numbers see Table I)

$\lambda = 2$  (the Euclidean metrics), standardizaion

35-29-42-41-43-40-38-39-37-36-44-34-48-12-11-10-13-14-15-19-20-22-27-21-26-24-25-23-46-33-28-47-50-49-30-31-32-51-9-8-6-7-5-4-16-18-17-45-3-2-1

$\lambda = 1$  (Hamming's metrics), standardization

51-4-5-7-6-8-9-3-2-1-45-17-18-16-40-38-39-37-36-44-35-29-42-41-43-28-33-46-23-25-24-26-27-21-22-20-19-15-14-13-10-11-12-48-34-47-50-49-30-31-32

solvents and nonpolar aromatics is illogical. The only similarity consists in the ability to donate electrons. On the other hand, in Hamming's metrics HMPA (32) and carbon disulphide appear at the opposite ends of the series, which indicates an incongruous connection, in the previous case only in the final phase of clustering. The sequence found corresponds less to the conventional ideas, and it is given by the sequence of polarizable solvents (aromatics) through inert aliphatic hydrocarbons, basic solvents (ethers), protic alcohols, followed by ketones, esters, chlorinated hydrocarbons, till polar aprotic solvents. Although some of the sequences are found in both the metrics, it is obvious that the solvents classified are not located in the parameter space along a curve. The perspective view rather shows a triangular formation with the vertexes formed by water, HMPA, and alkanes, the surface being bent from the vertex formed by water to the opposite vertexes. Therefore it is obviously difficult (even though, of course, possible) to define a general solvent sequence with continuous changes of their properties.

The classifications using the method of the furthest neighbour, of average bond, and the centroid method led roughly to the same results in terms of one metrics, somewhat greater differences being found between the results in the two metrics used within one classification method. As the Euclidean metrics provided better resolution of stable clusters (a more distinct separation of the phases with a small change in the command variables followed and large change of these quantities), the results are interpreted in terms of this metrics. Since the methods given rather prefer the clusters of "spherical" shape, the clustering proceeded uniformly in the sense of assignment to the individual classes. The method of average bond and the centroid method gave, in the region practically interesting (about 5–10 clusters), a stable arrangement with 10 clusters identical in both the methods. The objects which remained unclassified were Nos 32 (HMPA), 35 (water), a two-member cluster being formed by formamide (29) and glycol (42). If the solvents mentioned are ignored, one obtains seven clusters of relatively clean-cut properties (Table III). From the table it is obvious that the clusters can be divided by the P parameter (which primarily corresponds to polarity<sup>26</sup>) into four classes – entirely nonpolar solvents in the class I, nonpolar solvents in the class II, little polar solvents in III, and the polar ones in IV. This classification corresponds to the Kirkwood polarity function  $(\epsilon - 1)/(2\epsilon + 1)$ . Each of the classes II, III, and IV includes two subgroups. In the class II there is a subclass IIp with the dominant PPC parameter (polarity/polarizability) or the Onsager function  $(n^2 - 1)/(n^2 + 1)$ . The solvents included in this group are primarily polarizable. The subclass IIb is formed by the solvents with dominant (within the class) PBC parameter (polarity/basicity), i.e. with predominant basic properties. An analogous situation is also seen in class III. In all the classes given the value of the PAC parameter (polarity/acidity) is roughly comparable. The opposite phenomenon is observed in the class IV whose subclasses primarily differ just in the PAC parameter. The solvents in IVa are aprotic, whereas



in IVp are protic. From the analysis carried out it is possible to suggest a new classification of solvents whose terminology and the solvents classified are given in Table IV. The first class given includes nonpolar inert solvents, aliphatic hydrocarbons. The class of nonpolar polarizable solvents involves aromatic hydrocarbons and substituted methanes with one type of atom. Ethers together with triethylamine form the class of nonpolar basic solvents containing free electron pairs. The group of little polar polarizable solvents includes aliphatic halogen derivatives and substituted benzenes with heteroatoms in substituents. The class of little polar basic solvents is formed by cyclic ethers, ketones, esters, and pyridine. The classification of phenetole (20) is questionable and can be due to the low frequency of this solvent in source data<sup>26</sup>. The group of polar aprotic solvents includes the solvents with heteroatoms (forming strongly polarized bonds) in their molecules, e.g. amides, nitriles, nitro compounds, sulfoxides, and sulfones. Alcohols and acetic acid form an unambiguous group of polar protic solvents. The classification given cannot accommodate solvents with exceptional properties such as water, glycol, and formamide which probably form strong hydrogen bonds (due either to easy structurization or to number of hydroxyl groups) and HMPA (due obviously to high basicity accompanied by high polarity). These solvents constitute a heterogeneous group of so-called exceptional solvents.

The clusters created by the method of the farthest neighbour differ but little from the classification given. A stable arrangement was found for nine clusters out of which six ones were identical with the classification according to Table IV, the class IVa was decomposed into two clusters (28–33–47–50 and 30–31–49 together with 32), and a special cluster was formed by water, formamide, and glycol.

TABLE IV

A new classification of solvents based on chemometrical empirical scale of parameters — the terminology and solvents included

Class	Name	Solvents included
I	Nonpolar–inert	1-2-3
IIp	Nonpolar–polarizable	4-5-6-7-8-9-51
IIb	Nonpolar–basic	16-17-18-45
IIIp	Little polar–polarizable	10-11-12-13-14-15-19-34-48
IIIb	Little polar–basic	20-21-22-23-24-25-26-27-46
IVa	Polar–aprotic	28-30-31-33-47-49-50
IVp	Polar–protic	36-37-38-39-40-41-43-44
V	Exceptional	29-32-35-42

With the help of the classification according to Table IV it was possible to calculate the informativeness of the individual parameters forming the space of the objects. This quantity makes it possible to find which of the parameters has the greatest contribution to the given classification. In our case the greatest differentiating power is exhibited by the PAC parameter (the information content 8.4), followed by PBC (7.3) and finally PPC (5.2). Hence the solvents should be classified according to their polarity and acidity, which is usually done (as it was mentioned in Introduction). This conclusion is in accordance with the average sensitivity given by the regression coefficients, which was found in the earlier paper<sup>26</sup>.

In conclusion it must be stressed that the classification found, which corresponds to the chemical structure of the compounds, is based solely on the experimental data concerning the solvent effect, the information about chemical structure being not input into the calculation at any phase. Thus the results presented form a piece of evidence for the relation (often presumed a priori) between the chemical structure of a solvent and its effect on the processes taking place therein.

#### REFERENCES

1. Reichardt C.: *Solvent and Solvent Effects in Organic Chemistry*. 2nd ed. VCH Verlagsgesellschaft mbH, Weinheim 1988.
2. Brönsted J. N., *Rec. Trav. Chim.* **42**, 718 (1923).
3. Brönsted J. N.: *Chem. Ber.* **61**, 2409 (1928).
4. Brönsted J. N.: *Z. Angew. Chem.* **43**, 229 (1930).
5. Parker A. J.: *J. Chem. Soc.* **1961**, 1328.
6. Parker A. J.: *Quart. Rev.* **16**, 162 (1962).
7. Parker A. J.: *Chem. Rev.* **69**, 1 (1969).
8. Janz G. J., Danyluck S. S.: *Chem. Rev.* **60**, 209 (1960).
9. Bagno A., Scorrano G.: *J. Am. Chem. Soc.* **110**, 4577 (1988).
10. Rice P. D., Bobbitt D. R.: *J. Chromatogr.* **437**, 3 (1988).
11. Krygowski T. M., Fawcett W. R.: *Can. J. Chem.* **54**, 3283 (1976).
12. Bohle M., Kollecker W., Martin D.: *Z. Chem.* **17**, 161 (1977).
13. Chastrette M.: *Tetrahedron* **35**, 1441 (1979).
14. Cramer R. D.: *J. Am. Chem. Soc.* **102**, 1837 (1980).
15. Cramer R. D.: *J. Am. Chem. Soc.* **102**, 1849 (1980).
16. Sjöström M., Wold S.: *J. Chem. Soc., Perkin Trans. 2*, **1981**, 104.
17. Chastrette M., Carretto J.: *Tetrahedron* **38**, 1615 (1982).
18. Svoboda P., Pytela O., Večeřa M.: *Collect. Czech. Chem. Commun.* **48**, 3287 (1983).
19. Swain C. G., Swain M. S., Powel A. L., Alunni S.: *J. Am. Chem. Soc.* **105**, 502 (1983).
20. Carlson R., Lunstedt T., Albano C.: *Acta Chem. Scand.*, B **39**, 79 (1985).
21. Elguero J., Fruchier A.: *Ann. Quim.*, C **79**, 72 (1983).
22. Chastrette M., Rajzmann M., Chanon M., Purcell K. F.: *J. Am. Chem. Soc.* **107**, 1 (1985).
23. Maria P.-C., Gal J.-F., de Franceschi J., Fargin E.: *J. Am. Chem. Soc.* **109**, 483 (1987).
24. Zalewski R. I., Kokocinska H., Reichardt C.: *J. Phys. Org. Chem.* **2**, 232 (1989).
25. Pytela O.: *Collect. Czech. Chem. Commun.* **54**, 136 (1989).
26. Pytela O.: *Collect. Czech. Chem. Commun.* **55**, 634 (1990).



27. Anderberg M. R.: *Cluster Analysis for Application*. Academic Press, New York 1973.
28. Jambu M.: *Classification automatique pour l'analyse des données*. Bordas, Paris 1978.
29. Hartigan J. A.: *Clustering Algorithms*. Wiley, New York 1981.
30. Ajvazjan S., Bežajevová Z., Staroverov O.: *Metody vícerozměrné analýzy*. SNTL, Prague 1981.
31. Lukasová A., Šarmanová J.: *Metody shlukové analýzy*. SNTL, Prague 1985.
32. Sneath P. H. A. in: *Numerical Taxonomy* (A. J. Cole, Ed.), p. 144. Academic Press, New York 1969.
33. Sörensen T.: *Biol. Scr.* 5, 1 (1948).
34. Sokal R. R., Michener C. D.: *Univ. Kansas Sci. Bull.* 38, 1409 (1958).

Translated by J. Panchartek.