Statistical Analysis of Solvatochromic Shift Data

By **Michael Sjöström** and **Svante Wold**,* Research Group for Chemometrics, Department of Chemistry, Umeå University, S–901 87 Umeå, Sweden

Principal component (p.c.) analysis combined with a cross-validation technique (c.v.) was used in the analysis of some data sets composed of solvatochromic shift data for non-hydrogen bonding indicators. The data were basically the same as earlier used by Kamlet *et al.* to define the π^* solvent polarity scale. The c.v. showed that to adequately describe the data a two-component p.c. model was needed and consequently two parameters θ_{1k} and θ_{2k} were determined for each solvent k. The performance of these solvent parameters was compared with the π^* scale.

FROM the theoretical point of view the nature of solvent effects on chemical processes is not well understood. At present, chemists therefore rely on empirically defined solvent scales to obtain a quantitative description of these effects.[1,2]

Some of the most widely used and extensive scales are the π^* scale of Kamlet *et al.*,[3,4] the E_T scale of Dimroth *et al.*,[5] the χ_R scale of Brooker *et al.*,[6] and the $\mathscr{S}$ scale of Lassau *et al.*[7-9] The π^* scale is defined by the solvatochromic shifts of a large number of indicators, mainly *m*- and *p*-*NN*-dialkylanilines. In contrast, the E_T and χ_R scales are each defined by the solvatochromic shifts of a single indicator [4-(2,4,6-triphenylpyridinium)-2,6-diphenylphenoxide and ' Brooker's merocyanine ', respectively] and the $\mathscr{S}$ scale is defined by the log k value of the Menschutkin reaction in different solvents.

In order to avoid anomalies due to hydrogen bonding interactions in the π^* scale, Kamlet *et al.*[3] have exluded hydrogen bond acceptor (HBA) solvents (*e.g.* acetone) in combination with hydrogen bond donor (HBD) indicators (*e.g. meta-* and *para*-substituted anilines). For hydrogen bond acceptor–donor (HBA–D) solvents (*e.g.* alcohols), measurements for indicators without HBA properties were used. Initially seven non-HBD indicators (*e.g. NN*-diethyl-4-nitroaniline) were used to calculate a primary π^* scale for 28 non-hydrogen-bonding (NHB) solvents (*e.g.* hexane and chlorinated aliphatic solvents). From the solvatochromic shift measurements ν_{ik} (index i for the indicators and index k for the solvents) the parameters ν_{i0}, s_i, and π_k^* were calculated with a round-robin least-squares program which minimized the residuals ε_{ik}. The model used was (1).

$$\nu_{ik} = \nu_{i0} + s_i \pi_k^* + \varepsilon_{ik} \qquad (1)$$

Using the ν_{i0} and s_i values, secondary π_k^* values for additional solvents were calculated as averages [equation (2)] of the π^*_{ik} values from equation (3).

$$\pi_k^* = \sum_{i=1}^{n} \pi_{ik}^*/n \qquad (2)$$

$$\pi_{ik}^* = (\nu_{ik} - \nu_{i0})/s_i \qquad (3)$$

In a recent article Kamlet *et al.*[4] have estimated ν_{i0} and s_i for nine indicators using known π_k^* values. From equations (2) and (3), π_k^* for 23 additional solvents were

determined and wherein the highest and lowest π_i^* values for a solvent were excluded in the average.

A model like equation (1) is a special case with $A = 1$ of the general equation (4).[10,11] However, with aid of

$$\nu_{ik} = \alpha_i + \sum_{a=1}^{A} \beta_{ia}\theta_{ak} + \varepsilon_{ik} \qquad (4)$$

principal component (p.c.) analysis [10-13] the parameters α_i, β_{ia}, and θ_{ak} can be determined in a statistically more efficient way than with the above method. Furthermore α_i, β_{ia}, and θ_{ak} can be determined from the measurements ν_{ik} for an arbitrary A. Such a data matrix, with M indicators and N solvents, from which a p.c. model is calculated is in the following denoted a training-set.

To avoid nonsense parameters it is crucial not to use more terms, A, than needed to account for the systematic information in the data. On the other hand, too few terms means that systematic information remains in the residuals. The appropriate number of terms A can, for example, be determined for a matrix without missing data with cross-validation (c.v.) technique.[14]

With the aim to extract the systematic information contained in in the solvatochromic shift data, we here wish to present the p.c.–c.v. analysis of some representative data sets, from which the π^* scale previously was defined by Kamlet *et al.* We have also investigated the correlations between the resulting θ_{ak} scales and the E_T χ_R and $\mathscr{S}$ scales.

For the calculations the statistical package SIMCA was used.[13,15] The package includes p.c. analysis, c.v. and classification routines, and the methods are well documented in the literature.[13-16] The data analysis follow the same procedure as the recent analysis of chemical reactivity data by Sjöström *et al.*[17]

DATA

With the aim of excluding possible hydrogen bond interactions between solvents and indicators we have, like Kamlet *et al.*,[3,4] introduced some restrictions on the data. Thus, in the training-set seven indicators of non-HBD type and 65 solvents of HBA and NHB type were used. Solvents of HBA–D type were put in the test-set in all investigated data sets. Since chloroform and acetonitrile are supposed to act as HBA–D solvents in some cases,[3] these solvents were placed in the test-set. Thus only HBA and NHB solvents

were used in the p.c.–c.v. analysis. The indicators are restricted to those with measurements for at least 50 solvents of HBA and NHB type. Only solvents with at least four measurements among the indicators (1)—(7) were chosen.

<div align="center">TABLE 1</div>

<div align="center">Analysed data sets</div>

Training sets	Indicators [a]	Solvents [b]
M1	(2), (3), (4), (6)	1—3, 5—29, 33, 34, 39, 42, 47—49, 51, 52, 56, 63, 79, 88—99, 130, 133—135
M2	(1), (2), (3), (6)	1, 2, 5—15, 17, 19—29, 33, 34, 37—39, 42—44, 48, 52, 56, 63, 79, 88—94, 96, 97, 99, 130, 134, 135
M3	(2), (3), (4), (6), (7)	1, 2, 5—7, 9—13, 16—29, 34, 39, 47, 48, 49, 51, 52, 56, 63, 79, 88—99, 130, 133, 134, 135
M4	(1)—(7)	1—3, 5—29, 33, 34, 37—44, 46—49, 51, 52, 56, 63, 79, 88—99, 130—135 [c]

[a] Indicator numbers, see Table 3. [b] Solvent numbers, see Table 4. [c] Missing data, indicator (1), solvents 3, 16, 18, 41, 46, 47, 49, 51, 95, 98, 131, and 133; (3), 41, 131, and 106, (4), 37, 38, 41, 43, 44 and 50; (5), 3, 16, 18, 37, 38, 41—44, 46, 50, 51, 132, 101—107, 109, 111—113, 201, and 202; (6) 46 and 132; (7), 3, 8, 14, 15, 33, 37, 38, 41, 42, 44, 50, 201, and 202.

The analysed data sets M1—M4 are given in Table 1, the solvents in Table 4, and the indicators in Table 2. No measurements have been excluded in the p.c.–c.v. analysis.

Classification of a Data Vector for a Solvent.—Once a p.c. model is determined for a training-set, a data vector v_{ip} for

$$v_{ip} - \alpha_i = \sum_{a=1}^{A} \beta_{ia}\theta_{ap} + \varepsilon_{ip} \qquad (5)$$

a solvent p from the training or test-set can be fitted by multiple regression to the p.c. model (α_i and β_{ia}).[13]

Depending on the size of the residual standard deviation s_p [equation (7)], such a data vector v_{ip} can be classified as

RESULTS

Since c.v. only operates on complete data sets, initially three data sets M1—M3 without missing data were analysed. The p.c.–c.v. analysis showed that two components models [*i.e.* $A = 2$ in equation (4)] were needed to correlate the data in M1—M3 (see Table 2). With the intention to calculate θ_{ak} values for as many solvents as possible a p.c. analysis was done on M4, a data set with 9% missing data. A two-component model was used since M4 contained all measurements of M1—M3 and was only expanded with measurements for five new solvents. In the p.c. model of M4, the $\beta_{i1}\theta_{1k}$ term describes 86% of the standard deviation around the means α_i. A minor part, 4%, is described by the $\beta_{i2}\theta_{2k}$ term and the remaining 10% is contained in the residuals. However, θ_2 is substantial for numerous solvents, such as acetone, methyl formate, acetonitrile, dibenzyl ether, brombenzene, *etc.* The resulting θ_{1k} and θ_{2k} values are presented together with the classification results in

<div align="center">TABLE 2</div>

R Values from the c.v. of M1—M3. Component A is significant as long as $R < R_{\rm crit.}$. $R_{\rm crit.}$ values are given in parentheses

Data set	Components A in equation (4)		
	$A = 1$	$A = 2$	$A = 3$
M1	0.01 (0.88)	0.36 (0.50)	1.85 (0.13)
M2	0.01 (0.88)	0.35 (0.50)	0.32 (0.13)
M3	0.02 (0.92)	0.46 (0.88)	0.83 (0.32)

Table 4 and the p.c. model of M4 is given in Table 3. The F values in Table 4 show that except for the solvents 22, 93, and 96 the data in the training-set is well described by the model of M4. The enhanced F values for most of the test-set solvents show that the data vectors for the HBA–D solvents are not well described by the model of M4. Among the exceptions are acetonitrile and chloroform and thus no deviating behaviour is detected for these solvents.

<div align="center">TABLE 3</div>

The p.c. model from M4. The residual standard deviations for each indicator are compared with an F test when the measurements are fitted to the θ_{ak} ($s_{i\theta}$) and π^* ($s_{i\pi}$) scales

Indicator [a]	α_i	β_{i1}	β_{i2}	$s_{i\theta}$	$s_{i\pi}$	$F = s_{i\theta}^2/s_{i\pi}^2$
4-Nitroanisole (1)	32.67	0.346	−0.078 [b]	0.078	0.081	(1.1) [b]
NN-Diethyl-3-nitroaniline	24.14	0.326	−0.116 [b]	0.072	0.069	(0.9) [b]
4-Methoxy-β-nitrostyrene (3)	28.56	0.338	0.635	0.052	0.099	3.6
NN-Diethyl-4-nitroaniline (4)	25.53	0.462	0.324	0.046	0.091	2.5
NN-Dimethyl-4-aminobenzophenone (5)	29.16	0.301	−0.500	0.069	0.094	1.9
NN-3,5-Tetramethyl-4-nitroaniline (6)	25.65	0.398	0.285	0.061	0.096	2.6
NN-Diethyl-3-methyl-4-nitroaniline (7)	25.76	0.444	0.376	0.043	0.095	4.9

[a] Data from refs. 3, 4, 21, 24, and 25 and personal communication from Dr M. J. Kamlet. [b] Not significant at the 95% level.

showing normal behaviour or not. This is made with an F test by comparing s_p with the overall standard deviation S_0

$$F = s_p^2/S_0^2 \qquad (6)$$

$$s_p^2 = \sum_{i=1}^{M} \varepsilon_{ip}^2/(M - A) \qquad (7)$$

$$S_0^2 = \sum_{i=1}^{M}\sum_{k=1}^{N} \varepsilon_{ik}^2/(N - A - 1)(M - A) \qquad (8)$$

of the data vectors in the training-set. Notice that equation (5) also can be used to determine secondary θ_{ak} values for solvents in the test-set.

The performance of the π^* and θ_{ak} scales is compared in Tables 3 and 5. In Table 5 the solvatochromic shift data for some further indicators including Dimroth's betaine ($E_{\rm T}$) and Brooker's merocyanine ($\chi_{\rm R}$) are fitted to the π^* and θ_{ak} scales. Analysed in the same way are the log k values for the Menschutkin reaction between tripropylamine and methyl iodide ($\mathscr{S}$), measured in a large number of solvents. For the indicators (1), (2), (11), (14), and (15) the fit to θ_{ak} is equal or slightly better than to the π^* scale. Except for (14), this result is expected since θ_{2k} is not significant for (1), (2), (11), and (15) and since θ_{1k} and π^* are highly correlated ($r = 0.997$, $n = 65$). For all other in-

dicators and the Menschutkin reaction the θ_{ak} values give a significantly better fit than the π^* values (see Tables 3 and 5).

DISCUSSION

From the results it is apparent that two parameters θ_{1k} and θ_{2k} for each solvent of the HBA and NHB types are needed to describe the systematic change in the

TABLE 4

Resulting θ_{1k} and θ_{2k} values from M4

No.[a] Solvent		θ_{1k}	θ_{2k}	F[b]
Training-set				
1	n-Hexane	4.53	0.032	2.8
2	Cyclohexane	4.03	0.157	0.8
3	Triethylamine	3.39	0.059	0.6
5	Di-n-butyl ether	2.82	−0.003	0.8
6	Carbon tetrachloride	2.37	0.170	1.3
7	Diethyl ether	2.33	−0.189	3.1
8	Toluene	0.716	0.040	0.8
9	Dioxan	0.529	−0.057	0.6
10	Trichloroethylene	0.601	−0.020	0.1
11	Ethyl acetate	0.590	−0.188	1.1
12	1,1,1-Trichloroethane	0.864	−0.057	0.6
13	Tetrahydrofuran	0.363	−0.063	1.1
14	Benzene	0.295	0.033	1.0
15	Chlorobenzene	−0.476	0.179	0.9
16	Butan-2-one	−0.353	−0.191	0.1
17	Anisole	−0.645	0.200	0.6
18	Acetone	−0.457	−0.310	0.1
19	Triethyl phosphate	−0.759	−0.145	0.4
20	1,2-Dichloroethane	−1.08	−0.008	1.2
21	Methylene chloride	−1.25	−0.029	0.4
22	1,1,2-Trichloroethane	(−1.43)[c]	(0.090)[c]	(4.0)
23	Dimethylacetamide	−1.73	−0.060	0.2
24	Pyridine	−1.74	0.079	2.3
25	Dimethylformamide	−1.86	−0.074	0.3
26	Hexamethylphosphoramide	−1.78	−0.041	2.7
27	Butyrolactone	−1.98	−0.105	0.6
28	N-Methylpyrrolidone	−2.11	0.121	1.0
29	Dimethyl sulphoxide	−2.70	0.057	0.3
33	Bromobenzene	−0.916	0.268	0.6
34	Acetic anhydride	−0.910	−0.238	1.0
37	Benzonitrile	−1.86	0.076	0.0
38	Butyl acetate	0.909	−0.081	0.3
39	Ethyl chloroacetate	−0.621	−0.072	1.6
40	Tetrahydropyran	0.688	−0.140	1.7
41	Cyclohexanone	−0.832	0.036	0.3
42	Tri-n-butyl phosphate	−0.236	−0.068	0.4
43	Tetrachloroethylene	2.36	0.152	0.1
44	1,1,2,2-Tetrachloroethane	−2.09	0.209	0.7
46	Dibenzyl ether	−1.22	0.405	0.6
47	Ethyl benzoate	−0.673	0.181	0.4
48	Tri-n-butylamine	3.46	0.161	1.2
49	NN-Dimethylbenzylamine	0.977	0.266	0.9
51	Cyclopentanone	−0.889	−0.013	0.3
52	Methyl acetate	0.578	−0.258	0.1
56	Sulpholan	−2.41	−0.060	0.6
63	Phenylacetonitrile	−2.30	0.118	0.6
79	2,6-Lutidine	−0.242	0.105	0.8
88	Bis-(2-methoxyethyl) ether	−0.232	−0.105	0.9
89	Butyronitrile	−0.576	−0.260	0.6
90	Cumene	1.46	0.110	0.6
91	1,2-Dibromoethane	−0.823	0.188	0.3
92	m-Dichlorobenzene	−0.314	0.239	0.4
93	trans-1,2-Dichloroethylene	(1.18)	(−0.064)	(8.2)
94	Diethyl malonate	−0.179	−0.110	0.2
95	Ethyl acetoacetate	−0.923	−0.126	0.5
96	Ethyl trichloroacetate	(0.061)	(0.103)	(5.6)
97	Fluorobenzene	−0.067	0.029	1.4
98	Heptan-3-one	0.137	−0.187	0.1
99	Methyl acetate	0.040	−0.298	0.7
130	Dimethoxymethane	1.81	−0.185	0.2
131	Phenylacetone	−1.72	0.024	0.2
132	Pentachloroethane	−0.004	0.045	0.3
133	Trimethyl orthoacetate	1.90	−0.053	0.8
134	Trimethyl orthoformate	0.263	−0.198	0.5
135	m-Xylene	1.16	0.121	2.3

TABLE 4 (continued)

Test-set				
30	Chloroform	−0.956	−0.070	1.4
50	Acetonitrile	−0.851	−0.385	0.7
101	t-Butyl alcohol	(0.310)	(−0.199)	(5.3)
102	isobutyl alcohol	(0.179)	(−0.373)	(7.8)
103	n-Butanol	(0.210)	(−0.510)	(8.7)
104	Ethanol	(0.056)	(−0.392)	(4.4)
105	Methanol	(−0.427)	(−0.549)	(5.8)
106	2-Phenylethanol	(−1.72)	(−0.090)	(7.3)
107	Ethylene glycol	(−2.51)	(−0.281)	(5.3)
109	Benzyl alcohol	−2.58	0.013	2.1
111	Water	(−3.91)	(−0.961)	(8.9)
112	n-Propanol	(−0.033)	(0.394)	(6.4)
113	Trifluoroethanol	−3.31	−0.872	2.4
201	Acetic acid	0.345	−0.590	2.6
202	Formamide	−3.54	−0.263	3.6

[a] Numbers for the solvents are those used as refs. 3 and 4. [b] F Values according to equation (6). $F_{calc.}$ is given in parentheses if $F_{calc.} > F_{crit.}$. [c] θ-Values in parentheses are less reliable due to a poor fit to the M4 model and should be used with caution

solvatochromic shifts for the studied indicators of non-HBD type. We have also shown that these two θ_{ak} scales better describe solvatochromic shifts and chemical reactivity than the single π^* scale. This comparision was made on data sets which were not included in the estimation of the θ scales. Furthermore, for solvents of the HBA and NHB types it was found that the θ scales well could describe the π^*, X_R, E_T, and $\mathscr{S}$ polarity scales. This implies that the number of solvent scales for HBA and NHB solvents can be drastically reduced without loss of information and that the data base for the θ scales can be expanded to include measurements from the E_T, X_R, $\mathscr{S}$, and probably numerous other solvent scales.

Empirical solvent scales like π^* (or θ_{1k}), E_T, X_R, and $\mathscr{S}$ are often interpreted as caused by a single effect, the polarity of the solvent. However, it is noteworthy that if a general set of solvents are studied, correlations between empirical polarity measures and physical properties of the solvents [for example the dielectric constant ε or $(\varepsilon - 1)/(2\varepsilon + 1)$] are too poor to be of predictive value.[1,2] This is also the case for measures calculated from classical theories.[1] This means that that we cannot eliminate the possiblity that a large number of properties of the solvents combine in a complex manner to form such parameters.

The θ_{2k} values on the other hand seem to be related to the polarizability of the solvents. This can be seen from the two correlations in Figure 1 where a linear combination of θ_{1k} and θ_{2k} is highly correlated with the refractive index n_D^{20}. In these correlations the θ_{1k} vector makes only a small contribution. The polarizability property also seem to play an important part for the HBA-D solvents even if some of these show a somewhat different trend from the HBA and NHB solvents.

An extension of equation (1) was introduced by Kamlet

$$v_{ik} = v_{i0} + s_i'(\pi_k^* + d\delta) \qquad (9)$$

et al.[3,4] to compensate for the polarizability of the solvents and was used to correlate the E_T and $\mathscr{S}$ scales and some other reactivity data. In model (9) δ is a polarizability

correction factor, chosen to be 0.0 for all non-chlorinated aliphatic, solvent, 0.5 for polychlorinated aliphatic solvents, and 1.0 for aromatic compounds. Since the δ values reflect the correct trend of the behaviour of the θ_2

Also a subclassification will not give p.c. models of lower complexity (*i.e.* $A = 0$ or 1). This is so because neither θ_{1k} or θ_{2k} are constant nor are linearly related to each other within these classes (see Figure 2). However, such

TABLE 5

A comparision of the correlations when spectral data for some indicators and the Menschutkin reaction in different solvents are fitted to the π_k^* and θ_{ak} scales

Reaction or indicator	n^a	$v_k = a + b\,\pi_k^*$			$v_k = a' + b'\theta_{1k} + b''\theta_{2k}$				$F = \dfrac{s_{t\pi}^2}{s_{t\theta}^2}$	Ref.
		a	b	$s_{e\pi}$ (r)	a'	b'	b''	$s_{e\theta}$ (r)		
$\log k_k$ [N(C$_3$H$_7$)$_3$ + CH$_3$I] (8)	41	−4.22	4.49	0.21 (0.977) [b]	−1.40	−0.67	−0.76	0.13 (0.991)	2.6	7, 9
Dimroth's betaine (E_T) (9)	43	30.00	13.72	2.15 (0.850) [c]	38.36	−1.95	−10.43	1.32 (0.948)	2.7	5
NN-Dimethyl-4-nitroaniline [d] (10)	40	28.10	−3.43	0.12 (0 990)	25.99	0.51	0.44	0.08 (0.995)	2.0	4, 26
N-(4-Nitrophenyl)aziridine [d] (11)	17	32.12	−2.54	0.11 (0.985)	30.54	0.38	e	0.10 (0.989)	(1.2) [e]	23
N-(4-Nitrophenyl)pyrrolidine (12)	18	27.55	−3.30	0.13 (0.988)	25.54	0.48	0.48	0.07 (0.996)	3.2	23
N-(4-Nitrophenyl)piperidine (13)	18	27.92	−3.41	0.12 (0.990)	25.83	0.50	0.36	0.07 (0.997)	3.1	23
4,4′-Bis(dimethylamino)benzophenone (14)	21	30.08	−2.23	0.07 (0.986)	28.67	0.32	−0.36	0.06 (0.989)	(1.2) [e]	4
Brooker's merocyanine [d] (15)	27	50.71	−7.97	0.58 (0.972)	45.76	1.17	e	0.53 (0.979)	(1.2) [e]	6

[a] Solvents from the training-set (see Table 4) plus acetonitrile and chloroform if measurements were available. [b] Equation (9) gives s_e 0.15, r 0.989. [c] Equation (9) gives s_e 1.44, r 0.938. [d] Ethyl acetoacetate is excluded for indicator (10), chloroform for (11), and dioxan for (15). The residuals are larger than 2.5 s_e when included in the regressions. [e] Not significant on the 95% level.

values (see Figure 2) equation (9) improves the correlations of the E_T and $\mathscr{S}$ scales. However, the θ_1 and θ_2 scales still give a better description of the E_T and $\mathscr{S}$ scales (see Table 5, footnotes b and c).

When the E_T or $\mathscr{S}$ scales are fitted to the π^* scale the fit can be improved by a division of the data set into the

a division of a data set into separate classes can be justified when a data set is not adequately described by a simple class model and if it is simultanously possible to find a classification in which separate p.c. models for each class describe the data better.
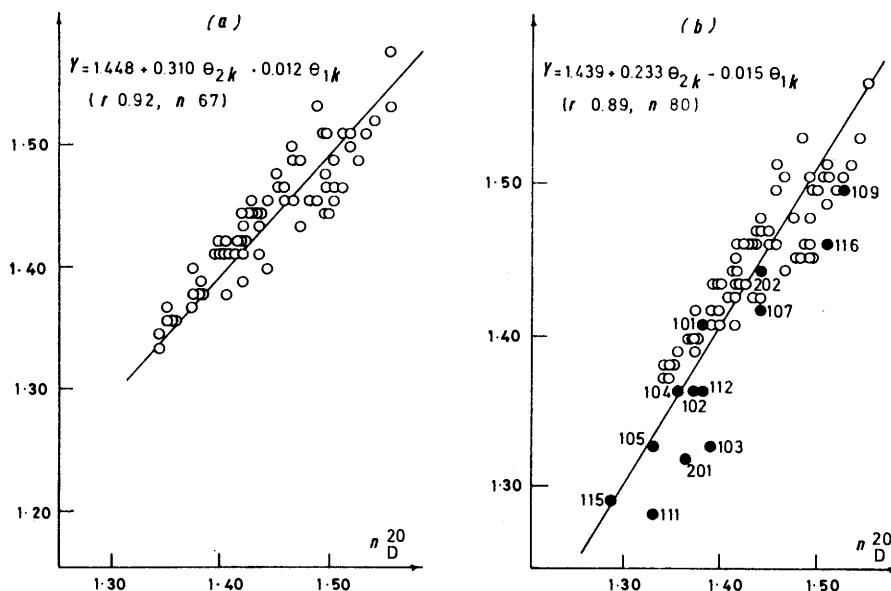
A division of the data set has been done in this in-

FIGURE 1   (a) Linear combination of θ_{1k} and θ_{2k} for the HBA and NHB solvents is highly correlated with the refractive index n_D^{20} (r 0.92, n 67). Acetonitrile and chloroform are included in the correlation. (b) If also the HBA–D solvents are considered the correlation is similar (r 0.89, n 80). However, some of the HBA–D solvents show a somewhat different trend. The correlation coefficients are not improved if the transformation $(n_D^2 - 1)/(n_D^2 + 2)$ is used instead of n_D. The reason is the high correlation (r 0.9998) between n_D and $(n_D^2 - 1)/(n_D^2 + 2)$ for the solvents considered (see also ref. 2)

three subclasses mentioned.[3,4] This improvement of the fit for the separate classes can be explained from the present study as caused by a smaller variation in θ_{2k} within these classes (see Figure 2). This naturally improves the fit if a one parameter model is used. For the M4 data set such a division into three classes will be of less value since the data is already well described by a single model.

vestigation, in that the HBA–D solvents have been excluded in the p.c. analysis. The reason is that most of the HBA–D solvents are not well described by the same p.c. model as the HBA and NHB solvents, as seen from the classification results in Table 4. From the classification results for acetonitrile and chloroform the conclusion is that these solvents show the same behaviour

as the HBA and NHB solvents. The poor classification of most of the HBA–D solvents show that either the non-systematic information in their measurements is larger than for the HBA and NHB solvents or that a more complex model is needed to treat the HBA–D solvents together with the HBA and NHB solvents. The θ_{ak} values for the HBA–D solvents also seem to have limited utility compared with those for the HBA and NHB solvents. For example when the HBA–D solvents are included in the regressions of the E_T and the Menschutkin reaction, the residual standard deviations increases with a factor two and three. In the case of the Menschutkin reaction
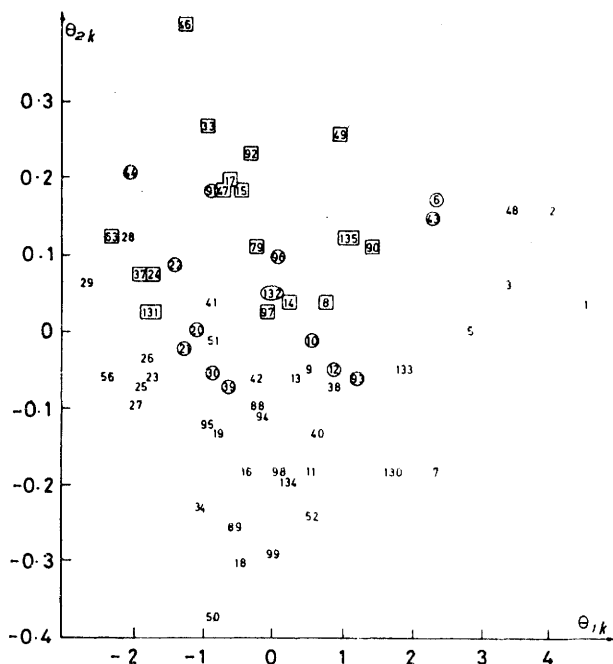
FIGURE 2 Plot of θ_{1k} against θ_{2k} for the training-set solvents plus acetonitrile and chloroform. The aromatic and chlorinated aliphatic solvents are characterised by □ and ○, respectively

this is expected since the protic solvents follow another mechanism than the non-protic.[18] In the case of the E_T scale, hydrogen-bond interactions between the solvent and indicator have been proposed.[3]

To deal with solvatochromic shifts in cases where the solvents are of the HBA–D type and the indicators of HBA type extensions of equations (1) and (9) have been formulated.[19,20] In equations (10) and (11) α is the hydrogen-bond donor acidity for the HBA–D solvents. However, it has been found that non-chlorinated alipha-

$$\nu_{ik} = \nu_{i0} + s_i\pi_k^* + a_i\alpha_k \qquad (10)$$

$$\nu_{ik} = \nu_{i0} + s_i'(\pi_k^* + d\delta) + a_i\alpha_k \qquad (11)$$

tic, chlorinated aliphatic, and aromatic solvents had to be treated separately in equation (10) in correlations with the E_T scale.[20] Equation (11) as an extension of equation (9) has not been further investigated. From these findings it is evident that a complex treatment is needed to force the HBA–D solvents into the same model as the

HBA and NHB solvents. This indicates that the HBA–D solvents should better be reated with a separate p.c. model.

To correlate solvatochromic shift data when the solvents are of HBA and NHB type and the indicators are of HBD type another extension (12) of equation (1) was used .[21,22] Here the β_k values are introduced to

$$\nu_{ik} = \nu_{i0} + s_i\pi_k^* + b_i\beta_k \qquad (12)$$

express the hydrogen-bond acceptor ability of the HBA solvents. A preliminary p.c.–c.v., reveals, however, that a three-component model is needed if the M4 data set is expanded to contain measurements for HBD indicators.

The α and β scales in equations (10) and (12) were estimated with the so-called solvatochromic comparison method.[19,21] Briefly, with this method the difference is calculated between observed and theoretical ν_k values, the latter calculated assuming no HBA (or (NHB) properties. Averages of these shift differences are then used to calculate the α and β scales. From the present investigation it is evident that with p.c. analysis such parameter scales can be calculated directly from the raw data. Equally important with c.v., it is possible to determine effectively the complexity of the model needed to correlate a data set.

REFERENCES

[1] C. Reichardt, ' Solvent Effects in Organic Chemistry ', Verlag Chemie, Weinheim, 1979.
[2] I. A. Koppel and V. A. Palm in ' Linear Free Energy Relationships ', eds. N. B. Chapman and J. Shorter, Plenum Press, London, 1972.
[3] M. J. Kamlet, J. L. Abboud, and R. W. Taft, J. Amer. Chem. Soc., 1977, 99, 6027.
[4] M. J. Kamlet, T. N. Hall, J. Boykin, and R. W. Taft, J. Org. Chem., 1979, 44, 2599.
[5] C. Reichardt, Angew. Chem. Internat. Edn., 1979, 18, 98.
[6] L. G. S. Brooker, A. C. Craig, D. W. Heseltine, P. W. Jenkins, and L. L. Lincoln, J. Amer. Chem. Soc., 1965, 87, 2443.
[7] C. Lassau and J.-C. Jungers, Bull. Soc. chim. France, 1968, 2678.
[8] Y. Drougard and D. Decroocq, Bull. Soc. chim. France, 1968, 2972.
[9] G. Berrebi and D. Decroocq, J. chim. Phys., 1974, 71, 675.
[10] S. Wold and M. Sjöström in ' Correlation Analysis in Chemistry: Recent Advances ', eds. J. Shorter and N. B. Chapman, Plenum Press, London, 1978.
[11] S. Wold, Chemica Scripta, 1973, 6, 97.
[12] S. Wold and M. Sjöström, Chemica Scripta, 1972, 2, 49.
[13] S. Wold and M. Sjöström in ' Chemometrics, Theory and Practice ', ACS Symposium Series No 52, ed. B. R. Kowalski, Amer Chem. Soc., Washington, 1977.
[14] S. Wold, Technometrics, 1978, 20, 397.
[15] S. Wold, SIMCA-2T manual, available from S. Wold.
[16] S. Wold, Pattern Recognition, 1976, 8, 127.
[17] M. Sjöström and S. Wold, J.C.S. Perkin II, 1979, 1274.
[18] M. Auriel and E. de Hoffmann, J. Amer. Chem. Soc., 1975, 96, 7433.
[19] R. W. Taft and M. J. Kamlet, J. Amer. Chem. Soc., 1976, 98, 2886.
[20] R. W. Taft and M. J. Kamlet, J.C.S. Perkin II, 1979, 1723.

[21] M. J. Kamlet and R. W. Taft, *J. Amer. Chem. Soc.*, 1976, **98**, 377.

[22] M. J. Kamlet, M. E. Jones, R. W. Taft, and J.-L. Abboud, *J.C.S. Perkin II*, 1979, 342.

[23] M. J. Kamlet, R. R. Minsinger, E. G. Kayser, M. H. Aldrige, and J. W. Eastes, *J. Org. Chem. Soc.*, 1971, **25**, 3852.

[24] R. R. Minesinger, M. E. Jones, R. W. Taft, and M. J. Kamlet, *J. Org. Chem.*, 1977, **42**, 1929.

[25] M. J. Kamlet, E. G. Kayser, J. W. Eastes, and W. H. Gillingan, *J. Amer. Chem. Soc.*, 1973, **95**, 5210.

[26] T. Yokoyama, *Austral. J. Chem.*, 1976, **29**, 1469.