Simultaneous Description of the Influence of Solvent, Reaction Type, and Substituent on Equilibrium Constants by Means of Three-mode Factor Analysis

Martien C. Spanjer* and C. Leo de Ligny*
Laboratory for Analytical Chemistry, University of Utrecht, Croesestraat 77A, 3522 AD Utrecht, The Netherlands
Hans C. van Houwelingen and Jeroen M. Weesie
Institute for Mathematical Statistics, University of Utrecht, Budapestlaan 6, 3584 CD Utrecht, The Netherlands

The equations, proposed by Hammett, Taft, and Nieuwdorp *et al.*, respectively, for the simultaneous description of the influence of reaction type and substituent on equilibrium and reaction rate constants are discussed. The latter equation represents an example of factor analysis. This mathematical–statistical technique has also been applied to describe simultaneously the influence of solvent and substituent and the influence of solvent and reaction type. It is thus a logical step to classify equilibrium and reaction rate constants with respect to three modes, solvent, reaction type, and substituent, and to try to describe the influence of these three variables by three-mode factor analysis. Two examples of the application of this technique to literature data are given. The first concerns data on ionization constants for 15 series of substituted compounds in three solvents. The second example concerns data on phase equilibrium constants of six series of substituted compounds in nine two-phase systems. The two-phase systems comprise gas–liquid as well as liquid–liquid and solid–liquid systems. The precision of the fit of the observations and the precision of the prediction of the missing data are discussed. In the first example 237 data are missing. Among them are 90 data that cannot be predicted at all by the Hammett, Taft, or Nieuwdorp equations (*viz.*, for reactions on which no measurements at all are available in a particular solvent). The standard deviation of the prediction of the latter data by three-mode factor analysis ranges from 0.1 to 0.2. In the second example nearly all missing data are for reactions on which no measurements at all are available in a particular solvent. They can be predicted by three-mode factor analysis with a standard deviation that ranges from 0.09 to 0.13. Further, it is shown that the number of parameters that is required to fit the observations by three-mode factor analysis is far less than the number of parameters in the corresponding regression analysis model, *viz.*, the Taft model.

Theory

The description of the influence of the solvent on equilibrium and reaction rate constants is an important problem in chemistry, especially in physical organic chemistry. A satisfactory solution of this problem has not yet been obtained. In this paper we present a new approach, and illustrate its merits by applying it to data on chemical equilibrium constants of 15 series of substituted aromatic compounds in three solvents, and to data on phase equilibrium constants of six series of substituted aromatic compounds in nine widely different two-phase systems.

Our approach is an extension of the work by Nieuwdorp *et al.*[1] on the simultaneous description of the influence of the reaction type and substituents on equilibrium and reaction rate constants. It leads to the simultaneous description of the influence of solvent, reaction type, and substituents. Nieuwdorp *et al.*'s study, in its turn, is an extension of the well known Hammett[2] and Taft[3] approaches.

The Hammett relationship is given in equation (1). The

$$\log(K_X/K_H) = \rho\sigma_{m,p} \qquad (1)$$

subscripts X and H refer to the substituted and the unsubstituted compounds, respectively, ρ is a parameter that depends on the reaction type, and σ is a variable that depends on the substituent and its position (*i.e.*, whether it is *m* or *p*). It was soon recognized that the Hammett relation could describe substituent effects in only a relatively small area of reaction types. To widen the scope of applications Taft proposed equation (2). The subscripts *I* and *R* refer to the inductive effect and the resonance effect of a substituent, respectively.

$$\log(K_X/K_H) = \rho_I\sigma_I + \rho_R\sigma_R \qquad (2)$$

For each substituent only one σ_I value is given (independent of its position), but the σ_R value depends on the type of reaction. In a review of Ehrenson *et al.*[4] four types of reactions, and thus four sets of σ_R values, were proposed: the well known σ_R^0, σ_R (benzoic acids), σ_R^-, and σ_R^+ sets.

Nieuwdorp *et al.* showed that data on the four reaction types that correspond with the above mentioned sets of σ_R values can be described with universal σ values when a third term is added to the right-hand side of equation (2). In this equation each substituent has only one, universal σ_I, σ_R, and σ_E value, that is independent of its position. The subscript *E* refers to the direct (through, exalted) resonance effect of a substituent on the reaction centre. As in the Taft equation, the ρ parameters have different values for *m*- and *p*-substituted series. The wide scope of applicability of equation (3) has been demonstrated recently.[5]

$$\log(K_X/K_H) = \rho_I\sigma_I + \rho_R\sigma_R + \rho_E\sigma_E \qquad (3)$$

Further, Nieuwdorp *et al.* stressed the point that, in principle, in equations (1)—(3) both σ and ρ are adjustable parameters. Previously, in the application of equations (1) and (2) reference reaction series were chosen to define a scale of σ values. *E.g.*, the σ values in the Hammett equation (1) are defined by the ionization constants of benzoic acids in water at 25 °C, for which, by definition, $\rho \equiv 1$. Thus, the σ values in equation (1), and the σ_I and σ_R values in equation (2), are considered to be known *a priori*, *i.e.*, σ is considered as a variable instead of as a parameter, and equations (1) and (2) are applied in regression analysis. Nieuwdorp *et al.*, on the other hand, considering both σ and ρ as adjustable parameters, determined their optimal values by factor analysis[6] of 570 data on 76 reactions and 17 substituents. To do so, a procedure had to be developed for the application of factor analysis on data sets in which a large fraction of the data is missing.[7] [Of course, once optimal values of σ_I, σ_R, and σ_E have been calculated, they can be used in conjunction with equation (3) for regression analysis of other data.]

In the language of factor analysis, equation (3) can be reformulated as (4). The symbol y stands for $\log(K_X/K_H)$, r' and s

$$y_{r',s} = \sum_{l=1}^{w} R_{r',l}\, S_{s,l} \qquad (4)$$

denote the reaction and the substituent, respectively, and R and S are adjustable parameters, dependent on the reaction and the substituent, respectively. It is the purpose of factor analysis to give a precise description of the data y with a small number w of terms. For Nieuwdorp *et al.*'s data, three terms were sufficient to reduce the standard deviation to 0.03.

In equation (4) reactions of different type (*e.g.*, the ionization of phenols and of benzoic acids in water) and reactions of the same type in different solvents (*e.g.*, the ionization of phenols in water and in 50% ethanol) are all considered as 'different' reactions. In other words, r' in equation (4) is a combination mode of reaction type and solvent. If we denote the reaction type by r and the solvent (or, for a phase-equilibrium constant, the two-phase system) by p, equation (4) can be reformulated as (5). Then, it is a logical step to classify the data y in three modes

$$y_{(p,r),s} = \sum_{l=1}^{w} R_{(p,r),l}\, S_{s,l} \qquad (5)$$

(with respect to solvent, reaction type, and substituent, respectively) and to try to describe them by three-mode factor analysis. Extension of (5) to three modes is formulated in (6). The symbol P represents a solvent-dependent parameter and c

$$y_{p,r,s} = \sum_{j=1}^{u} \sum_{k=1}^{v} \sum_{l=1}^{w} c_{j,k,l}\, P_{p,j}\, R_{r,k}\, S_{s,l} \qquad (6)$$

represents the elements of the so-called three-mode core matrix of scaling constants. As far as we know, equation (6) has only been applied in the field of the social sciences, and for the case where data exist for each combination of p, and r, and s.[8] Two of the present authors have recently devised a method to estimate $P_{p,j}$, $R_{r,k}$, and $S_{s,l}$ for the case of missing data,[9] and the present investigation is one of the first applications of this method in the field of the natural sciences.

We believe that it is obvious to try the merits of equation (6) for the simultaneous description of the influence of solvent, reaction type, and substituent on equilibria and reaction rates, for the following three reasons. (1) The influence of reaction type and substituent on $y_{r,s}$ can be described simultaneously by two-mode factor analysis, *i.e.*, by equation (5), with three factors.[1] (2) The influence of solvent and substituent on $y_{p,s}$ can also be described simultaneously by two-mode factor analysis, *i.e.*, by equation (7). This was demonstrated by Weiner[10] for data on

$$y_{p,s} = \sum_{j=1}^{u} P_{p,j}\, S_{s,j} \qquad (7)$$

the ionization of substituted benzoic acids in a variety of solvents; four factors were required to reproduce the data within the experimental precision of 0.01. (3) The combined influence of solvent and reaction type on $y_{p,r}$ has also been described by two-mode factor analysis, *i.e.*, by equation (8), for data on the

$$y_{p,r} = \sum_{j=1}^{u} P_{p,j}\, R_{r,j} \qquad (8)$$

solvent dependence of the n.m.r. chemical shift of ^{1}H, ^{13}C, and ^{29}Si nuclei in 36 solvents.[11] Two factors were sufficient to reproduce the data within experimental error. Thus, it is logical to combine equations (5), (7), and (8) to obtain equation (6). It may be expected that not more than four factors for each mode are required to describe adequately the influence of solvent, reaction type, and substituent.

It should be realized that the same data set (*i.e.*, data on substituent effects on a number of reactions in a number of solvents) may be either described by three-mode factor analysis, equation (6), or by two-mode factor analysis, equation (5), or, if optimal values of the parameters $S_{s,l}$ have been estimated from other data, by regression analysis. In regression analysis, equation (5) is also used, but now w is fixed and $S_{s,l}$ is a variable. Roughly speaking, the difference between equation (5), whether used in factor analysis or in regression analysis, and equation (6) is that the parameters $R_{(p,r),l}$ in equation (5) are freely adjustable, whereas in equation (6) the constraint (9) is posed on their values.

$$R_{(p,r),l} = \sum_{j=1}^{u} \sum_{k=1}^{v} c_{j,k,l}\, P_{p,j}\, R_{r,k}, \text{ for } l = 1\text{—}w \qquad (9)$$

Thus it can be expected that the residual standard deviation with equation (5) is smaller than with equation (6). What then is the advantage of equation (6)? It is twofold. (1) More important than the standard deviation $\hat{\sigma}$ is the length of the prediction interval $\hat{l}$. This length is equal to $2t\hat{\sigma}$, where the factor t increases with decreasing number of degrees of freedom DF. With equation (5), used in regression analysis, DF is equal to the number of observations minus the number of estimated parameters for a particular reaction in a particular solvent. Thus, DF may be quite small, and accordingly t may be rather large. With equation (6), DF is equal to the number of observations minus the number of estimated parameters in the whole data set, and this is likely to be a large number. Thus, even if $\hat{\sigma}$ is smaller with equation (5), $\hat{l}$ may be smaller with equation (6). (2) If no observations are available in a particular solvent, they cannot be predicted with equation (5), whether used in factor analysis or in regression analysis. On the other hand, equation (6) offers the possibility to predict these missing data as well. An example will be worked out in the Discussion.

Data

We shall apply equation (6) to two cases.

(1) The first comprises data on chemical equilibria, *i.e.*, ionization constants of 15 series of substituted aromatic compounds in three (admittedly rather closely related) solvents. These data are given in Table 1. (2) The second includes data on phase equilibria of six series of substituted aromatic compounds, *i.e.*, net retention volumes in reversed-phase h.p.l.c. and in gas liquid chromatography and partition coefficients in several widely different two-phase systems. These data are given in Table 2. It has been proved previously[12] that data on the influence of substituents on phase equilibria are, in principle, equivalent with data on chemical equilibria, when a correction is applied for the influence of the substituent itself on the phase equilibrium constant. In other words, the left-hand side of equations (1)—(3) reads for phase equilibria as in (10) where X

$$\Delta\Delta \log K \equiv \log(K_{X\bullet Y}/K_{H\bullet Y}) - \log(K_{X\bullet H}/K_{H\bullet H}) \qquad (10)$$

is a variable substituent, φ a common hydrocarbon skeleton, and Y a common reaction centre. It has been demonstrated also that data on liquid–liquid partition equilibria[13] and retention data in gas chromatography[14] and in normal-phase[15] and reversed-phase[16] liquid chromatography can be fitted excellently by equations (1), (2), or (3).

Results and Discussion

Case 1.—The results of three-mode factor analysis with missing data,[9] applied to the data in Table 1c, are presented in Table 3. The results of an analysis of variance, in which only the main effects of the solvents, reaction types, and substituents (but not their interactions) were taken into account, are also presented. An analysis of variance with first-order interactions is impossible because the number of parameters in the model (198) approaches the number of observations (213).

It follows from the penultimate column of Table 3 that the present data can be fitted better by the factor analysis model (even with only one factor for each mode) than by an (additive) analysis of variance model. With two factors for each mode the residual standard deviation is 0.10 and this figure can be improved to 0.07 by introduction of a third factor for the reaction types and the substituents.

However, the ability of a statistical model to fit observations is not its most useful property. Far more important is its ability

Table 1. (a) Investigated solvents

Code	Solvent
1	water, 25 °C
2	50 v/v % ethanol-water, 20—25 °C
3	95 v/v % ethanol-water, 20—25 °C

Table 1. (b) Investigated reactions (ionization of the tabulated compounds)

Code	Compounds	References
1	*trans*-4-X-cyclohexanecarboxylic acids	17
2	*m*-X-phenylacetic acids	18—21
3	*p*-X-phenylacetic acids	18—21
4	*p*-X-β-phenylpropionic acids	22,23
5	*m*-X-benzoic acids	18—20, 24—28
6	*p*-X-benzoic acids	18—20, 24—28
7	*p*-X-benzenephosphonic acids, first step	29
8	*p*-X-benzenephosphonic acids, second step	29
9	*m*-X-phenols	30
10	*p*-X-phenols	30
11	*m*-X-thiophenols	30, 31
12	*p*-X-thiophenols	31
13	2-X-thiophene-5-carboxylic acids	32
14	*m*-X-anilines	33, 34
15	*p*-X-anilines	33, 35

Table 1. (c) Data on log (K_X/K_H), where K is an ionization constant

		Substituent									
Solvent	Reaction	F	Cl	Br	I	CH_3	NO_2	CN	OCH_3	$COCH_3$	$COOCH_3$
1	1		0.32					0.42	0.24		0.24
	2		0.172		0.153	−0.051	0.345				
	3	0.066	0.122	0.124	0.134	−0.058	0.461		−0.048		
	4		0.053			−0.024	0.187		−0.029		
	5	0.338	0.373	0.391	0.352	−0.069	0.710		0.115	0.376	
	6	0.062	0.226	0.232		−0.17	0.778		−0.268	0.502	
	7		0.31	0.34		−0.17			−0.27		
	8		0.17			−0.15	0.59				
	9		0.32	0.24		−0.17	0.84				
	10		0.76	0.83		−0.20	1.50		0.24		
	11		0.54	0.62	0.68		2.90	2.04	−0.23	1.88	
	12		0.835			−0.045	1.374		0.230		
	13		0.480	0.595		−0.205	1.900		−0.160	1.285	
	14	1.09	1.13			−0.11	2.12	1.83	0.39	1.03	1.04
	15	0.02	0.68	0.77	0.86	−0.45	3.67	2.92	−0.73		2.14
2	1		0.50					0.66	0.38		0.36
	2	0.18	0.19	0.19	0.16	−0.09	0.49	0.45	0.04		
	3	0.11	0.19	0.19	0.22	−0.09	0.61	0.52	−0.06		
	4		0.096			−0.053	0.323		−0.034		
	5		0.53	0.58		−0.08	1.14	0.95		0.59	0.64
	6		0.44	0.45		−0.20	1.27	1.10	−0.32	0.70	0.51
	7		0.44	0.43		−0.22			−0.39	0.84	
	8		0.22	0.27		−0.17	0.85				
	9		0.27	0.25		−0.12	0.84				
	10		1.11	1.07	1.08	−0.15			0.29	0.85	0.78
	11		0.68	0.78	0.87	−0.33	3.60		−0.24	2.07	1.70
	12		0.93	0.88	0.93	−0.21			0.24	0.85	0.80
	13		0.72	0.78	0.79	−0.29	2.79		−0.30	1.85	1.61
	14	1.08	1.17	1.20	1.22	−0.11					
	15	0.02	0.64	0.76	0.92	−0.49	3.68				
3	5		0.58			−0.20	1.35				
	6		0.43			−0.20	1.45				
	10		1.22	1.30	1.33	−0.17			0.17	0.99	0.87
	11		0.88	0.98	1.06	−0.33	3.04		−0.27	2.23	1.88
	12		1.15	1.10	1.24	−0.24			0.12	0.97	0.92
	13		0.87	0.95	1.00	−0.38	2.90		−0.44	2.04	1.82

to predict missing data. The average standard deviation of the calculated values of the missing data can be computed by extension of the procedure, developed for the two-mode factor analysis case,[38] to the three-mode situation. Using a linearization of the model around the true parameters, the σ^2_{pred} = [variance of (new observation − prediction)] can be computed for all missing data. The square root of its average is given in the last column of Table 3. It follows that the factor analysis model (2,2,2) is the best one for the prediction of missing data, and that model (2,3,3) is of no use because it gives imprecise predictions. Detailed information on the fit of the observations by this model is given in Table 4. It appears that there are no large discrepancies in the fit for the individual solvents, reactions, and substituents. The rather large values of the root of the mean residual sum of squares for reaction 11 and for the NO_2 substituent correspond with the large substituent effects observed for this reaction and for this substituent. The residuals (*i.e.*, observation–estimation), calculated by the factor analysis model (2,2,2), are given in Table 5. As can be expected from the discussion of Table 4, large residuals are found for the NO_2-substituted compound in reaction 11, in the three solvents. For reaction 6 in solvent 1 all residuals are negative, but otherwise

Table 2. (a) Investigated two-phase systems

Code	Two-phase system	References
4	50 v/v % methanol–water/octadecylsilylsilica at 25 °C	16
5	30 v/v % acetonitrile–water/octadecylsilylsilica at 25 °C	16, 36
6	32.5 v/v % tetrahydrofuran–water/octadecylsilylsilica at 25 °C	36
7	helium/Apiezon M at 206 °C	14
8	helium/polyethylene glycol 20,000 at 206 °C	14
9	n-octanol/water	37
10	50 v/v % methanol–water/n-hexadecane at 25 °C	16
11	30 v/v % acetonitrile–water/n-hexadecane at 25 °C	16,36
12	32.5 v/v % tetrahydrofuran–water/n-hexadecane at 25 °C	36

Table 2. (b) Investigated reactions (partition of the tabulated compounds)

Code	Compounds	Code	Compounds	Code	Compounds
16	*m*-benzoic acids	18	*m*-phenols	20	*m*-anilines
17	*p*-benzoic acids	19	*p*-phenols	21	*p*-anilines

Table 2. (c) Data on ΔΔ log V_N, where V_N is a chromatographic net retention volume, and ΔΔlog *P*, where *P* is a partition coefficient

		Substituent									
Solvent	Reaction	F	Cl	Br	I	CH_3	NO_2	CN	OCH_3	$COCH_3$	$COOCH_3$
4	16	0.158	0.162	0.100	0.034	−0.024	0.229	0.258	0.150	0.104	
	17	0.144	0.181	0.134	0.148	−0.026	0.295	0.223	0.118	0.151	
	18	0.120	0.154	0.099	0.073	−0.107	0.363	0.368	−0.035	0.210	0.170
	19	0.028	0.126	0.086	0.080	−0.096	0.314	0.314	−0.158	0.146	0.186
	20	0.071	0.117	0.081	0.051	−0.074	0.317	0.351	−0.045	0.161	
	21	−0.005	0.085	0.076	0.060	−0.065	0.176	0.274	−0.125	0.105	0.110
5	16	0.126	0.129	0.125	0.120	−0.003	0.184	0.161	0.076	0.079	
	17	0.154	0.161	0.163	0.142	−0.004	0.237	0.211	0.087	0.102	
	18	0.166	0.072	0.108	0.076	−0.078	0.324	0.362	0.036	0.236	0.165
	19	0.052	0.054	0.089	0.073	−0.068	0.276	0.284	−0.072	0.128	0.129
	20	0.239	0.244	0.244	0.211	0.007	0.417	0.407	0.040	0.161	
	21	0.077	0.185	0.204	0.205	−0.002	0.317	0.332	−0.152	0.014	0.177
6	16	0.223	0.228	0.224	0.182	−0.006	0.471	0.463	0.154	0.302	
	17	0.206	0.261	0.264	0.247		0.590	0.564	0.059	0.401	
	18	0.269	0.274	0.257	0.247	−0.030	0.544	0.538	0.066	0.221	0.259
	19	0.136	0.219	0.213	0.216	−0.035	0.497	0.421	−0.123	0.066	0.237
	20	0.321	0.347	0.350	0.347	0.057	0.667	0.611	0.028	0.123	
	21	0.175	0.232	0.243	0.270	0.018	0.460	0.451	0.159	0.136	0.263
7	18	0.056	0.065	0.051	0.062	0.030	0.104	0.108	0.023		
	19	0.054	0.042	0.042	0.049	0.000	0.232	0.213	0.015		
	20	0.030	0.029	0.029	0.028	−0.015	0.093	0.092	−0.001		
	21	0.016	0.024	0.033	0.051	−0.025	0.297	0.206	−0.034		
8	18	0.130	0.150	0.166	0.158	−0.045	0.299	0.258	−0.007		
	19	0.096	0.147	0.157	0.157	−0.047	0.513	0.425	−0.056		
	20	0.135	0.146	0.158	0.146	−0.028	0.265	0.213	0.016		
	21	0.051	0.134	0.146	0.151	−0.056	0.688	0.493	−0.085		
9	16	0.140	0.100	0.140	0.140	−0.060	0.240	0.180	0.170		−0.030
	17	0.070	0.070	0.130	0.030	−0.160	0.300	0.260	0.110		
	18	0.330	0.330	0.310	0.350	−0.060	0.820	0.810	0.140	0.480	
	19	0.170	0.220	0.270	0.330	−0.080	0.730	0.710	−0.100	0.440	0.510
	20	0.260	0.230	0.340		−0.060	0.750		0.050		
	21	0.110	0.160	0.500		−0.070	0.770		0.070		
10	18	0.741	0.782	0.802	0.749	−0.074	1.344	1.221	0.264	0.662	0.688
	19	0.533	0.626	0.621	0.603	−0.091	1.109	1.053	0.108	0.597	0.613
11	18	0.126	0.204	−0.054	0.511	−0.119	0.643	0.654	−0.369	0.360	0.024
	19	−0.422	0.258	0.151	0.054	−0.133	0.488	0.508	−0.334	0.300	0.012
12	18	0.998	0.893	0.850	0.637	−0.070	0.930	1.010	0.394	0.711	0.810
	19	0.847	0.937	1.058	0.665	−0.102	1.120	1.090	0.391	0.840	0.876

Table 3. Summary of the results of an analysis of variance and of three-mode factor analysis of the data, presented in Table 1c

Model	NFP	NFR	NFS	NPAR	DF	$\hat{\sigma}$	$\hat{\sigma}_{pred}$
Analysis of variance				26	187	0.68	
Three-mode factor analysis	1	1	1	26	187	0.24	0.27
Three-mode factor analysis	2	2	2	52	161	0.10	0.13
Three-mode factor analysis	2	2	3	61	152	0.10	0.17
Three-mode factor analysis	2	3	2	66	147	0.08	0.16
Three-mode factor analysis	2	3	3	77	136	0.07	0.36

Number of observations: 213. Number of missing values: 237. NFP = number of factors for the solvents. NFR = number of factors for the reactions. NFS = number of factors for the substituents. NPAR = number of estimated parameters. DF = degrees of freedom. $\hat{\sigma}$ = standard deviation of the model. $\hat{\sigma}_{pred}$ = average standard deviation of (new observation − prediction) for the missing data.

Table 4. Fit of the three-mode factor analysis model (2,2,2) for the individual solvents, reactions and substituents

Solvents			Reactions			Substituents		
Code	NO	RMRSS	Code	NO	RMRSS	Symbol	NO	RMRSS
1	84	0.09	1	8	0.09	F	9	0.08
2	93	0.08	2	12	0.03	Cl	36	0.06
3	36	0.10	3	15	0.00	Br	26	0.05
			4	8	0.03	I	17	0.06
			5	18	0.05	CH_3	33	0.05
			6	18	0.12	NO_2	27	0.16
			7	9	0.10	CN	9	0.07
			8	7	0.05	OCH_3	26	0.11
			9	8	0.06	$COCH_3$	16	0.08
			10	19	0.07	$COOCH_3$	14	0.11
			11	23	0.15			
			12	18	0.06			
			13	22	0.12			
			14	13	0.12			
			15	15	0.08			

NO = number of observations. RMRSS = $(RSS/NO)^{\frac{1}{2}}$ = root of the mean residual sum of squares.

the signs of the residuals appear to be distributed randomly. It can thus be concluded that the model gives an adequate fit of the observations. The values, predicted by the factor analysis model (2,2,2) for the 237 missing data, are given in Table 6, together with their standard deviation. This Table gives a striking illustration of the merits of the proposed method. For, 90 of the missing data are from reactions on which no observations at all are available in solvent 3 (95 v/v % ethanol). The traditional way to give a mathematical–statistical description of the data is by the Hammett (1), Taft (2), or Nieuwdorp equation (3), *i.e.*, for each reaction and each solvent separately. Of course, this approach does not offer the possibility to predict data on a reaction in solvents where no measurements at all are available. Further, for 53 missing data that can be predicted by the Nieuwdorp equation, the 0.90 probability interval is larger than for the predictions by three-mode factor analysis, for reasons set forth in the theoretical part. In only 73 cases is the 0.90 probability interval for predictions by the Nieuwdorp equation smaller. For 3 series only 3 data points are available, which makes it impossible to calculate the intervals for the 21 missing data in these series with the Nieuwdorp equation.

We conclude from Tables 3—6 that the combined influence of solvent, reaction type, and substituent on chemical equilibrium constants can be described satisfactorily by equation (6), *i.e.*, by the factor analysis model (2,2,2). The values of the factors and of the elements of the three-mode core matrix in this model are given in Tables 7 and 8, respectively.

Case 2.—It must be admitted that the range of solvents in case 1 is rather small: water, 50, and 95 v/v % ethanol. Therefore we applied equation (6) to a set of phase equilibrium constants where the range of two-phase systems is very large (see Table 2a). The results of three-mode factor analysis with missing data, applied to these data (listed in Table 2c), are given in Table 9. Also given are the results of an analysis of variance. These results are analogous to the results shown in Table 3. Again the three-mode factor analysis model is better than the analysis of variance model. With the factor analysis model, and one factor for the reactions and two factors for the two-phase systems and for the substituents, the residual standard deviation of the observations is equal to 0.09, and the standard deviation of the predictions is equal to 0.10. Thus, even in this case where the phases range from a gas, *via* apolar liquids and water, to a solid, three-mode factor analysis gives a good description of the observations and an accurate prediction of the 37% missing data. The values of $\hat{\sigma}$ and $\hat{\sigma}_{pred.}$ for the (2,1,2) model are even better than those for the (2,2,2) model in Table 3.

Regression analyses of the data in Table 2c with the Taft equation (2) yield a pooled value for the residual standard deviation that is equal to that obtained with the factor analysis model (2,1,2). However, the number of parameters in the factor analysis model is much less than that in the regression analysis model, *viz.*, 39 *versus* 76.

Detailed information on the fit of the observations by the factor analysis model (2,1,2) is given in Table 10. No large discrepancies are noted in the fit for the individual solvents, reactions, and substituents. The residuals, calculated by the factor analysis model (2,1,2), are given in Table 11. The signs of the residuals appear to be distributed randomly. It can thus be concluded that the model gives an adequate fit of the observations.

Nearly all missing data in this case are from reactions on which no observations at all are available in some two-phase systems. As mentioned above, these missing data cannot be

Table 5. Residuals [*i.e.*, observation − estimation by the three-mode factor analysis model (2,2,2)] of the data in Table 1c

		Substituent									
Solvent	Reaction	F	Cl	Br	I	CH_3	NO_2	CN	OCH_3	$COCH_3$	$COOCH_3$
1	1		−0.06					−0.01	0.05		0.13
	2		0.02		−0.01	−0.02	−0.02				
	3	−0.02	−0.02	−0.03	−0.03	0.00	−0.03		0.00		
	4		−0.02			0.00	−0.05		−0.01		
	5	−0.04	−0.01	0.01	−0.04	0.00	−0.12		0.04	0.02	
	6	−0.13	−0.07	−0.08		−0.06	−0.21		−0.19	−0.11	
	7		0.02	0.02		−0.03			−0.11		
	8		−0.02			−0.08	−0.07				
	9		0.11	0.01		−0.08	0.07				
	10		−0.07	0.03		−0.12	0.09		−0.06		
	11		−0.12	−0.10	−0.10		0.24	−0.14	0.11	0.09	
	12		0.09			0.04	0.02		−0.01		
	13		−0.13	−0.06		0.05	−0.36		0.07	−0.17	
	14	0.15	0.16			0.07	−0.05	0.05	0.23	0.05	0.11
	15	0.01	0.08	0.06	0.04	0.00	0.17	0.05	−0.03		−0.05
2	1		−0.02					−0.06	0.19		
	2	0.01	−0.01	−0.01	−0.05	−0.05	−0.01	0.04	0.03		0.07
	3	0.00	0.02	0.01	0.02	−0.02	0.01	0.03	−0.01		
	4		0.01			−0.02	0.03		−0.01		
	5		0.03	0.08		0.02	−0.03	−0.01		0.04	0.12
	6		0.08	0.07		−0.07	0.05	0.10	−0.23	−0.05	−0.14
	7		0.10	0.05		−0.05			−0.20	−0.14	
	8		−0.02	0.02		−0.08	0.04				
	9		0.01	−0.02		−0.02	−0.09				
	10		0.01	−0.01	−0.02	0.01			−0.01	0.02	−0.05
	11		−0.10	−0.07	−0.06	0.03	0.45		0.15	−0.05	−0.10
	12		−0.05	−0.09	−0.06	−0.06			0.01	0.02	−0.01
	13		−0.02	−0.02	−0.07	0.01	0.06		−0.03	0.09	0.10
	14	−0.09	−0.09	−0.07	−0.09	0.16					
	15	0.02	−0.02	−0.02	0.02	0.01	−0.19				
3	5		0.02			−0.08	0.01				
	6		0.04			−0.06	0.13				
	10		−0.01	0.09	0.09	0.03			−0.11	−0.06	−0.15
	11		0.04	0.07	0.06	0.06	−0.32		0.14	−0.02	−0.03
	12		0.05	0.01	0.12	−0.05			−0.10	−0.05	−0.06
	13		0.07	0.09	0.07	−0.05	−0.03		−0.15	0.16	0.20

predicted by the Hammett equation or its extensions. However, they can be predicted by the present approach, and the predictions by the factor analysis model (2,1,2) are given in Table 12, together with their standard deviation, which ranges from 0.09 to 0.13. The values of the factors and of the elements of the three-mode core matrix in the (2,1,2) model are given in Tables 13 and 14. It appears from Tables 9—12 that the combined influence of two-phase system, reaction type, and substituent on phase equilibrium constants can be described adequately by equation (6), even when the two-phase systems differ widely. We believe that an analogous conclusion holds for chemical equilibrium constants and reaction rate constants, in widely different solvents.

It would be interesting to learn how the substituent factors S_l in Tables 7 and 13 are related to the mechanisms by which the substituents are believed to exert their influence on equilibrium constants: induction, resonance between the substituent and the aromatic ring, and direct resonance between the substituent and the reaction centre. However, this is not easily possible for the following reason. The inductive and resonance effects of substituents are quantitatively described by the σ_I and σ_R^0 values given by Taft[3] or Nieuwdorp.[1] The direct resonance effect is reflected in Nieuwdorp's σ_E values, but not separately: the σ_E values contain unknown contributions of the inductive and resonance effect. In the author's opinion a good measure for the direct resonance effect does not exist.

So, if the regression of S_l on σ_I, σ_R, and σ_E is investigated, the values of the regression coefficients do not tell us the relative importance of the three mechanisms. Only if the regression coefficient of S_l on σ_E is zero would the ratio of the regression coefficients on σ_I and σ_R give the relative importance of the inductive and resonance mechanisms. However, this will clearly not be the case, as in most investigated reactions the direct resonance effect is definitely present.

Conclusions

Three-mode factor analysis is a good mathematical–statistical model for the simultaneous description of the influence of solvent, reaction type, and substituents on equilibrium constants. The residual standard deviation of the observations is of the same magnitude as obtained by a regression analysis model, *viz.*, by applying the Taft equation to each reaction in each solvent separately. However, the number of parameters in the factor analysis model is far less than that in the regression analysis model. The advantage of the factor analysis model is that it is better suited for the prediction of missing data. It can even predict data that cannot be predicted at all with the Taft equation, *viz.*, for reactions on which no data at all are available in a particular solvent.

Table 6. Predictions of the missing data in Table 1c by the three-mode factor analysis model (2,2,2) and their standard deviations

Solvent	Reaction	Substituent									
		F	Cl	Br	I	CH$_3$	NO$_2$	CN	OCH$_3$	COCH$_3$	COOCH$_3$
1	1	0.46 ± 0.13		0.37 ± 0.12	0.36 ± 0.12	−0.02 ± 0.11	0.52 ± 0.13			0.07 ± 0.14	
	2	0.14 ± 0.12		0.15 ± 0.11				0.30 ± 0.19	0.02 ± 0.11	0.18 ± 0.13	0.16 ± 0.12
	3							0.41 ± 0.12		0.31 ± 0.12	0.27 ± 0.12
	4	0.05 ± 0.13		0.08 ± 0.11	0.08 ± 0.11			0.19 ± 0.12		0.14 ± 0.13	0.12 ± 0.12
	5							0.68 ± 0.12			0.34 ± 0.12
	6				0.33 ± 0.11			0.81 ± 0.11			0.53 ± 0.11
	7	0.13 ± 0.12			0.35 ± 0.11		1.22 ± 0.16	1.00 ± 0.14		0.84 ± 0.14	0.71 ± 0.13
	8	0.13 ± 0.13		0.20 ± 0.11	0.22 ± 0.11			0.54 ± 0.13	−0.05 ± 0.12	0.41 ± 0.15	0.35 ± 0.13
	9	0.12 ± 0.13			0.24 ± 0.11			0.63 ± 0.13	−0.08 ± 0.12	0.50 ± 0.14	0.43 ± 0.13
	10	0.91 ± 0.15			0.81 ± 0.12			1.15 ± 0.16		0.42 ± 0.15	0.46 ± 0.13
	11	0.32 ± 0.13				−0.31 ± 0.11					1.52 ± 0.12
	12	0.80 ± 0.14		0.72 ± 0.12	0.73 ± 0.12			1.10 ± 0.15		0.45 ± 0.14	0.48 ± 0.13
	13	0.36 ± 0.12			0.71 ± 0.11			1.85 ± 0.12			1.25 ± 0.12
	14			0.97 ± 0.12	1.00 ± 0.12						
	15									2.65 ± 0.17	
2	1	0.57 ± 0.14		0.50 ± 0.12	0.51 ± 0.12	−0.05 ± 0.11	0.88 ± 0.15			0.26 ± 0.13	
	2									0.25 ± 0.12	0.23 ± 0.12
	3									0.38 ± 0.12	0.33 ± 0.12
	4	0.06 ± 0.14		0.09 ± 0.12	0.10 ± 0.12			0.24 ± 0.12		0.18 ± 0.13	0.15 ± 0.12
	5	0.48 ± 0.12			0.52 ± 0.11				0.07 ± 0.11		
	6	0.24 ± 0.12			0.41 ± 0.11						
	7	0.16 ± 0.13			0.42 ± 0.12		1.44 ± 0.17	1.18 ± 0.15	0.83 ± 0.13		
	8	0.16 ± 0.15			0.27 ± 0.12			0.66 ± 0.12	−0.06 ± 0.12	0.50 ± 0.14	0.43 ± 0.12
	9	0.15 ± 0.14			0.30 ± 0.12			0.76 ± 0.12	−0.09 ± 0.12	0.60 ± 0.14	0.51 ± 0.12
	10	1.14 ± 0.16					2.16 ± 0.17	1.77 ± 0.17			
	11	0.39 ± 0.13						2.58 ± 0.14			
	12	0.10 ± 0.15					2.03 ± 0.16	1.66 ± 0.16			
	13	0.44 ± 0.13						2.23 ± 0.13			
	14						3.02 ± 0.19	2.47 ± 0.17	0.14 ± 0.12	1.46 ± 0.14	1.36 ± 0.13
	15							3.17 ± 0.17	−0.79 ± 0.14	2.94 ± 0.17	2.42 ± 0.14
3	1	0.62 ± 0.15	0.58 ± 0.13	0.57 ± 0.13	0.58 ± 0.13	−0.07 ± 0.11	1.07 ± 0.17	0.87 ± 0.14	0.19 ± 0.12	0.37 ± 0.13	0.38 ± 0.12
	2	0.19 ± 0.13	0.22 ± 0.12	0.22 ± 0.12	0.23 ± 0.11	−0.05 ± 0.11	0.56 ± 0.13	0.48 ± 0.12	0.01 ± 0.11	0.29 ± 0.12	0.26 ± 0.12
	3	0.12 ± 0.12	0.18 ± 0.11	0.20 ± 0.11	0.21 ± 0.11	−0.07 ± 0.11	0.65 ± 0.13	0.53 ± 0.12	−0.06 ± 0.11	0.41 ± 0.12	0.36 ± 0.12
	4	0.07 ± 0.14	0.10 ± 0.13	0.10 ± 0.12	0.11 ± 0.12	−0.03 ± 0.11	0.31 ± 0.14	0.26 ± 0.12	−0.02 ± 0.12	0.19 ± 0.13	0.17 ± 0.12
	5	0.52 ± 0.13		0.56 ± 0.12	0.58 ± 0.12			1.10 ± 0.12	0.06 ± 0.11	0.65 ± 0.12	0.60 ± 0.11
	6	0.26 ± 0.12		0.41 ± 0.11	0.44 ± 0.11			1.08 ± 0.12	−0.10 ± 0.11	0.82 ± 0.12	0.71 ± 0.11
	7	0.17 ± 0.14	0.37 ± 0.13	0.40 ± 0.12	0.44 ± 0.13	−0.18 ± 0.11	1.52 ± 0.18	1.25 ± 0.16	−0.20 ± 0.12	1.03 ± 0.14	0.88 ± 0.14
	8	0.17 ± 0.15	0.26 ± 0.13	0.27 ± 0.13	0.29 ± 0.13	−0.10 ± 0.11	0.88 ± 0.14	0.72 ± 0.13	−0.07 ± 0.12	0.54 ± 0.13	0.47 ± 0.12
	9	0.17 ± 0.15	0.28 ± 0.13	0.30 ± 0.12	0.32 ± 0.12	−0.11 ± 0.11	1.00 ± 0.14	0.82 ± 0.13	−0.10 ± 0.12	0.64 ± 0.13	0.55 ± 0.12
	10	1.24 ± 0.16					2.54 ± 0.18	2.08 ± 0.18			
	11	0.42 ± 0.14						2.75 ± 0.14			
	12	1.08 ± 0.15					2.36 ± 0.17	1.93 ± 0.17			
	13	0.47 ± 0.13						2.40 ± 0.14			
	14	1.27 ± 0.15	1.40 ± 0.13	1.41 ± 0.13	1.46 ± 0.13	−0.31 ± 0.11	3.43 ± 0.21	2.81 ± 0.19	0.12 ± 0.13	1.70 ± 0.15	1.57 ± 0.14
	15	−0.01 ± 0.15	0.67 ± 0.14	0.80 ± 0.14	0.92 ± 0.14	−0.51 ± 0.12	3.99 ± 0.19	3.27 ± 0.20	0.82 ± 0.15	3.03 ± 0.19	2.50 ± 0.16

Table 7. Values of the factors in the three-mode factor analysis model (2,2,2)

Solvents			Reactions			Substituents		
Code	Factor 1	Factor 2	Code	Factor 1	Factor 2	Symbol	Factor 1	Factor 2
1	0.479	0.828	1	0.115	−0.268	F	0.129	0.561
2	0.594	−0.061	2	0.065	−0.048	Cl	0.186	0.360
3	0.646	−0.557	3	0.079	0.015	Br	0.196	0.315
			4	0.038	0.001	I	0.209	0.293
			5	0.153	−0.157	CH_3	−0.065	0.060
			6	0.160	0.013	NO_2	0.609	−0.009
			7	0.188	0.103	CN	0.499	−0.009
			8	0.106	0.011	OCH_3	−0.040	0.364
			9	0.122	0.032	$COCH_3$	0.370	−0.405
			10	0.282	−0.472	$COOCH_3$	0.322	−0.263
			11	0.414	0.199			
			12	0.264	−0.393			
			13	0.358	0.100			
			14	0.395	−0.364			
			15	0.509	0.575			

Table 8. Elements of the three-mode core matrix of the model (2,2,2)

p	*r*	*s*	Element
1	1	1	21.059
1	1	2	0.019
1	2	1	0.008
1	2	2	−4.314
2	1	1	−0.057
2	1	2	0.031
2	2	1	1.299
2	2	2	−0.277

Table 9. Summary of the results of an analysis of variance and of three-mode factor analysis of the data presented in Table 2c

Model	NFP	NFR	NFS	NPAR	DF	$\hat{\sigma}$	$\hat{\sigma}_{pred.}$
Analysis of variance				23	319	0.37	
Three-mode factor analysis	1	1	1	23	319	0.11	0.12
Three-mode factor analysis	2	1	2	39	303	0.09	0.10
Three-mode factor analysis	3	1	3	53	289	0.09	0.10
Three-mode factor analysis	2	2	2	46	296	0.07	1.21
Three-mode factor analysis	3	3	3	75	267	0.06	0.77

Number of observations: 342. Number of missing data: 198.

Table 10. Fit of the three-mode factor analysis model (2,1,2) for the individual two-phase systems, reactions and substituents

Solvents			Reactions			Substituents		
Code	NO	RMRSS	Code	NO	RMRSS	Symbol	NO	RMSS
4	57	0.06	16	36	0.09	F	38	0.09
5	57	0.05	17	34	0.10	Cl	38	0.07
6	56	0.09	18	85	0.08	Br	38	0.06
7	48	0.11	19	86	0.08	I	36	0.08
8	32	0.08	20	49	0.06	CH_3	37	0.04
9	32	0.04	21	52	0.09	NO_2	38	0.11
10	20	0.06				CN	36	0.09
11	20	0.14				OCH_3	38	0.09
12	20	0.09				$COCH_3$	26	0.10
						$COOCH_3$	17	0.08

NO = number of observations. RMRSS = $(RSS/NO)^{\frac{1}{2}}$ = root of the mean residual sum of squares.

Table 11. Residuals of the data in Table 2c, estimated by the three-mode factor analysis model (2,1,2)

		Substituent									
Solvent	Reaction	F	Cl	Br	I	CH_3	NO_2	CN	OCH_3	$COCH_3$	$COOCH_3$
4	16	0.12	0.09	0.03	−0.05	0.00	0.04	0.08	0.17	0.01	
	17	0.09	0.09	0.05	0.05	0.00	0.06	0.01	0.14	0.04	
	18	0.05	0.02	−0.03	−0.08	−0.07	0.02	0.05	−0.01	0.04	0.05
	19	−0.04	0.00	−0.03	−0.06	−0.06	0.00	0.02	−0.13	−0.01	0.08
	20	0.00	−0.02	−0.04	−0.10	−0.03	−0.02	0.03	−0.02	−0.01	
	21	−0.08	−0.04	−0.04	−0.08	−0.03	−0.14	−0.02	−0.10	−0.05	0.00
5	16	0.05	0.04	0.03	0.03	0.02	−0.01	−0.02	0.07	−0.02	
	17	0.06	0.05	0.05	0.03	0.02	0.00	−0.01	0.08	−0.02	
	18	0.03	−0.10	−0.06	−0.09	−0.04	−0.02	0.03	0.02	0.05	0.01
	19	−0.07	−0.11	−0.07	−0.08	−0.04	−0.05	−0.02	−0.08	−0.04	−0.02
	20	0.11	0.07	0.07	0.05	0.04	0.07	0.08	0.03	−0.02	
	21	−0.05	0.03	0.04	0.05	0.03	−0.01	0.03	−0.16	−0.16	0.03
6	16	0.09	0.06	0.06	0.03	0.03	0.15	0.16	0.13	0.13	
	17	0.04	0.06	0.06	0.05		0.19	0.19	0.03	0.19	
	18	0.03	−0.02	−0.05	−0.04	0.03	−0.04	−0.01	0.03	−0.09	−0.02
	19	−0.09	−0.06	−0.07	−0.05	0.02	−0.05	−0.10	−0.16	−0.22	−0.02
	20	0.08	0.05	0.05	0.06	0.11	0.09	0.06	−0.01	−0.19	
	21	−0.05	−0.05	−0.04	0.01	0.07	−0.08	−0.07	0.12	−0.15	0.00
7	18	0.03	0.00	0.00	−0.01	0.05	−0.06	−0.05	0.04		
	19	0.03	−0.02	−0.01	−0.02	0.02	0.08	0.07	0.03		
	20	0.00	−0.03	−0.03	−0.04	0.01	−0.07	−0.06	0.02		
	21	−0.01	−0.03	−0.02	−0.02	−0.01	0.14	0.06	−0.02		
8	18	0.04	−0.01	0.01	−0.02	0.00	−0.11	−0.13	0.03		
	19	0.01	−0.01	0.01	−0.01	0.00	0.13	0.07	−0.03		
	20	0.04	−0.02	0.00	−0.04	0.02	−0.15	−0.17	0.05		
	21	−0.04	−0.02	0.00	−0.02	−0.01	0.30	0.13	−0.06		
9	16	0.01	−0.08	−0.04	−0.04	−0.02	−0.15	−0.19	0.17		−0.20
	17	−0.09	−0.15	−0.09	−0.20	−0.11	−0.18	−0.19	0.11		
	18	0.09	0.00	−0.01	0.02	0.01	0.12	0.15	0.14	0.12	
	19	−0.05	−0.09	−0.03	0.02	−0.01	0.07	0.09	−0.10	0.10	0.23
	20	0.02	−0.10	0.02		0.01	0.04		0.05		
	21	−0.11	−0.15	0.20		0.00	0.11		0.07		
10	18	0.09	0.07	0.05	0.11	0.03	0.10	0.03	0.08	−0.04	0.00
	19	−0.08	−0.05	−0.09	0.00	0.01	−0.06	−0.06	−0.07	−0.06	−0.03
11	18	0.25	0.10	−0.09	0.28	−0.01	−0.04	0.03	−0.10	0.09	−0.03
	19	−0.31	0.16	0.11	−0.16	−0.03	−0.15	−0.08	−0.08	0.05	−0.04
12	18	0.03	0.02	−0.12	−0.03	−0.02	−0.19	−0.08	−0.07	0.00	−0.07
	19	−0.06	0.12	0.15	0.04	−0.05	0.07	0.07	−0.05	0.17	0.06

Table 12. Predictions of the missing data in Table 2c by the three-mode factor analysis model (2,1,2)

		Substituents									
Solvent	Reaction	F	Cl	Br	I	CH_3	NO_2	CN	OCH_3	$COCH_3$	$COOCH_3$
7	16	0.02 ± 0.09	0.03 ± 0.09	0.03 ± 0.09	0.04 ± 0.09	−0.01 ± 0.09	0.09 ± 0.09	0.08 ± 0.09	−0.01 ± 0.09	0.04 ± 0.09	0.03 ± 0.09
	17	0.02 ± 0.09	0.04 ± 0.09	0.04 ± 0.09	0.05 ± 0.09	−0.01 ± 0.09	0.11 ± 0.09	0.10 ± 0.09	−0.01 ± 0.09	0.05 ± 0.09	0.04 ± 0.09
8	16	0.05 ± 0.09	0.09 ± 0.09	0.09 ± 0.09	0.10 ± 0.09	−0.03 ± 0.09	0.23 ± 0.09	0.21 ± 0.09	−0.02 ± 0.09	0.11 ± 0.09	0.08 ± 0.09
	17	0.06 ± 0.09	0.11 ± 0.09	0.11 ± 0.09	0.12 ± 0.09	−0.03 ± 0.09	0.28 ± 0.09	0.26 ± 0.09	−0.02 ± 0.09	0.14 ± 0.09	0.10 ± 0.09
10	16	0.36 ± 0.10	0.40 ± 0.10	0.42 ± 0.10	0.36 ± 0.10	−0.06 ± 0.09	0.69 ± 0.11	0.66 ± 0.11	0.10 ± 0.09	0.39 ± 0.10	0.38 ± 0.10
	17	0.45 ± 0.10	0.49 ± 0.10	0.51 ± 0.10	0.44 ± 0.10	−0.07 ± 0.09	0.85 ± 0.12	0.81 ± 0.12	0.13 ± 0.09	0.48 ± 0.10	0.47 ± 0.10
	20	0.66 ± 0.11	0.72 ± 0.11	0.76 ± 0.11	0.64 ± 0.10	−0.11 ± 0.10	1.25 ± 0.13	1.19 ± 0.13	0.19 ± 0.10	0.70 ± 0.11	0.69 ± 0.11
	21	0.61 ± 0.11	0.67 ± 0.10	0.71 ± 0.11	0.60 ± 0.10	−0.10 ± 0.09	1.17 ± 0.12	1.12 ± 0.12	0.18 ± 0.10	0.65 ± 0.10	0.65 ± 0.11
11	16	−0.07 ± 0.09	0.06 ± 0.09	0.02 ± 0.09	0.13 ± 0.09	−0.06 ± 0.09	0.38 ± 0.10	0.35 ± 0.10	−0.15 ± 0.09	0.15 ± 0.09	0.03 ± 0.09
	17	−0.08 ± 0.10	0.07 ± 0.09	0.03 ± 0.09	0.16 ± 0.10	−0.08 ± 0.09	0.47 ± 0.10	0.42 ± 0.10	−0.19 ± 0.10	0.18 ± 0.10	0.04 ± 0.10
	20	−0.12 ± 0.10	0.11 ± 0.10	0.04 ± 0.10	0.23 ± 0.10	−0.11 ± 0.10	0.68 ± 0.11	0.62 ± 0.11	−0.27 ± 0.10	0.27 ± 0.10	0.05 ± 0.10
	21	−0.12 ± 0.10	0.10 ± 0.10	0.04 ± 0.10	0.22 ± 0.10	−0.10 ± 0.10	0.64 ± 0.11	0.58 ± 0.11	−0.25 ± 0.10	0.25 ± 0.10	0.05 ± 0.10
12	16	0.54 ± 0.11	0.48 ± 0.10	0.54 ± 0.11	0.37 ± 0.10	−0.03 ± 0.09	0.62 ± 0.11	0.61 ± 0.11	0.26 ± 0.10	0.40 ± 0.10	0.49 ± 0.11
	17	0.66 ± 0.11	0.59 ± 0.11	0.66 ± 0.11	0.46 ± 0.10	−0.04 ± 0.10	0.76 ± 0.12	0.75 ± 0.11	0.32 ± 0.10	0.49 ± 0.10	0.60 ± 0.11
	20	0.97 ± 0.12	0.87 ± 0.12	0.97 ± 0.12	0.67 ± 0.11	−0.05 ± 0.10	1.12 ± 0.13	1.10 ± 0.13	0.47 ± 0.11	0.72 ± 0.11	0.88 ± 0.12
	21	0.91 ± 0.12	0.81 ± 0.11	0.91 ± 0.12	0.63 ± 0.11	−0.05 ± 0.10	1.05 ± 0.12	1.02 ± 0.12	0.44 ± 0.11	0.67 ± 0.11	0.82 ± 0.12

Table 13. Values of the factors in the three-mode factor analysis model (2,1,2)

Solvents			Reactions		Substituents		
Code	Factor 1	Factor 2	Code	Factor 1	Symbol	Factor 1	Factor 2
4	0.132	−0.169	16	0.261	F	0.277	0.500
5	0.150	−0.088	17	0.321	Cl	0.298	0.189
6	0.256	−0.115	18	0.470	Br	0.315	0.307
7	0.062	−0.091	19	0.440	I	0.263	−0.033
8	0.160	−0.196	20	0.471	CH_3	−0.042	0.113
9	0.296	−0.223	21	0.440	NO_2	0.507	−0.411
10	0.587	−0.045			CN	0.485	−0.349
11	0.191	−0.751			OCH_3	0.086	0.496
12	0.630	0.536			$COCH_3$	0.286	−0.062
					$COOCH_3$	0.288	0.257

Table 14. Elements of the three-mode core matrix of the model (2,1,2)

p	*r*	*s*	Element
1	1	1	8.817
1	1	2	0
2	1	1	0
2	1	2	1.939

References

1 G. H. E. Nieuwdorp, C. L. de Ligny, and J. C. van Houwelingen, *J. Chem. Soc., Perkin Trans. 2*, 1979, 537.

2 L. P. Hammett, 'Physical Organic Chemistry,' McGraw-Hill, New York, 1940, p. 184.

3 R. W. Taft and I. C. Lewis, *J. Am Chem. Soc.*, 1958, **80**, 2436.

4 S. Ehrenson, R. T. C. Brownlee, and R. W. Taft, *Prog. Phys. Org. Chem.*, 1973, **10**, 1.

5 M. C. Spanjer and C. L. de Ligny, submitted to *J. Chem. Res.*

6 H. H. Harman, 'Modern Factor Analysis,' University of Chicago Press, Chicago, 1970.

7 C. L. de Ligny, G. H. E. Nieuwdorp, W. K. Brederode, W. E. Hammers, and J. C. van Houwelingen, *Technometrics*, 1981, **23**, 91.

8 P. M. Kroonenberg and J. de Leeuw, *Psychometrica*, 1980, **45**, 69.

9 H. M. Weesie and J. C. van Houwelingen, GEPCAM Users's Manual, Institute of Mathematical Statistics, Utrecht, 1983 (available on request).

10 P. H. Weiner, *J. Am. Chem. Soc.*, 1973, **95**, 5845.

11 M. R. Bacon and G. E. Maciel, *J. Am. Chem. Soc.*, 1973, **95**, 2413.

12 C. L. de Ligny, N. J. Koole, H. D. Nelson, and G. H. E. Nieuwdorp, *J. Chromatogr.*, 1975, **114**, 63.

13 T. Fujita, J. Iwasa, and C. Hansch, *J. Am. Chem. Soc.*, 1964, **86**, 5175.

14 G. H. E. Nieuwdorp, C. L. de Ligny, and N. G. van der Veen, *J. Chromatogr.*, 1978, **154**, 133.

15 L. R. Snyder, 'Principles of Adsorption Chromatography,' Marcel Dekker, New York, 1968.

16 M. C. Spanjer and C. L. de Ligny, *J. Chromatogr.*, 1982, **253**, 23.

17 S Siegel and J. M. Komarmy, *J. Am. Chem. Soc.*, 1960, **82**, 2547.

18 J. F. J. Dippy and F. R. Williams, *J. Chem. Soc.*, 1934, 1888.

19 J. W. Baker, J. F. J. Dippy, and J. E. Page, *J. Chem. Soc.*, 1937, 1774.

20 J. F. J. Dippy, S. R. C. Hughes, and B. C. Kitchiner, *J. Chem. Soc.*, 1964, 1275.

21 A. J. Hoefnagel, J. C. Monshouwer, E. C. G. Snorn, and B. M. Wepster, *J. Am. Chem. Soc.*, 1973, **95**, 5350.

22 J. F. J. Dippy and J. E. Page, *J. Chem. Soc.*, 1938, 357.

23 R. Fuchs, C. A. Kaplan, J. J. Bloomfield, and L. F. Hatch, *J. Org. Chem.*, 1962, **27**, 733.

24 J. F. J. Dippy, F. R. Williams, and R. H. Lewis, *J. Chem. Soc.*, 1935, 343.

25 J. F. J. Dippy and R. H. Lewis, *J. Chem. Soc.*, 1936, 644.

26 L. G. Bray, J. F. J. Dippy, and R. S. C. Hughes, *J. Chem. Soc.*, 1957, 265.

27 D. H. McDaniel and H. C. Brown, *J. Org. Chem.*, 1958, **23**, 420.

28 W. L. Bright and H. T. Briscoe, *J. Phys. Chem.*, 1933, **37**, 787.

29 H. H. Jaffé, L. D. Freedman, and G. O. Doak, *J. Am. Chem. Soc.*, 1953, **75**, 2209.

30 G. Schwarzenbach and E. Rudin, *Helv. Chim. Acta*, 1939, **22**, 360.

31 P. de Maria, A. Fini, and F. M. Hall, *J. Chem. Soc., Perkin Trans. 2*, 1973, 1969.

32 A. R. Butler, *J. Chem. Soc. B*, 1970, 867.

33 A. Bryson, *J. Am. Chem. Soc.*, 1960, **82**, 4858.

34 G. Baddeley, J. Chadwick, and H. T. Taylor, *J. Chem. Soc.*, 1954, 2405.

35 A. de Courville, *Compt. Rend. Acad. Sci., Paris*, 1966, **262**, 1196.

36 M. C. Spanjer and C. L. de Ligny, *Chromatographie*, 1985, **20**, 120.

37 A. Leo, C. Hansch, and D. Elkins, *Chem. Rev.*, 1971, **71**, 525.

38 J. C. van Houwelingen, 'Principal Components of Large Matrices with Missing Elements,' Proceedings of the 3rd Prague Symposium on Asymptotic Statistics, 1983, Elsevier, Amsterdam, 1984, p. 295.

Received 29th August 1984; *Paper* 4/1487