Multivariate Data Analysis of Carbon-13 Nuclear Magnetic Resonance Substituent Chemical Shifts of 2-Substituted Naphthalenes

Dan Johnels, Ulf Edlund,* and Svante Wold
Department of Organic Chemistry and Research Group of Chemometrics, Institute of Chemistry, Umeå University, S-901 87 Umeå, Sweden

The ^{13}C n.m.r. chemical shifts of several 2-substituted naphthalenes have been analysed by principal components data analysis. In addition, the shift values have been related to the ^{13}C n.m.r. chemical shifts of the corresponding monosubstituted benzenes by means of partial least-squares data analysis. This latter analysis showed that nearly all of the systematic variation in the naphthalene shift data set could be predicted from the shift data of the monosubstituted benzenes. Moreover, it is shown that the grouping of substituent effects noticed for most benzene derivatives is also present in the naphthalene data. The advantage of using partial least-squares data analysis compared with the conventional dual substituent parameter analysis is also demonstrated. In fact, the overall description of the ^{13}C shifts using the σ_I, σ_R dual-substituent parameter correlation is not improved compared with the use of the mean values of σ_I and σ_R for the donor, acceptor, alkyl, and halogen subclasses. Finally, two existing methods for the determination of the relevant number of substituent parameters is discussed. A slightly conservative cross-validation method is argued to be better than methods where the variance is explained as being due to the experimental error.

The mechanisms for the transmission of substituent effects have been an important aspect of physical organic chemistry for a long time.[1,2] A major breakthrough for these types of studies came with the improvements in ^{13}C FT-n.m.r. technology in the early 1970s. By measuring ^{13}C substituent chemical shifts (SCS) not only at certain probe positions, but also along the molecular framework at intermediate positions, multivariate information about transmission mechanisms is obtained.

Unfortunately, the problem is normally approached by considering one position at a time. Hence, the shift data are usually analysed using a fixed, predetermined dual-substituent parameter (DSP) model such as in equation (1), where σ_R is one out

$$SCS = \rho_I\sigma_I + \rho_R\sigma_R \quad (1)$$

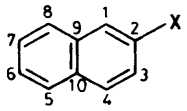
of four mesomeric scales (σ_R^-, σ_R^0, σ_R^{BA}, or σ_R^+). The practice is to perform four separate correlations, each with a different resonance scale, and then pick the scale that affords the best fit to the experimental shifts.[1e,2g]

However, as we previously pointed out, there are certain limitations with the use of multiple regression (MR).[3] First, it must be assumed that all independent variables (σ values) are precisely known and relevant to the shift variables studied. Secondly, MR analysis demands almost orthogonal substituent scales, if the regression coefficients, in this case called transmission coefficients (ρ), should be used to separate the contributing effects. The third problem is concerned with the choice and spread of substituents. This topic has been considered by some workers in this field, and several substituent basis sets have been suggested, representing as wide a range as possible of the substituent categories.[2a-c] However, the most commonly quoted minimum substituent basis sets are all very similar and contain only second-row elements, except for one halogen (Cl or Br) and hydrogen. Other non-second row elements are only included occasionally, but a misfit of one or two such substituents would be hard to detect using MR. If the regression parameters are to be used for interpretation purposes, a fundamental requirement of any statistical model is the existence of a homogeneously distributed data set. Hence, before performing the data analysis, it should be checked that the data are uniformly and continuously distributed in the variable space. If data are clustered, separate local models should preferably be used for each cluster or subset.

In an early study of the ^{13}C n.m.r. SCS of 4-substituted styrenes, we analysed the SCS data by a principal component (PC) data analytical method using all the reported shift data of 4-substituted styrenes.[4] The choice of n.m.r. variables was based on two objectives: (1) to separate three classes of 4-substituted styrenes (α-H, α-Me, and α-Bu^t, seven derivatives in each class); and (2) to study to what extent each variable takes part in the modelling of each class.

The C-1, C-β, and the two β-hydrogens were the variables that best fulfilled these demands. One component (substituent parameter) was found to be statistically significant by cross-validation. This work was later criticised[5] and it was claimed that the loss of the expected second component was caused by the applied preprocessing of our n.m.r. data (global scaling, *i.e.*, giving each variable a variance equal to unity). Recently, this argument was reconsidered by the same authors[6] and the data matrix used in our study was re-analysed using a PC method. It was argued that scaling was unimportant for the discrepancies between these two analyses. The major cause of the discrepancy was suggested instead to be the way the number of components was determined and, that in our analysis, with the limited number of derivatives at hand, hydrogen was excluded ($\Delta\delta$ values were used) and Bu^t was chosen instead of CN in the substituent set. In this paper, we consider this criticism in more detail and explain the different philosophies behind the two different methods used to determine the number of relevant factors: cross-validation and percentage trace (PT) methods as advanced by Malinowski and Howery.[7]

The second proposition indicates a heavy dependence of the number of significant components in the resulting model on the choice of substituents. This has already been discussed, to some extent, using a larger shift matrix, the ^{13}C chemical shifts of monosubstituted benzenes.[3a] The initial hypothesis was that the indicated dependence of the parameterization on the selection of substituents in the styrene might be caused by a grouping of the SCS data. If one group, such as acceptors, is underrepresented, this might explain the collapse of the DSP model into a single-parameter model. In a similar vein, the generality of DSP models to account for the SCS in alternate conjugated



(1)

systems could have been biased by a selection of substituents, mainly from three substituent clusters, alkyls, donors, and acceptors. Three clusters in an M-dimensional variable space can always be explained by a plane, *i.e.*, by a two-parameter model.

A PC analysis was thus performed on the chemical shifts of 82 monosubstituted benzenes,[3a] where only second-row elements were selected to be the directly substituted atoms, except for some halogens and hydrogen. *ca.* 90% of the substituents were shown to belong to four clusters, acceptors, donors, alkyls, and halogens. Most subclasses showed extensions in space that were not parallel. In fact, single-parameter models for the subclasses gave better predictions than any two- or three-component models based on the total data set (global model).

The objective of this paper is to extend this study to another alternant system, where more remote positions can be included. It is of interest to see whether the clustering observed in the benzene data is also prevalent in this system, and, if so, whether this grouping can be predicted from the ^{13}C n.m.r. SCS in the benzenes. As already mentioned, we will also discuss our views on variable scaling in multivariate data analysis and some approaches to the determination of the appropriate number of components or parameters in the applied model.

Methods

Choice of Data.—The ^{13}C n.m.r. chemical shifts of 21 2-substituted naphthalenes and the corresponding monosubstituted benzenes were taken from the literature. The substituents in (1) used were in the order: H, Me, Bu^t, CH_2Br, F, Cl, Br, I, COMe, CHO, CO_2H, CN, NO_2, NH_2, NMe_2, OMe, OH, NHCOMe, and OCOMe. The shift values of the 2-substituted naphthalenes were collected from reference 8*b*, except for the following: Bu^t and CO_2H from reference 8*a*; CHO from reference 8*c*; and CH_2Br from reference 8*d*. All shift data of monosubstituted benzenes were from reference 2*d*.

The shift data include all reported derivatives to date conforming to the selection criteria discussed below. Shift values obtained in $CDCl_3$ were used throughout the data analysis. In a more inert medium, such as cyclohexane, aggregation effects and solubility limitations may affect the accuracy of the chemical shift values. To test the relevance of the clustering found for the ^{13}C SCS data of monosubstituted benzenes, the substituents were selected as in the former analysis.[3a] Except for H, Cl, Br, and I only such substituents have been included that have second-row elements as the directly attached atoms. The choice conforms to the pattern given in the minimum basis sets and our data set contains most of the substituents used in linear free energy relationships (LFER).[9]

Data Analysis.—Principal components analysis. Principal components data analysis[10] has been used to model the systematic ^{13}C SCS of the 2-substituted naphthalenes [equation (2)]

$$y_{ik} = \bar{y}_i + \sum_{a=1}^{A} c_{ia}u_{ak} + e_{ik} \qquad (2)$$

where y_{ik} is the scaled SCS value of variable i (position in the naphthalene system) and object k (substituent), $\bar{y}_i$ is the mean of variable i, c_{ia} is the component loading, and u_{ak} is the score or component value. The c_{ia} and u_{ak} are estimated by minimizing the sum of squared residuals, Σe^2_{ik}. The appropriate number of terms, A, is estimated by a cross-validation procedure described below. The resulting components will, by definition, be orthogonal, *i.e.*, if interpreted as 'substituent effects', truly independent.

In a geometric description, each 2-substituted naphthalene is represented by a point in an M-dimensional space spanned by the ten ^{13}C SCS axes. A PC model with $A = 1$ then constitutes a line, while a model using $A = 2$ will be represented by a plane in this SCS M-space.

Partial least-squares analysis. In the partial least-squares (PLS) data analysis,[11] the scaled data were divided into two matrices, X (^{13}C SCS of monosubstituted benzenes) and Y (^{13}C SCS of 2-substituted naphthalenes) and the two X and Y blocks were then related to one another in a PC-like manner. The naphthalene matrix is modelled as in equation (2) and the benzene matrix X is modelled as in equation (3).

$$x_{jk} = \bar{x}_j + \sum_{a=1}^{A} b_{ja}t_{ak} + e_{jk} \qquad (3)$$

The two matrices are related as given in equation (4) where r_a

$$u_{ak} = r_a t_{ak} + f_{ak} \qquad (4)$$

is the least-squares inner regression coefficient and f_{ak} the residual. In normal PC data analyses, separately performed on Y and X, the t and u vectors are eigenvectors to XX' and YY', respectively, while in the PLS analysis, these vectors are eigenvectors to the matrices $YY'XX'$ and $XX'YY'$. This approach introduces a relation between the two SCS blocks, while still preserving most of the PC projection properties of the resulting components. As in PC, the relevant number of components, A, is determined by cross-validation.

Preprocessing of Shift Data for PC Analysis.—Bilinear least-squares methods, such as PC analysis, are dependent on the scaling of the analysed variables.[12] The applied scaling should reflect the information in each variable. This information is a function of both the experimental error of the variable and its coherence to the common pattern of all the variables. A global scaling where each variable is given a variance of unity over the whole data set is by far the most commonly applied scaling using PC methods. Reynolds *et al.*[6] scaled the data so that the variables have equal 'experimental' accuracy (see below). It should be mentioned that factor analyses, as defined by statisticians, are scaling independent.[13]

In this analysis, where the experimental errors are the same for each variable, global scaling will introduce noise into the analysis, owing to the increased importance of variables having a small initial variance. In spite of this limitation, this is done to avoid a dominating influence of the *ipso*- and *ortho*-like positions. These positions have the largest initial variance, which can hide systematic variation in the other positions. We have retained the scaling parameters from the global analysis in the analysis of the subclasses. This was done to facilitate a comparison between the global and subclass modelling. The significance of clustering would be further enhanced if the subclasses were scaled separately. We mention at this point that the preprocessing of observed data is connected to the criteria used to determine the appropriate number of parameters in the model used.

Determination of the Relevant Number of Components.—In all our PC analyses, a cross-validation method has been used to estimate the number of significant components. This approach differs fundamentally from the PT methods based on the residual standard deviation that have recently been sugges-

ted.[6,7] The latter descriptive methods correspond to the use of enough components to reproduce the data matrix down to the experimental error, thereby justifying the choice of the aforementioned scaling,[6] *i.e.*, giving equal 'experimental' accuracy to the variables. The cross-validation method is, contrary to the PT criteria, based on the predictive ability of the model. Consequently, if the aim is to find models that can be used to predict properties, *i.e.*, reactivity, chemical shifts, *etc.*, of new compounds, cross-validation is the preferred approach.[14] In the physical organic field, as in other areas of chemistry, the essential demand on a model, whether empirical or theoretical, is that it should have the power to predict properties of new compounds in the series.

The experimentalist normally has a tendency to underestimate the experimental error and, more seriously, often ignores the fact that there always exists a model error. This emphasizes the need to use a slightly conservative criterion, that does not lead to the inclusion of components that only account for random noise. The danger with PT criteria, leading to models that explain the variance down to the experimental error, is thus obvious. Especially in the LFER area, where the flood of substituent scales is more than adequate, PT criteria could be counterproductive. Several new 'effects' are likely to be proposed since many short-range substituent effects are substituent specific (anisotropy, steric, 'heavy atom' effects, *etc.*). Moreover, spurious or misassigned peaks, aggregation, or steric effects in a minority of compounds might show up as new 'effects'. In this context it must be stressed that all parameters claimed to be of chemical significance must also be statistically significant.

The cross-validation in the PC analysis proceeds as follows.[15] A group of data elements, with representatives from all rows and columns in the data matrix, is left out from the data matrix, for instance 1/4 of the matrix. With these elements held out, the PC parameters with a given number of components A are estimated from the remaining data. The estimated parameters are then used to calculated predictions for the deleted elements. The procedure is repeated by excluding another part of the matrix, and estimating parameters from the remaining data matrix, *etc.* Enough rounds are made (in this case four) as one needed to keep each data element out once and only once. Hence, a measure of the predictive error for the PC model for a given A is then obtained. This is compared with the residual standard deviation with $A = A - 1$. If the prediction error is significantly smaller than the residual standard deviation, this component is added to the model. This procedure of adding components one at a time is repeated for as long as the components are statistically significant.

A similar scheme is used in the cross-validation in the PLS analysis.

Results and Discussion

PC Analysis of SCS Data of 2-Substituted Naphthalenes.—Initially the ^{13}C n.m.r. SCS of 2-substituted naphthalenes were analysed by the PC method. According to cross-validation, three components were needed to model the SCS data ($A = 3$). This global model describes 87% of the chemical shift variance. By comparing the component plots in the naphthalene system with the corresponding plot for monosubstituted benzenes, a similarity in clustering is indicated (Figure 1).

In order to test if this clustering is statistically significant, the mean values were calculated and a comparison was made between the residuals from: (1) the whole naphthalene data set, except such substituents that were found to be outliers in the PC analysis of the benzenes (numbers 1, 14, 15, 20, 21);[3a] and (2) the four subsets formed by separating the SCS into

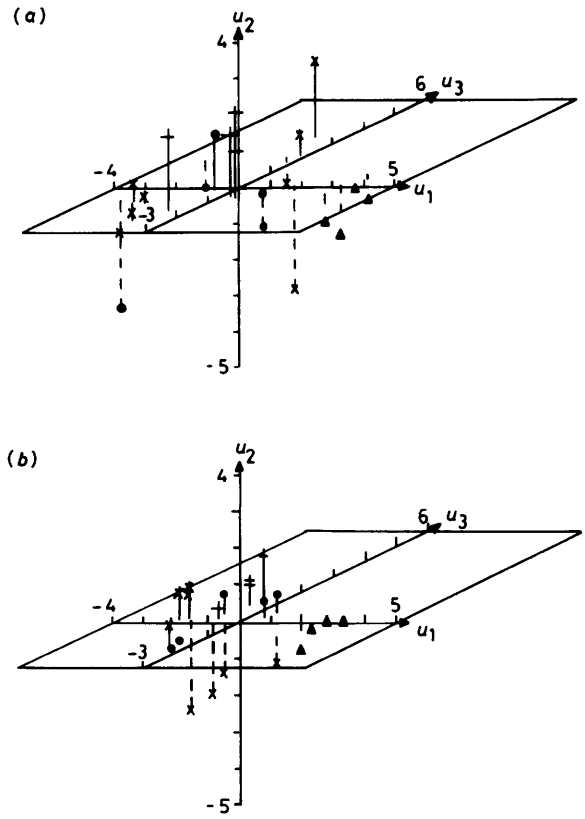


Figure 1. (*a*) Plot of the three significant components from the PC model derived from the SCS data of the 2-substituted naphthalenes. (*b*) Plot of the three significant components from the PC model derived from the SCS data of the corresponding monosubstituted benzenes. (✱) acceptors, (+) alkyls, (▲) donors, (×) halogens, and (●) 'outliers'

alkyls (numbers 2—5), halogens (6—9), acceptors (10—13), and donors (16—19).

The residual variance in these two analyses was then compared by an approximate F-test, equation (5).

$$F = s_1^2(\text{whole})/s_2^2(\text{pooled}) = 3.06 \qquad (5)$$
$$F_{\text{crit.}} = 1.53 \quad (p = 0.01)$$

A statistically significant clustering is also thus verified in the case of 2-substituted naphthalenes. Contrary to the prior analysis of the benzene data set, the number of substituents in each subclass is too few at present to justify a derivation of local PC models for each subset.

PLS Relationship between Substituted Naphthalenes and Benzenes.—A single global model. It would also be interesting to know whether the substituent positions in the naphthalene ^{13}C SCS space could be predicted from the SCS behaviour in the benzenes. To relate these two data matrices, we used the PLS method described above. As a first step, the data set formed by all 21 2-substituted naphthalenes and the corresponding benzene data set were compared. 83% of the variance in the naphthalene shift data was described by the three-component model derived from the PLS analysis ($A = 3$) (compared with 87% in the PC analysis). In other words, nearly all the systematic information explained by the PC model in the naphthalene SCS can be predicted from the benzene data. This indicates that the relative location of the clusters in the naphthalenes is similar to that in the benzenes. The 'outliers'

Table. Pooled and residual variances of the PLS and DSP analysis

	Position										
Model	1	2	3	4	5	6	7	8	9	10	S_{pooled}^2
PLS_{local}	0.03	0.05	0.04	0.05	0.42	0.03	0.09	0.05	0.03	0.03	0.101 [a]
DSP [b]				0.26	0.73	0.02	0.38	0.13	0.26	0.02	0.256
DSP_{mean} [c]				0.27	0.63	0.09	0.36	0.12	0.18	0.06	0.242

[a] Calculated for positions 4—10. [b] σ_I and σ_R^0 values were taken from reference 17. [c] The mean values of σ_I and σ_R^0 of each class were used.

also have the same positions relative to the clusters in the naphthalene space as in the benzene space. This similarity in substituent behaviour in aromatic systems has been utilised to predict ^{13}C n.m.r. shifts in 2-substituted naphthalenes and in 4-substituted *p*-terphenyls using the shifts of the corresponding monosubstituted benzenes.[16]

It must be stressed, however, that both CN (number 14) and NO_2 (number 15) were found within the acceptor group if proximate positions, the *ipso*- and *ortho*-like positions, were excluded in the analysis. It is known that magnetic anisotropy effects influence the positions close to these substituents.

Separate local PLS models of each subset. Having confirmed the similar substituent clustering in the benzene and naphthalene SCS data, we wanted to investigate the relationship between the SCS within each subclass in the two aromatic systems. The ^{13}C SCS data were thus divided into four subsets, as above, each described by a single component ($A = 1$).

The descriptions of the naphthalene ^{13}C SCS data by these four local models and by a global three-component model were subsequently compared by an approximate *F*-test, equation (6),

$$F = s_y^2(\text{whole})/s_y^2(\text{pooled}) = 1.99 \qquad (6)$$
$$F_{crit.} = 1.84 \quad (p = 0.01)$$

i.e., a better description using simple 'one effect' ($A = 1$) local models is obtained compared with a 'complicated' ($A = 3$) global model. This means that the positions of the substituents within each cluster is similar in both the naphthalene and benzene space. However, it should be pointed out that the local models are only approximate descriptions of the local substituent behaviour, owing to the limited number of substituents in each subclass.

Comparison between Local PLS Models and a Global DSP Model.—To demonstrate the potency of the present PLS approach in comparison with the more conventional DSP analysis, we have compared the application of these two methods to the scaled naphthalene data.

Since DSP models cannot account for substituent effects in positions proximate to the substituents, the *ipso*- and *ortho*-like positions were excluded. If these positions are differently affected, this will disfavour PLS relative to DSP, since they were included in the derivation of the PLS models. The fact that the significance of the parameters in the DSP model was disregarded will also favour the DSP model.

An inspection of the Table reveals that the PLS method gives a better description of all but two positions (numbers 6 and 10), where the DSP model gives a slightly, but not significantly, better description. The *F*-test of the pooled variances is given in equation (7). Hence the local PLS models are

$$F = S_{DSP}^2/S_{PLS}^2 = 2.53 \qquad (7)$$
$$F_{crit.} = 1.84 \quad (p = 0.01)$$

significantly better than the DSP model.

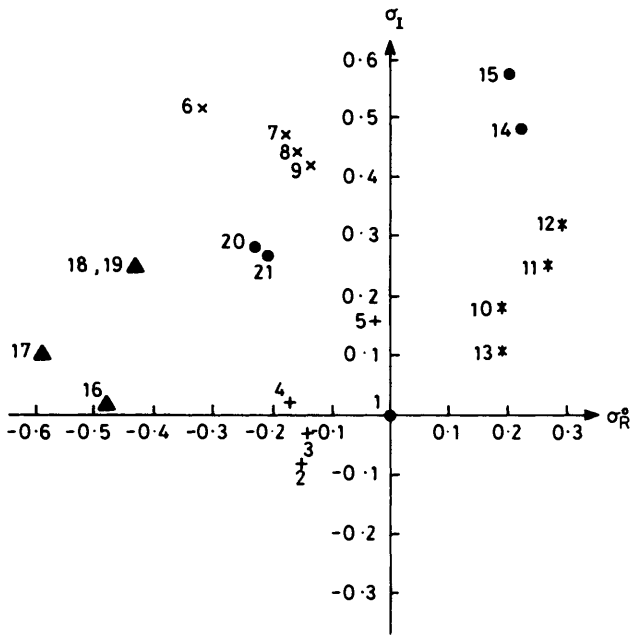


Figure 2. Plot of σ_I *versus* σ_R^0 for the 21 substituents used. The same symbols are used as in Figure 1

Relevance of the DSP Parameters for the Intra-class Substituent Effects.—To test to what extent the σ_I and σ_R values model the subgroup behaviour, *i.e.*, the differences in substituent effects within each substituent cluster, such as donors, acceptors, *etc.*, we have compared the conventional DSP treatment with the simplified 'clustering' model obtained by only using the mean of σ_I and σ_R for each subclass. The residual variance of the two treatments in each position is reported in the Table. Interestingly, an *F*-test clearly shows that there is no significant improvement in the overall modelling by the use of the σ_I and σ_R for the given substituent set, relative to the use of the mean values of σ_I and σ_R in each subclass, equation (8).

$$F = S_{DSP}^2/S_{DSP}^2(\text{mean}) = 1.06 \qquad (8)$$
$$F_{crit.} = 1.65 \quad (p = 0.01)$$

These observations confirm our earlier findings that the existence of substituent clustering must be born in mind using MR methods, as with DSP. The correlation between ^{13}C SCS obtained for various positions within a given structure could as a whole or in part be caused by a grouping of substituent effects. If a majority of the substituents actually studied were found to be within three classes, a two-parameter model defining a plane through the class means will approximately describe these classes.[3a] In such a case, and where the σ_I and σ_R values do not contain information of the intra-class behaviour, the use of the

mean values of σ_I and σ_R should give a model almost as potent. This is observed in the analysis above.

It is often argued that the σ_I and σ_R scales are truly independent.[6] This statement is rather meaningless if substituents are clustered in their interactions with the molecular system. It is however important to show that this independence also persists if one or two of the indicated subclasses are deleted from the analysis, or more drastically, that the 'fundamental' properties of σ_I and σ_R also could be identified in each subclass. A plot of these scales as in Figure 2 does not clearly reveal such an independence of the field and mesomeric effects within the subclasses.

Conclusions

To conclude, we have shown that the ^{13}C SCS of 2-substituted naphthalenes are clustered and positioned in the shift space in an analogous way to the corresponding monosubstituted benzenes. Thus, the use of statistical methods, aimed at relating matrices of this kind, such as PLS, is far better to predict ^{13}C SCS in general than the more conventional DSP approach. We note that both the PLS and DSP analyses need a 'calibration set' of compounds to determine the model parameters, *b*, *c*, and *r* in PLS, and ρ_I and ρ_R in DSP. Hence, the DSP analysis is not more 'fundamental' than PLS from the theoretical information point of view.

Finally, we have discussed the methods used to determine the relevant number of parameters in the substituent effect models. A descriptive philosophy where the aim is to explain all variance as being due to the experimental error will most likely create more 'effects' than an approach based on cross-validation. In the latter method the predictive power of the model is actually tested. Using PT criteria, there is a risk that model errors, underestimated experimental errors, misassigned peaks, unforeseen solvent or bulk effects, anisotropy effects, *etc.*, will appear as extra parameters.

If, for interpretative purposes, a universal model is desirable, clustering of SCS should be avoided, still maintaining maximum spread in substituent behaviour. At least 20% of the analysed SCS data should be taken from 'non-second row' derivatives, if the model is claimed to be of any generality, *i.e.*, hydrogen and substituents having sulphur atoms, Group IV elements *etc.* directly attached to the carbon framework should be included. This problem will be considered in a future paper.

An easy test to confirm the absence of grouping would be to exclude donors or acceptors from the analysis once, to confirm that the transmission coefficients (ρ) do not vary significantly. If they are unchanged upon such deletion, the generality of the found effects is more reliably demonstrated.

Acknowledgements

Grants from the Swedish Natural Science Research Council and the Swedish Board for Planning and Coordination of Research are gratefully acknowledged. We are indebted to Professor W. F. Reynolds for critical discussions and correspondence.

References

1 (*a*) D. F. Ewing, in 'Correlation Analysis in Chemistry,' eds. N. B. Chapman and J. Shorter, Plenum Press, New York, 1978, chap. 8; (*b*) M. T. Tribble and J. G. Traynham, in 'Advances in Linear Free Energy Relationships,' eds. N. B. Chapman and J. Shorter, Plenum Press, New York, 1972, ch. 4; (*c*) R. D. Topsom, *Prog. Phys. Org. Chem.*, 1976, **12**, 1; (*d*) G. L. Nelson and E. A. Williams, *ibid*, 1976, **12**, 229; (*e*) J. Bromilow, R. T. C. Brownlee, D. J. Craik, P. R. Fiske, E. J. Rowe, and M. Sadek, *J. Chem. Soc., Perkin Trans. 2*, 1981, 753; (*f*) D. J. Craik and R. T. C. Brownlee, *Prog. Phys. Org. Chem.*, 1983, **14**, 1.

2 (*a*) W. Adcock and T.-C. Khor, *J. Am. Chem. Soc.*, 1978, **100**, 7799; (*b*) R. T. C. Brownlee and D. J. Craik, *Org. Magn. Reson.*, 1981, **15**, 248; (*c*) J. Bromilow, R. T. C. Brownlee, V. O. Lopez, and R. W. Taft, *J. Org. Chem.*, 1979, **44**, 4766; (*d*) D. F. Ewing, *Org. Magn. Reson.*, 1979, **12**, 499; (*e*) E. A. Hill and H. E. Gunther, *ibid.*, 1981, **16**, 177; (*f*) G. K. Hamer, I. R. Peat, and W. F. Reynolds, *Can. J. Chem.*, 1973, **51**, 897; (*g*) N. K. Wilson and R. D. Zehr, *J. Org. Chem.*, 1982, **47**, 1184.

3 (*a*) D. Johnels, U. Edlund, H. Grahn, S. Hellberg, M. Sjöström, S. Wold, S. Clementi, and W. J. Dunn III, *J. Chem. Soc., Perkin Trans. 2*, 1983, 863; (*b*) S. Alunni, S. Clementi, U. Edlund, D. Johnels, S. Hellberg, M. Sjöström, and S. Wold, *Acta Chem. Scand., Ser. B*, 1983, **37**, 47.

4 U. Edlund and S. Wold, *J. Magn. Reson.*, 1980, **37**, 183.

5 W. F. Reynolds, P. Dias, D. W. MacIntyre, G. K. Hamer, and I. R. Peat, *J. Magn. Reson.*, 1981, **43**, 81.

6 W. F. Reynolds, A. Gomes, A. Maron, D. W. MacIntyre, R. G. Maunder, A. Tanin, H. E. Wong, G. K. Hamer, and I. R. Peat, *Can. J. Chem.*, 1983, **61**, 2367.

7 E. R. Malinowski and D. G. Howery, in 'Factor Analysis in Chemistry,' Wiley Interscience, New York, 1980.

8 (*a*) W. Kitching, M. Bullpitt, D. Gartshore, W. Adcock, T. C. Khor, D. Doddrell, and I. D. Rae, *J. Org. Chem.*, 1977, **42**, 2411; (*b*) H. Takai, A. Odani, and Y. Sasaki, *Chem. Pharm. Bull.*, 1978, **26**, 1966; (*c*) J. Seita, J. Sandström, and T. Drakenberg, *Org. Magn. Reson.*, 1978, **11**, 239; (*d*) M. Bullpitt, W. Kitching, D. Doddrell, and W. Adcock, *J. Org. Chem.*, 1976, **41**, 760.

9 G. W. Klumpp, in 'Reactivity in Organic Chemistry,' Wiley Interscience, New York, 1982.

10 For a recent review see: S. Wold, C. Albano, W. J. Dunn III, U. Edlund, K. Esbensen, P. Geladi, S. Hellberg, E. Johansson, W. Lindberg, and M. Sjöström, in 'Proceedings NATO Advanced Study Institute on Chemometrics, Cosenza, Italy, September 1983,' ed. B. R. Kowalski, Riedel, Dardrecht, Holland, 1984.

11 H. Wold, in 'Systems under Indirect Observation,' eds. K. G. Jöreskog and H. Wold, North Holland, Amsterdam, 1982.

12 M. P. Derde, D. Coomans, and D. L. Massart, *Anal. Chim. Acta*, 1982, **141**, 187.

13 K. V. Mardia, J. T. Kent, and J. M. Bibby, in 'Multivariate Analysis,' Academic Press, New York, 1979.

14 H. T. Eastment and W. J. Krzanowski, *Technometrics*, 1978, **24**, 73.

15 S. Wold, *Technometrics*, 1978, **20**, 397.

16 D. Johnels, U. Edlund, E. Johansson, and S. Wold, *J. Magn. Reson.*, 1983, **55**, 316.

17 The substituent constants were collected from the compilation of O. Exner, in 'Correlation Analysis in Chemistry,' eds. N. B. Chapman and J. Shorter, Plenum Press, New York, 1978, ch. 10.

Received 18*th September* 1984; *Paper* 4/1618