Acta Chemica Scandinavica B 37 (1983) 47–53

Multivariate Data Analysis of Substituent Descriptors

SERGIO ALUNNI,[a] SERGIO CLEMENTI,[a] ULF EDLUND,[b] DAN JOHNELS,[b] SVEN HELLBERG,[b] MICHAEL SJÖSTRÖM[b] and SVANTE WOLD[b]

[a] Department of Chemistry, University of Perugia, Perugia, Italy and [b] Department of Organic Chemistry and Research Group for Chemometrics, Institute of Chemistry, Umeå University, S-901 87 Umeå, Sweden

A matrix containing seven often used substituent descriptors (Hammett sigmas *etc.*) for twenty-eight substituents is studied by multivariate statistical analysis.

The results show: (a) Strong grouping of the substituents into four separate classes; alkyls, donors, acceptors and halogens. (b) Separate models for the classes are superiour for describing the intra class structures compared to a whole set model. (c) A high collinearity between some of the descriptors.

As discussed, the found grouping and the collinearity can be limiting factors in the use of multiple regression in quantitative structure-activity and structure-reactivity studies.

In some of our previous work we investigated the application of multivariate statistics in physical organic chemistry. Substituent effects[1–6] in linear free energy relationships (LFERs), has been an area where statistical tools have shown to be valuable.[7–9] Thus, a principal component (*PC*) model with one component ($A=1$ in eqn. (1) below) has the same form as a simple linear regression equation ($M=1$ in eqn. (7) below) and as the Hammett equation.[10] Hence a *PC* model has been used to derive a unified substituent scale for isolated benzene systems, independently of a single reference reaction.[2]

The main difference between the linear regression and the *PC* analysis lies in the assumptions about the substituent parameters. In the former analysis, the values of the independent variable(s) x_i (the substituent constants) are assumed to be exactly known and 100 % relevant to the description of the data set under examination.[11] In the latter approach no assumption about the relevance of the variables x_i is required, since this relevance is obtained from the statistical analysis.

The two philosophies diverge even more when the number of explanatory variables is increased, as, for instance, in the multiple regression analysis (MRA) of dual or multiple parameter equations [see eqn. (7)]. To get statistically sound results, the application of MRA requires that the x variables are independent and fairly non-collinear of each other. However collinearity can be tested for. In contrast, *PC* analysis is insensitive to collinearities and in fact uses them to estimate the components θ. In other words in MRA one first needs to define "fundamental" effects. Thereafter one tries to interpret the results in terms of the definitions previously agreed. With *PC* analysis no such definitions are required. The results are usually interpreted just in terms of the components needed to model the data set. They can also be related to the effects currently believed to be measured by the variables used in the input.

A further statistical problem with MRA is the heavy dependence of the results upon the number of observations in relation to the number of independent variables. As discussed by Topliss and Edwards[12] this problem also arises when variables are selected or screened from a larger ensemble. Thus when the number of screened independent variables exceeds the number of observations, the risk for spurious correlations is serious. For example, with 5 independent variables and 10 observations, the probability for chance correlation, P_c, ($r^2 > 0.8$) is 0.05. If the number of screened variables is increased to 10, then $P_c = 0.30$. This problem is amplified if the independent and independent variables are grouped in the same way. If we have, say, twenty observations grouped in five classes the chance correlation approaches the case with five observations. We note (see Ref. 12) that with five observa-

0302-4369/83 $2.50

tions and five variables $P_c = 0.30$, but with twenty observations and five variables the chance correlation is negligible.

We have previously pointed out the problem of grouping for the substituent constants measuring the inductive and mesomeric effects. In the two-dimensional space defined by σ_I and σ_R the points for some twenty-four of the most common substituents are located in such a way that *ca.* 50 % of the points lie within a narrow area, and the others are spread out in various directions.[5,6] Owing to this situation, it follows that a certain number of appropriate substituents has to be used in MRA to obtain significant results.[13]

In order to investigate whether the same situation also applies to the case with an increased number of variables (substituent scales), we were tempted to make a *PC* analysis of a matrix containing some of the mostly used descriptors of substituents. In this report we describe the result obtained by applying the *PC* model to a data set formed by values of seven variables $\sigma_m^{0\,2d}$, $\sigma_p^{0\,2d}$, $\sigma_R^{0\,14}$, σ^+ or σ^{-15}, E_S^{16}, π^{16}, and molar refractivity (MR)[16] for twenty-eight common substituents. Thus the variables used are measures of various electronic, steric and polarizability effects.

METHODS

The *PC* method is presented in detail in Refs. 1, 2, 7, 17 and 18. Therefore, we will limit the presentation to a brief summary. For an introduction to multiple regression see Ref. 19.

Principal components analysis

The descriptor matrix Y contains the elements y_{ik}, where index i is used for the descriptors (variables) and index k for the substituents (objects). From this data matrix the number of cross-terms, A, and then the parameters α_i, β_{ia} and θ_{ak} in eqn. (1) are estimated by minimizing the squared residuals ε_{ik}.

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

In this model α_i and β_{ia} are constants, which only are dependent on the descriptors and θ_{ak} are the substituent dependent parameters. The deviations from the model are expressed by the residuals ε_{ik}.

Before applying any statistical analysis, the descriptor values were auto-scaled. This means that the variables are given the same variance (fixed to unity). With this scaling all variables are given the same importance in the *PC* analysis.

First a model with $A = 0$ is fitted to the data, which means that each descriptor is given as its mean value α_i. Then the α_i value for each variable is subtracted from the matrix elements y_{ik} thus giving the residuals of dimension zero. If these residuals now contain systematic information, the $\beta_{i1}\theta_{1k}$ term is estimated. Whether the residuals contain information or not is determined by cross-validation [for the details see Ref. (18)]. Then new residuals are calculated by subtracting the term $\beta_{i1}\theta_{1k}$. If the new residuals contain systematic information additional $\beta_i\theta_k$ terms are then estimated one after the other, until the residuals just contain noise.

After a model has been determined with autoscaling, it can be refined by a reweighting of the variables, in this case by multiplying each variable with its modelling power ψ_i, defined in eqn. (2).

$$\psi_i = (1 - s_i/S_{iY}) \qquad (2)$$

Here s_i and S_{iY} are the residual standard deviations for variable i with A significant components and with $A = 0$, respectively. This means that variables for which the $\beta\theta$ terms contain no or little information, will have modelling powers close to zero. Thus with this type of reweighting, such variables are given small weights.

Once a class model is determined, a data vector y_{ip} of a substituent p can be fitted to the class parameters by MRA as in eqn. (3). The class model is characterized by the α_i and β_{ia} parameters estimated from a data set with M variables and N substituents.

$$y_{ip} - \alpha_i = \sum_{a=1}^{A} t_{ap}\theta_{ak} + e_{ip} \qquad (3)$$

How well the data vector for the substituents fits the model is expressed by the residual standard deviation s_p in eqn. (4). By comparing the size of

$$s_p = \left[\sum_{i=1}^{M} e_{ip}^2/(M - A) \right]^{1/2} \qquad (4)$$

s_p^2 with S_A^2 [eqn. (5)], with the F-test in eqn. (6), we can decide if a substituent data vector belongs to a certain class or not. S_A is the total residual standard deviation for a class with A significant components.

$$S_A=\left[\sum_i\sum_k \varepsilon_{ik}^2/(N-A-1)(M-A)\right]^{1/2} \quad (5)$$

$$F=s_p^2/S_A^2 \quad (6)$$

Multiple regression analysis

In the multiple regression model (7) a data vector y_k (the dependent variable) is fitted to a fixed number of M independent variables x_i. The model

$$y_k=c_0+\sum_{i=1}^{M} c_i x_{ik}+e_k \quad (7)$$

is characterized by the regression coefficients c_0 and c_i ($i=1-M$). For example, in structure-activity or structure-reactivity studies, y_k is a vector consisting of activities or reactivities for a set of similar compounds modified by changing substituents. The independent variables are then a set of substituent descriptors like those analyzed in this paper. Note that once a *PC* model is determined, the new θ_{ak} substituent descriptors can be used in eqn. (7) as independent variables instead of x_{ik}. The advantage is that usually $A \ll M$ and that the θ_{ak} parameters are orthogonal to each other thus avoiding collinearity problems.

RESULTS AND DISCUSSION

The application of a *PC* model to the whole data set shows that only two components are significant according to the cross-validation. *Ca.* 82 % of the total variance is described by the two components model. For the the optimized model, the parameters specific for the variables (α_i and β_{ia}) are given in Table 1 and the parameters specific for the substituents (θ_{ak}) are given in Table 2. In Fig. 1, the β_{i1} parameters are plotted against β_{i2} for each variable. It is noteworthy that the first component contains mainly the contributions of the electronic variables (seen from their high absolute values of β_{i1}) and the second component contains mainly information from the remaining three variables (seen from their high absolute values of β_{i2}). Fig. 1 also indicates the degree of collinearity between the variables. Variables highly correlated to each other lie either very near each other ($r \approx 1$) or in a symmetrical position with respect to the center of the diagram ($r \approx -1$). In the present case, we observe that the four electronic descriptors all lie within a small range, thus pointing out the redundancy of the σ scales for this data set, whereas E_S contains some non-redundant information and both π and MR fall in a third region of the diagram.

A plot of the values of the two components, θ_{1k} and θ_{2k} for each substituent, *i.e.* the values that each substituent assumes along the new dimensions of the reduced 7-dimensional space, is shown in Fig. 2. It clearly indicates that the substituents do not constitute a single homogeneous class, but are strongly grouped according to their chemical nature. Especially the first component, containing the "electronic" variables, seems to discriminate between the classes. Four separate subsets of substituents can be recognized: alkyls, halogens, acceptors and donors. Such strong grouping into four classes was also recognized in a multivariate analysis of ^{13}C NMR shifts of more than seventy monosubstituted benzenes.[20]

We note that Hansch *et al.*[21] have investigated a similar data set consisting of 8 variables and 90 common substituents by an hierarchical clustering analysis. The aim was to find substituents with similar properties. However, the number of clusters was determined in advance. Different analyses with 5, 10, 20 and 60 globular clusters were performed. This explains the difference between their results and ours. For example, in their analysis such diverse substituents as alkyls, donors and acceptors can be found within the same cluster. The present overall analysis shows that most of the variance in the data is described by a two-components model. Therefore, the grouping (clustering) of the substituents can be evaluated directly from Fig. 2, and no initial assump-

Table 1. Model parameters (α_i, β_{i1} and β_{i2}) for the whole data set.

	σ_m^0	σ_p^0	σ_R^0	$\sigma^{+/-}$	E_S	π	*MR*
w_i^a	2.80	2.23	2.03	1.06	0.537	0.410	0.056
α_i	0.719	0.415	−0.207	0.056	−0.789	0.117	0.676
β_{i1}	−0.460	−0.547	−0.314	−0.596	0.149	0.113	0.032
β_{i2}	−0.167	−0.040	0.182	0.084	−0.577	0.499	0.590

[a] The weights for the optimized model. The weights after autoscaling are; 4.30, 3.02, 4.06, 1.32, 1.10, 0.92 and 0.12.

Table 2. Components for the whole set model and residual standard deviations s_k for each substituent when (*i*) fitted to the whole class model, (*ii*) to its own class model (classes; alkyls, halogens, acceptors and donors) and (*iii*) to its next closest class. The residual standard deviations are denoted $s_k(i)$, $s_k(ii)$ and $s_k(iii)$.

k Subst.	θ_1	θ_2	$s_k(i)$	$s_k(ii)$	$s_k(iii)$	$F^b_{calc.}$
1 H	0.610	−0.698	0.423		0.32(A)	17
2 Me	0.890	−0.122	0.248	0.063	0.37(D)	4.0
3 Et	0.879	0.140	0.199	0.017	0.36(D)	3.8
4 *i*-Pr	0.879	0.515	0.164	0.061	0.37(D)	4.0
5 *t*-Bu	0.832	1.089	0.190	0.109	0.48(D)	6.8
6 CH_2Ph	0.707	1.179	0.334	0.020	0.52(D)	8.0
7 Ph	0.503	1.621	0.235	0.073	0.61(D)	11.
8 F	0.225	−0.840	0.218	0.027	0.46(D)	6.3
9 Cl	−0.119	−0.364	0.218	0.045	0.39(D)	4.5
10 Br	−0.175	−0.186	0.227	0.027	0.34(D)	3.4
11 I	−0.147	0.123	0.226	0.041	0.25(D)	1.8
12 CF_3	−1.249	0.154	0.276	0.30	0.48(H)	689
13 CO_2Me	−1.235	−0.006	0.222	0.13	0.60(H)	138
14 COPh	−1.129	1.415	0.173	0.10	0.40(H)	62
15 CHO	−1.479	−0.137	0.228	0.18	0.65(H)	162
16 CO_2R	−1.066	0.194	0.144	0.03	0.48(H)	89
17 CO_2H	−1.210	−0.108	0.287	0.18	0.62(H)	148
18 SO_2NH_2	−1.767	−0.491	0.324	0.27	0.78(D)	18
19 SO_2Me	−2.083	−0.430	0.315	0.22	0.88(D)	23
20 CN	−1.881	−0.451	0.103	0.11	0.72(H)	199
21 NO_2	−2.388	−0.055	0.193	0.18	0.82(H)	258
22 OMe	1.360	−0.557	0.083	0.16	0.39(A)	18
23 OH	1.614	−0.806	0.135	0.13	0.41(A)	120
24 OPh	1.074	1.181	0.322	0.035	0.35(A)	15
25 SMe	0.855	−0.032	0.134	0.14	0.33(A)	13
26 NH_2	2.111	−0.810	0.264	0.071	0.67(A)	53
27 NMe_2	2.374	−0.230	0.216	0.071	0.84(A)	83
28 NHAc	0.954	−0.147	0.367	0.21	0.51(A)	31

[a] Components for the whole set model. [b] *F*-Test according to eqn. (6) testing if a substituent data vector belongs to its next closest class. $F_{0.05} = 2.9$, 3.0 and 2.6 when a data vector is fitted to the alkyl, halogen or donor class models, respectively. [c] Next closest class given within parentheses A = alkyls, H = halogens and D = donors.

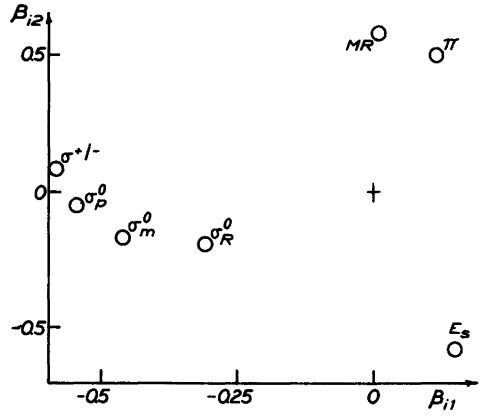

Fig. 1. A plot of β_{i1} against β_{i2} for the whole set analysis.

tion about the shape of the clusters has to be made.

Another investigation similar to the present one is that of Nieuwdorp *et al.*[22] In this paper factor analysis was applied to a data set consisting of 17 substituents and 76 reactions series, representing a wide span of reaction types. For each of the substituents three parameters were estimated from a model similar to eqn. (1) with A = 3. By plotting the three estimated constants for the substituents against each other, we find the same strong grouping of the substituents as in the present work.

If the four substituent classes are described by separate *PC* models, by pooling the variances[23] with $A = 0$ reported in Table 3, we see that 73 % of the total variance is described. The same scaling is retained as in the overall analysis to enable a comparison. Thus the conclusion is that 73 %

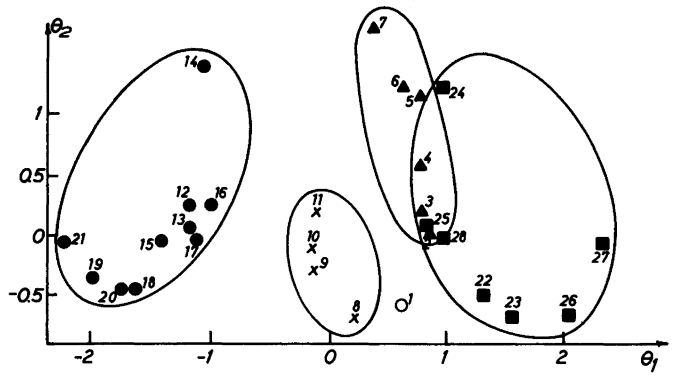

Fig. 2. Plot of the first against the second component for the whole data set model. Numbers identifying substituents as in Table 2; ●, acceptors; ■, donors; ▲, alkyls; ×, halogens.

of the total variance in the data consists of interclass variance while 27 % of the variance is due to intraclass variance. Since the overall analysis with 2 components explains 82 % of the total variance, the overall analysis mainly explains the interclass variation in data and only a minor part of the intraclass behaviour, see Fig. 3.

The high significance of the grouping is also confirmed by an analysis of variance.[23] Thus, the description of the data-set ($-H$ is not included) is significantly better ($P<0.01$) by the present four models with A = 0, than by an overall model with $A=0$.

The intraclass structures are better described by separate class models with $A=1$ in the case of the halogens and by $A=2$ for the other classes. These models describe 92 % of the variance in the whole set data or 70 % of the pooled variance for the separate class models with $A=0$ (see Fig. 3). This shows that for the given set of substituent descriptors separate class models are superior to a single overall model for describing the intraclass structure.

In addition, the separate class models can be further refined by using the same weighting strategy as for the whole class analysis. Thus as much as 93 % on an average of the intraclass structure is described if the data for each class first is autoscaled and then reweighted by multiplying each variable with its modelling power. The variance for the models with this weighting are given within paren-

Table 3. Residual standard deviation after model expansion for the whole set and for each subset with the same scaling. The value within parentheses for the separate class models refer to the optimized models with individual scaling and therefore the values in the different rows are not strictly comparable. The percentages of the variance with $A=0$, explained by models with $A=1$ and $A=2$ are denoted V_1 % and V_2 %.

Set	n^a	$S_0(A=0)^b$	$S_1(A=1)^b$	$S_2(A=2)^b$	V_1 %^{c}	V_2 %^{c}
Whole	28	0.605	0.361	0.257	64	82
Alkyls^d	6	0.316	0.197	0.092	61	92
		(0.471)	(0.252)	(0.091)	(71)	(96)
Halogens^d	4	0.174	0.051		92	
		(0.334)	(0.059)		(97)	
Acceptors^d	10	0.322	0.255	0.214	38	56
		(0.279)	(0.194)	(0.073)	(52)	(93)
Donors^d	7	0.367	0.253	0.184	52	75
		(0.473)	(0.177)		86	

^{a}Number of substituents. ^{b}Residual standard deviation for the *PC* models with $A=0(S_0)$, $A=1(S_1)$ and $A=2(S_2)$. ^{c}Percentages of the variance with $A=0$, explained by models with $A=1$ [$V_1\%=100(1-S_1/S_0)$] and by the models with $A=2$[$V_2\%=100(1-S_2/S_0)$]. ^{d}The substituents are for alkyls; 2–7, halogens; 8–11, acceptors; 13–21 and donors 22–28. For numbering see Table 2.

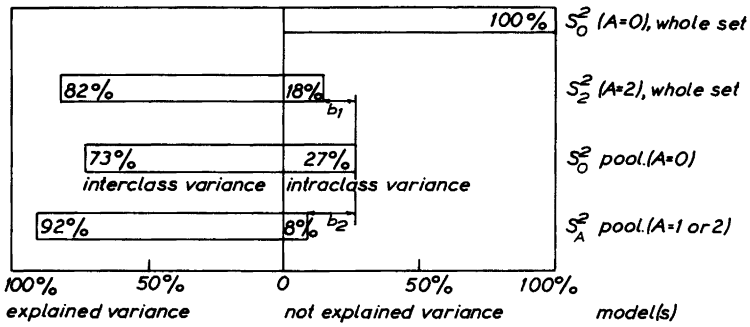

Fig. 3. Illustration of the sizes of the residual variances for different models expressed as % of the residual variance of the overall model with $A=0$. When the four separate classes are described by their mean values, 73% of the total variance is described. Thus the 73% can be assigned as interclass variance. Of the 27% intraclass variance, the whole set model only explains 9% (b_1) and the four separate models explain 19% (b_2). Thus the overall model with $A=2$ explains ~33% and the four class models ~70% of the intraclass variance.

theses in Table 3.

In the separate class models the contribution from the different variables varies considerably. For donors the "electronic" variables dominate (σ_m^0,σ_p^0 and $\sigma^{+/-}$) while for the halogens the "non-electronic" variables dominate (E_S, π and MR). For the alkyls σ_p^0, $\sigma^{+/-}$, E_S, π and MR contribute, while for the acceptors mainly σ_p^0, π and MR are important. This can be seen from the β-values and modelling powers for the different intraclass structures are not parallel to each other and consequently they are not parallel with the interclass behaviour.

The four subsets are well separated, as indicated by the residual standard deviations given in Table 2, obtained by the SIMCA[7] classification method. The standard deviations are given for each substituent (*i*) when fitted to the overall model, (*ii*) to its own class model and (*iii*) to its next closest class. The residual standard deviations in case (*i*) and (*iii*) are significantly larger ($P=0.05$) in all cases except one, compared to the standard deviation in case (*ii*). The exception is iodine that also fits the donor class, seen from the *F*-test in Table 2.

The class separation is clearly seen in Fig. 2. However, the figure is somewhat misleading with respect to substituents 25 and 28 (SMe and NHAc). These substituents are not close to the alkyls model. Their positions in the plot are due to an artifact of the projection of the data down on a plane. Hydrogen was not initially assigned to any of the classes. Indeed it does not belong to any class according to the *F*-tests (see Table 2). It is also noticable that Ph is well described by the same model as the alkyls, even if Ph formally not is an alkyl substituent.

To summarize the results: A strong grouping in the substituent descriptors is found. An overall *PC* model mainly explains interclass variation and little intraclass variation. Separate models for separate groups of substituents describe the intraclass behaviour much better than a single overall model. The separate models do not parallel with their interclass behaviour. In the overall model a high collinearity is found between some of the descriptors.

These findings will be important in structure-reactivity and structure-activity studies. With respect to the first area, we note that each subset has its own particular intraclass structure. A general theory that is able to cope with these class structures is not yet available. In structure-activity studies one rarely finds suitable model systems that are approximately linearly related to the properties of the system under investigation. Hence several substituent scales are used in a multiple regression model and the problems discussed above become serious. We also note that the present set of descriptors is widely used in structure-activity studies (see Ref. 16). If the present descriptors are used as independent variables in MRA, the result will be an unstable model which will have poor predictive ability due to the strong collinearity between some of the descriptors.[19] We also note that the statistical tests on the regression coefficients in MRA assume that the objects are not grouped. If they are grouped, the confidence intervals of the regression coefficients will be deceptively small and the correlation coefficients deceptively high.

The collinearity problem could be circumvented

by using the present θ_1 and θ_2 substituent parameters as independent variables in MRA instead. However, also in this case, the limitations for MRA are present due to the grouping of the substituents. This means that the prediction of a compound with unknown behaviour will not be much better than by using the mean value of the dependent variable for the already measured compounds in this class. However, in order to reasonably well define separate class models, at least 4–5 substituents must be present in each class. If one or more of the separate classes not are represented in the dependent variable and we want to make a prediction for a compound in the missing class, this can only be obtained if the relative position of the classes is the same in the dependent variable and independent variable. Whether this assumption is valid or not remains to be investigated.

Supplementary material available. Data used, α_i, β_{ia} and θ_{ak} for the four subclasses and classification results, can be obtained on request from the authors.

Acknowledgements. The Italian National Research Council (*C.N.R.*) is thanked for a special grant to SC. Support from the Swedish National Science Research Council (*NFR*) and the Swedish Council for Planning and Coordination of Research (*FRN*) is gratefully acknowledged.

REFERENCES

1. Wold, S. and Sjöström, M. In Chapman, N. B. and Shorter, J., Eds., *Correlation Analysis in Chemistry, Recent Advances.* Plenum, London 1978, Chapter 1.
2. Wold, S. and Sjöström, M. *Chem. Scr. 2* (1972) 49; b. Sjöström, M. and Wold, S. *Chem. Scr. 6* (1974) 114; c. Sjöström, M. and Wold, S. *Acta Chem. Scand. B 30* (1976) 167; d. Sjöström, M. and Wold, S. *Chem. Scr. 9* (1976) 200; e. Sjöström, M. and Wold, S. *J. Chem. Soc. Perkin Trans. 2* (1979) 1274.
3. Sjöström, M. and Wold, S. *Acta Chem. Scand. B 35* (1981) 537.
4. Clementi, S., Fringuelli, F., Linda, P. and Savelli, G. *Gazz. Chim. Ital. 105* (1975) 281.
5. Clementi, S. Fringuelli, F. and Savelli, G. *Chim. Ind. (Milan) 60* (1978) 598.
6. Clementi, S. and Fringuelli, F. *Anal. Chim. Acta 103* (1978) 477.
7. Sjöström, M. and Wold, S. In Kowalski, B. R., Ed., *Chemometrics: Theory and Application, A.C.S. Symposium Series No. 52,* Washington 1977, Chapter 7.
8. Albano, C., Dunn, W. J., III, Edlund, U., Johansson, E., Nordén, B., Sjöström, M. and Wold, S. *Anal. Chim. Acta 103* (1978) 429.
9. Albano, C., Blomquist, G., Coomans, D., Dunn, W. J., III, Edlund, U., Eliasson, B., Hellberg, S., Johansson, E., Johnels, D., Nordén, B., Sjöström, M., Wold, H. and Wold, S. In Höskuldsen, A., *et al.* Eds., *Symposium i Andvent Statistik, NEUCU, RECAU and RECKU,* Copenhagen 1981, p. 183.
10. Wold, S. *Chem. Scr. 5* (1974) 97.
11. This second implicit assumption does not seem to be recognized yet, as indicated by the wide acritic use of regression; see for instance Reynolds, W. F., Dais, F., MacIntyre, D. W. and Peat, I. R. *J. Magn. Reson. 43* (1981) 81.
12. Topliss, J. G. and Edwards, R. P. *J. Med. Chem. 10* (1979) 1238.
13. For a comment on the use of a "minimal basis set" see Clementi, S. *CAOC Newsletter 2* (1981) 13.
14. Katritzky, A. R. and Topsom, R. D. *Angew. Chem. 82* (1970) 106.
15. Exner, O. In Chapman, N. B. and Shorter, J., Eds., *Correlation Analysis in Chemistry, Recent Advances,* Plenum, London 1978, Chapter 10.
16. Seydel, J. K. and Schaper, K. J. *Chemische Struktur und Biological Aktivität von Wirkstoffen,* Verlag Chem., Weinheim-New York 1979, 226.
17. Wold, S. *Pattern Recognition 8* (1976) 127.
18. Wold, S. *Technometrics 20* (1978) 397.
19. Draper, N. and Smith, H. *Applied Regression Analysis,* 2nd Ed., Wiley-Interscience, New York 1981.
20. Johnels, D., Clementi, S., Dunn, W. J., III, Edlund, U., Grahn, H., Hellberg, S., Sjöström, M. and Wold, S. *J. Chem. Soc. Perkin Trans. 2. In press.*
21. Hansch, C., Unger, S. H. and Forsythe, A. B. *J. Med. Chem. 16* (1973) 1217.
22. Nieuwdorp, G. H. E., deLigny, C. L. and van Houwelingen, H. C. *J. Chem. Soc. Perkin Trans. 2* (1979) 537.
23. Box, E. P., Hunter, W. G. and Hunter, J. S. *Statistics for Experimenters.* Wiley-Interscience, New York 1978.

Received April 5, 1982.