The ^{13}C Nuclear Magnetic Resonance Substituent Chemical Shifts of 2-Substituted Indenes. Interpretation by a Multivariate Data Analysis Method

By **Bertil Eliasson** and **Ulf Edlund**,* Department of Organic Chemistry, Institute of Chemistry, Umeå University, S-901 87 Umeå, Sweden

The ^{13}C n.m.r. chemical shifts of several 2-substituted indenes are reported and the induced substituent shieldings of the remote proton-bearing aromatic carbons are analysed by a multivariate data analysis method. A dual parameter model is necessary to describe the shift variations and the two components obtained were found to correlate to a mesomeric $\sigma_R°$ and field–inductive σ_I scale, respectively. The two dual parameter models are tested on a structurally similar heterocyclic system. The limitations of multiparameter equations in predicting ^{13}C chemical shifts and in separating electronic effects are briefly discussed.

In the past ten years there has been a considerable interest in the correlation between reactivity parameters, originally intended to apply to equilibria and reaction rates, and spectroscopic properties such as ^{13}C n.m.r. substituent chemical shifts (s.c.s.).[1] A successful correlation between these parameters and s.c.s. would indicate that the effect of any other substituent on the equilibria or the reaction rate could be satisfactorily predicted by simply measuring a chemical shift of the actual derivative. The given approach will be appropriate provided; (*a*) that the experimental values (σ) are some measure of the electronic ground state perturbations and (*b*) that the ^{13}C s.c.s. will reflect these electronic perturbations at the nucleus.

If we restrict the discussion to n.m.r. investigations, it was recognized at an early stage that in order to get a reasonable correlation, an extension of the Hammett equation was often needed.[1,2a] However, these so-called dual substituent parameter equations (d.s.p.) put tight limits on data sets, with respect to the number and the choice of substituents.[1c] The most popular treatments applied in the n.m.r. field are the models suggested by Taft ($\Delta p = \rho_I\sigma_I + \rho_R\sigma_R$, where σ_R could be chosen from the sets $\sigma_R°$, $\sigma_R{}^+{}_{(\mathrm{BA})}$, $\sigma_R{}^+$, or $\sigma_R{}^-$) [1b] and Swain and Lupton ($\Delta p = fF + rR$).[3] The aim of these treatments was not to get a better fit than any other suitable linear combination of σ values, but to achieve a separation of the polar–field (σ_I,*F*) and mesomeric (σ_R, *R*) effects. Both these extensions of the simple Hammett equation, together with the d.s.p. model of Yukawa–Tsuno,[4] have been compared and criticized in several papers.[2a,5] The *F–R* model is considered to be the most questionable, not because it does not work as a correlation model, but because it claims to afford a separation of substituent effects that is illusory. In all these equations the regression parameters are taken as susceptibility or transmission coefficients for the polar–field and mesomeric substituent effects.

In spite of the severe criticism of the *F–R* model these substituent constants have recently reappeared in a three-component model introduced by Smith.[6] The extra parameter, *Q*,[7] was originally included to get acceptable plots of carbons of expected complexicity such as *ipso*- and *ortho*-carbons ($Q = P/Ir$,[3] where *P* is the polarizability, *r* the length of the C–X bond, and *I* the first ionization potential of the X atom). However, since many n.m.r. data sets better accommodate this equation with the extra parameter, the equation was quickly accepted as a general model for the correlation of carbon shifts (see below), unfortunately without proper consideration of the statistical significance. The degree of parameterization has been claimed to be a matter of subjective judgment,[2a] but from a statistical point of view this is certainly not the case.[8] Too much concern has been centred on the correlation coefficient *r* as a criteria for goodness of fit. Unfortunately correlation coefficients get larger with decreasing sample size,[2b] and a small data set is a very common situation in n.m.r. s.c.s. studies. The degree of parameterization should be checked by an *F*-test (variance analysis) on the squared residual standard deviations obtained when fitting the data to the two models.[2b,8]

Another problem arises as to whether or not one kind of effect is significantly correlated to the other, *i.e.* if there are two (or more) independent mechanisms for the transmission. The answer will have important consequences on the relevance of separating mesomeric and polar–field effects.

Finally, for a universal d.s.p. model to prevail, we must assume that there are no discontinuous changes in the transmission framework (core region) during the perturbation of the system. This is quite possible, especially if strongly interacting substituents (*e.g.* NO_2, NMe_2) are attached.

In two earlier reports we have introduced a pattern recognition method, based on multivariate data analysis (principal component analysis) as a tool for handling spectroscopic parameters, in our case n.m.r. parameters.[8,9] This approach proved valid for signal assigning purposes, classification of unknown structures, and for interpretation of n.m.r. s.c.s. For a specific class of 4-substituted styrenes we found a one-component model (strongly correlated to σ_p) to be statistically sufficient to describe the s.c.s. as probed by the C-1, C-β, and the vinyl hydrogen chemical shifts.[8] Similar results showing a behaviour ' simpler than predicted ' are quite common. Except in n.m.r. studies of monoaromatic systems such correlations to a single substituent constant have been noticed for the substituent effects on the cleavage of compounds YC_6H_4-$(C{\equiv}C)_nMEt_3$ (n = 1—3; M = Si or Ge),[10] and on various physical and chemical properties of substituted polyvinylenes $Y(CH = CH)_nX$.[11] The ratio $\rho_n : \rho_{n-1}$ was in both these cases unaffected by n, a result which would imply a similar distance and angle dependence of field and resonance effects. According to the theory of the transmission of substituent effects this result cannot be true. In order to shed further light on these problems and to test the generality of the d.s.p. treatments, we have continued a ^{13}C n.m.r. study of 2-substituted indenes.[12] This framework seemed well suited for this purpose since it is rigid, *i.e.* the orientation of the substituent to the probe centres is fixed. Moreover, the orientation of the C-4–C-7 carbons should facilitate the identification of a blend of effects, if these exist.

EXPERIMENTAL

N.M.R.—The ^{13}C n.mr. chemical shifts of the alkyl-substituted indenes and (5) were measured earlier.[12] All other spectra were obtained on a JEOL PFT-60 n.m.r. spectrometer, except for the spectrum of (15), which was recorded on a Varian XL-100 instrument. $CDCl_3$ was used as solvent for all compounds, except for (11) and (12), where $[^2H_6]$DMSO was used. The concentrations were 1.0 ± 0.3M. The probe temperature was 27 °C. The chemical shifts are given in p.p.m. downfield from Me_4Si with an accuracy of ±0.1 p.p.m.

Compounds.—The indene derivatives (7),[13] (8),[14] (9),[15] (11),[16] (12),[17] (14),[18b] and (15) [19] were all synthesized according to earlier methods and identified by ^1H n.m.r. and mass spectra. Compound (6) was prepared in the same way as (7) using MeOH and HCO_2CH_3, m.p. 38—40 °C. The ketone (10) was prepared starting from the acid chloride (14) and CH_3MgI, using inverse addition. The m.p. and spectroscopic data were similar to those earlier reported.[18] The methyl ester (13) was obtained by reducing 2-methoxycarbonylindan-1-one [20] with $NaBH_4$, followed by dehydration of the alcohol with $KHSO_4$. Crystallization from MeOH gave crystals, m.p. 78—80.5 °C, *m/e* 174, 143, 129, and 115, δ_H ($CDCl_3$) 7.6 (1 H, t), 7.5—7.1 (4 H, m), 3.8 (3 H, s), and 3.6 (2 H, d).

1-Methyl-2-bromoindene was synthesized from 2-bromoindenyl-lithium and CH_3I. This compound was rearranged to 3-methyl-2-bromoindene in basic solution.[21a] The 1-methyl and 3-methyl derivatives of (13) were synthesized starting from 3-methylindan-1-one.[21b] The 3-methyl derivative of (15) was also prepared starting from 1-methylindene. These compounds were used for assignment purposes.

RESULTS AND DISCUSSION

Assignments.—The ^{13}C n.m.r. chemical shifts of 14 2-substituted indenes are given in Table 1. The carbon shifts of the vinyl ethers and the halogenated compounds were easily assigned by comparisons to the absolute shifts of benzothiazoles [22a] and benzoxazoles,[22b] and by variable-by-variable plots. The unique substituent effects observed by introducing a methyl group in the 1- or 3-position afforded the assignment of the five-membered ring positions, and also the assignment of C-4 and C-7.[12] Characteristic shielding effects were noticed for these aromatic carbons. Additional confirmation of the proposed assignments of C-5 and -6 of compounds (8)—(15), was obtained by having compounds (1)—(7) in a training set, and the other structures in a test set and performing a SIMCA classification data analysis.[8,23] The assignments of these carbons were then reversed and the data analysis procedure was repeated. A comparison of the two classifications strongly supported the given assignment.

Correlation Analysis.—As earlier shown for 4-substituted styrenes [8] a statistical analysis of n.m.r. s.c.s. can be achieved by principal component (p.c.) analysis, combined with a cross-validation procedure. Our observed s.c.s., y_{ik}, will form a matrix Y described by equation (1)

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

where α_i is the variable mean, β_{ia} is the regression coefficient, and θ_{ak} constatutes the substituent constant. Deviations from the mathematical model are given as the residuals ε_{ik}. The degree of parameterization, *i.e.* the rank of the matrix, is given as A and will be determined by a cross-validation technique.[24] A data program system, SIMCA, is used for the total analysis, and a complete description of this package has been given earlier.[23] The most important steps will briefly be mentioned here. In the present case, the data analysis will be restricted to a single class. To measure to what extent the variation in one variable i is explained by a given p.c. model, we have used ψ_i, the so-called modelling power. This measure is defined as $\psi_i = 1 - (s_{i\varepsilon}/s_{iy})$ where equations (2a and b)

$$s_{i\varepsilon} = [M/(M - A) \sum_{k=1}^{N} \varepsilon_{ik}^2/(N - A - 1)]^{1/2} \qquad (2a)$$

apply. M is the number of variables and N is the number of objects.

$$s_{iy} = \left[\sum_{k=1}^{N} (y_{ik} - \bar{y}_i)^2/(N - 1)\right]^{1/2} \qquad (2b)$$

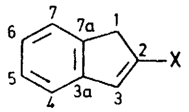
A good fit of a model having A components will result in a modelling power ψ_i close to unity.

Scaling is another important concept in cases where the modelling powers differ between variables having significant differences in variances. A small but significant variation in one variable can be superseded by a large variation in another variable. Autoscaling is then a necessity, giving the participating variables the same variance provided that the experimental errors are small compared with the model error. For the 4-substituted styrenes scaling was unnecessary since the relevant variables had nearly the same modelling powers.[8]

Finally it can be of interest to compare the residual standard deviation s_k for a given substituent k and the total residual standard deviation s_0 of the substituents in the given p.c. model. A high standard deviation s_k compared to s_0 will indicate different behaviour, *i.e.* an ' outlier '.

Table 1

^{13}C N.m.r. chemical shifts of 2-substituted indenes [a]



(1) X = H
(2) X = Me
(3) X = Pr^i
(4) X = Bu^t
(5) X = NMe$_2$
(6) X = OMe
(7) X = OEt
(8) X = Cl
(9) X = Br
(10) X = COMe
(11) X = CO$_2$H
(12) X = CONH$_2$
(13) X = CO$_2$Me
(14) X = COCl
(15) X = NO$_2$

Compound	C-1	C-2	C-3	C-4	C-5	C-6	C-7	C-3a	C-7a	Others	Comments
(1)	39.0	133.8	132.1	120.9	126.1	124.5	123.6	144.7	143.5		*b*
(2)	42.6	145.9	127.2	119.7	126.1	123.2	123.4	145.7	143.3	16.6 (CH$_3$)	*b*
(3)	39.1	156.7	124.1	120.0	126.1	123.6	123.5	145.1	142.9	30.0 (CH), 22.5 (CH$_3$)	*b*
(4)	37.7	159.6	123.3	120.1	126.1	123.6	123.5	145.4	143.0	33.3 (C), 30.4 (CH$_3$)	*b*
(5)	40.3	158.0	97.1	116.9	126.6	119.7	122.7	147.9	136.9	37.5 (CH$_3$)	*b*
(6)	37.5	167.6	99.0	118.8	126.5	122.3	123.2	145.0	136.0	57.1 (CH$_3$)	
(7)	37.7	166.6	99.2	118.7	126.5	122.2	123.1	145.2	135.9	65.5 (CH$_2$), 14.4 (CH$_3$)	
(8)	43.2	137.1	128.4	120.2	126.6	124.6	123.1	143.2	140.9		
(9)	45.2	124.7	132.8	120.0	126.5	124.7	123.0	143.8	142.4		
(10)	37.3	146.1	141.2	123.7	126.8	128.1	124.4	142.7	144.9	195.4 (CO), 26.4 (CH$_3$)	
(11)	38.3	140.4	138.6	123.5	126.9	127.4	124.4	142.8	144.8	165.9 (CO)	*c*
(12)	38.4	142.3	135.4	122.8	126.8	126.6	124.2	143.2	144.1	166.1 (CO)	*c*
(13)	38.3	141.1	137.0	123.3	126.8	127.5	124.2	142.6	144.7	165.3 (CO), 51.5 (CH$_3$)	
(14)	38.4	149.2	140.0	124.7	127.4	129.5	124.5	141.3	145.6	162.8 (CO)	
(15)	36.5	153.5	136.2	125.2	127.9	129.9	124.6	139.4	141.8		

[a] In p.p.m. downfield from Me$_4$Si. Solvent CDCl$_3$. [b] Shifts from ref. 12. [c] Solvent [^{2}H$_6$]DMSO. The shifts were adjusted to the Me$_4$Si scale with $\delta([^2H_6]DMSO) - \delta(Me_4Si) = 39.6$ p.p.m.[26]

For the analysis of ^{13}C s.c.s. of 2-substituted indenes, the C-4–C-7 shielding differences (Table 2) were chosen as variables since it would be desirable to have the substituent scale based on similar types of carbons, and those carbons should also be as unperturbed as possible by steric effects from the 2-substituent.

Thus our measured ^{13}C s.c.s. formed a 4 × 15 observation matrix Y_{ik} where i represents a specific carbon and the index k stands for the substituent. The matrix was analysed with the SIMCA program, showing that 99% of the squared residuals of variables C-4–C-7 were explained by a one-component model, using unscaled s.c.s. However, differences in variances and modelling powers between these variables made scaling necessary. Repeated analysis on autoscaled[23] data showed clearly that a two-component model was significant for the description of the ^{13}C s.c.s. variation. The modelling powers [equation (2)] obtained in the final analysis for the variables C-4–C-7 were 0.916, 0.956, 0.750, and 0.713, respectively. In Table 3 the component values are given along with the earlier mentioned F test. By definition the two components will be orthogonal and thus probe separate ' effects '.

Table 2

^{13}C N.m.r. s.c.s. of 2-substituted indenes [a]

Compound	C-4	C-5	C-6	C-7	Comments
(1)	0	0	0	0	
(2)	−1.2	0	−1.3	−0.2	
(3)	−0.9	0	−0.9	−0.1	
(4)	−0.8	0	−0.9	−0.1	
(5)	−4.0	0.5	−4.8	−0.9	
(6)	−2.1	0.4	−2.2	−0.4	
(7)	−2.2	0.4	−2.3	−0.5	
(8)	−0.7	0.5	0.1	−0.5	
(9)	−0.9	0.4	0.2	−0.6	
(10)	2.8	0.7	3.6	0.8	
(11)	2.7	0.8	3.0	0.7	*b*
(12)	2.0	0.7	2.2	0.5	*b*
(13)	2.4	0.7	3.0	0.6	
(14)	3.8	1.3	5.0	0.9	
(15)	4.3	1.8	5.4	1.0	

[a] Shifts relative to indene in CDCl$_3$. [b] Shifts relative to indene in [^{2}H$_6$]DMSO.

The first component, θ_1, shows acceptable correlations to mesomeric scales, but in order to be comparable to the most well accepted d.s.p. treatment (σ_R°, σ_I)[1b,25] the θ_1 values were transformed by a least-square fit to σ_R° according to $\sigma_R^\circ = a_1\theta_1' + a_2\theta_2' + e$ where θ_1' and θ_2' are the normalized θ_1–θ_2 scales. Then equations (3) apply. The new component values θ_1^* and θ_2^* (Table 3) were

$$\begin{aligned} \theta_1^* &= a_1\theta_1' + a_2\theta_2' \qquad (3) \\ \theta_2^* &= a_2\theta_1' - a_1\theta_2' \end{aligned}$$

TABLE 3

θ Values from principal component analysis of C-4—C-7 ^{13}C n.m.r. s.c.s. of 2-substituted indenes. θ* Values obtained from adaption of the θ_1 scale to the $\sigma_R{}^\circ$ scale

Compound	θ_1	θ_2	$\theta_1{}^*$	$\theta_2{}^*$	s_k [a]	F [b]	Comments
(1)	0	0	0	0	0.053	0.072	
(2)	−0.65	0.30	−0.13	0.06	0.067	0.12	
(3)	−0.43	0.20	−0.09	0.04	0.079	0.16	
(4)	−0.41	0.19	−0.08	0.04	0.069	0.12	
(5)	−2.0	2.1	−0.58	0.50	0.26	1.7	
(6)	−0.82	1.3	−0.30	0.32	0.16	0.67	
(7)	−0.94	1.3	−0.33	0.34	0.085	0.19	
(8)	−0.11	1.2	−0.20	0.34	0.40	4.2	Outlier
(9)	−0.30	1.1	−0.21	0.30	0.52	7.0	Outlier
(10)	2.5	0.31	0.29	0.18	0.081	0.17	
(11)	2.4	0.59	0.23	0.25	0.11	0.30	
(12)	1.8	0.64	0.15	0.24	0.086	0.19	
(13)	2.1	0.50	0.22	0.22	0.005	0.001	
(14)	3.6	1.1	0.31	0.44	0.046	0.056	
(15)	4.3	1.9	0.30	0.68	0.054	0.075	

[a] Substituent residual standard deviation in the p.c. model. This value is compared to the total residual standard deviation of the model, s_0 0.196. [b] F Test value, used in assigning the substituents to the class, described by the p.c. model The F values are compared to the critical value F_{crit} of 3.4.

TABLE 4

Results of linear regressions, $Y = A + BX + e$

X	Y	A	B	$p(A)$ [a]	$p(B)$ [a]	r [b]	s [c]	F [d]	n [e]	Comments
$\theta_1{}^*$	$\sigma_R{}^\circ$	−0.04	0.96	0.04	0.15	0.967	0.071	188	15	
$\theta_2{}^*$	σ_I	0.00	0.80	0.12	0.36	0.797	0.118	23	15	
		−0.02	0.97	0.08	0.26	0.919	0.079	65	14	$N(CH_3)_2$ excluded

[a] 95% Confidence intervals of the regression coefficients. [b] Correlation coefficient. [c] Residual standard deviation of regression. [d] F values, used in test for significance of regression. [e] Number of points.

then given the same variances as the $\sigma_R{}^\circ$ and σ_I scales. The second component $\theta_2{}^*$ is reasonably correlated to σ_I, especially if NMe_2 is excluded in the regression analysis (r 0.919, n 14) (Table 4).

A comparison between the $\theta_1{}^*$–$\theta_2{}^*$ and $\sigma_R{}^\circ$–σ_I models using the total ^{13}C s.c.s. data of C-4–C-7 shows that the $\theta_1{}^*$–$\theta_2{}^*$ scales are considerably better than $\sigma_R{}^\circ$–σ_I, corrected for the differences in degrees of freedom (Table 5). The two models are tested on the individual positions using a normal multiple regression procedure. The superiority of the $\theta_1{}^*$–$\theta_2{}^*$ scales is most pronounced for the C-5 position. For the C-7 position only the first or mesomeric component is significant using any of the mentioned d.s.p. scales. The ratios between the regression coefficients ($\beta_1{}^*$: $\beta_2{}^*$) are the largest and of similar sizes for C-4 and -6, while the $\theta_2{}^*$ component dominates the C-5 data variation. This result is consistent with an alternating mesomeric contribution and the dominance of the $\theta_2{}^*$ component at C-5 is also very likely on a polar–field basis. Very similar ratios were obtained using Taft's $\sigma_R{}^\circ$–σ_I model as shown in Table 5.

To check the generality of the $\theta_1{}^*$–$\theta_2{}^*$ scales we applied this model to various monosubstituted aromatic systems, but no significant difference between the $\theta_1{}^*$–$\theta_2{}^*$ and the $\sigma_R{}^\circ$–σ_I models was found in any of the studied cases. The overall conclusion from this comparative study is, that the use of $\theta_1{}^*$–$\theta_2{}^*$, or any commonly used d.s.p. equation, to predict ^{13}C chemical shifts is strongly limited to molecular frameworks similar to the one for which the equation was originally derived. For the limited number of substituents available in the benzothiazole series,[22a] the ^{13}C s.c.s. variation of C-4, -6, and -7 was completely dominated by the ' mesomeric ' component using any of the two models. Again the $\theta_1{}^*$–$\theta_2{}^*$ model gave a significantly better fit to the C-5 values and the second (*i.e.* θ^*_2) parameter was dominant (Table 6). The reported

TABLE 5

Linear multiple regression on ^{13}C s.c.s. of 2-substituted indenes

$$Y = A + B_1X_1 + B_2X_2 + e$$

X_1	X_2	Y	A	B_1	B_2	$p(A)$ [a]	$p(B_1)$ [a]	$p(B_2)$ [a]	r	$F(B_1)$ [b]	$F(B_2)$ [c]	Σe^2	$\Sigma\Sigma e^2$	n [d]
$\theta_1{}^*$	$\theta_2{}^*$	C-4	−0.1	8.88	2.68	0.15	0.33	0.48	0.998	2 255	99	0.29		15
		C-5	0.0	1.05	2.19	0.02	0.04	0.05	0.999	2 527	5 227	0.004		15
		C-6	0.0	10.3	3.71	0.52	1.1	1.6	0.986	268	17	3.27		15
		C-7	0.0	2.21	0.423	0.12	0.27	0.38	0.983	219	3.8	0.19		15
													3.75	
$\sigma_R{}^\circ$	σ_I	C-4	0.4	8.22	2.55	0.58	1.4	2.0	0.971	157	7.3	4.92		15
		C-5	0.2	0.68	1.84	0.25	0.61	0.87	0.864	5.9	21	0.89		15
		C-6	0.4	9.38	4.26	0.60	1.5	2.1	0.978	194	20	5.17		15
		C-7	0.2	2.08	0.112	0.23	0.57	0.82	0.923	63	0.1	0.79		15
													11.77	

[a] 95% Confidence intervals of the regression coefficients. [b] F Value, used in testing the significance of regression when expanding the one-component model $Y = A + B_2X_2 + e$ to the two-component model. [c] F Value for expansion of $Y = A + B_1X_1$ to the two-component model. [d] Number of points.

investigation of ^{13}C s.c.s. in this series claimed that an extension of the *F–R* model was necessary to give a satisfactory description of the induced shielding variations. However, the use of an extended model is merely a consequence of the limited generality of the commonly used d.s.p. equations and illustrates the danger of using fixed, pre-determined scales for the interpretation of n.m.r. chemical shifts. Moreover, a correlation coefficient *r* 0.83 is obtained between the *R* and *Q* scales. This fact will strongly influence the accuracy of the transmission coefficients, and by excluding one structure from the analysis large deviations in the regression coefficients will occur.

Independently of the choice of the fixed d.s.p. model used for the correlation, the conditions for obtaining statistically significant results for the evaluation of the blend of field–inductive and mesomeric effects are quite severe. A sufficient large range and number of substituents (8—10) must be chosen to avoid multicollinearity.[5a,b] Furthermore the transmission coefficient (λ) must be significantly different from unity to mention, on a statistical basis, anything about the variation of the blend of effects.[5a,b]

TABLE 6

Linear multiple regression on ^{13}C s.c.s. of 2-substituted benzothiazoles [22a]

X_1	X_2	Y	A	B_1	B_2	$p(A)$ [a]	$p(B_1)$ [a]	$p(B_2)$ [a]	r	$F(B_1)$ [b]	$F(B_2)$ [b]	Σe^2	$\Sigma\Sigma e^2$	n [c]
θ_1*	θ_2*	C-4	−0.2	8.03	0.774	0.85	1.8	2.2	0.982	133	0.8	1.26		8
		C-5	−0.1	3.15	3.83	0.27	0.56	0.70	0.994	208	199	0.12		8
		C-6	0.4	7.67	1.24	1.3	2.8	3.5	0.954	50	0.8	3.09		8
		C-7	0.1	3.50	0.580	0.43	0.91	1.1	0.976	97	1.7	0.33		8
													4.80	
σ_R°	σ_I	C-4	0.0	8.58	1.77	0.85	2.1	2.3	0.982	106	3.9	1.24		8
		C-5	0.3	2.52	3.31	1.1	2.9	3.1	0.889	5.1	7.7	2.19		8
		C-6	0.5	7.87	2.85	1.0	2.5	2.7	0.975	66	7.5	1.68		8
		C-7	0.4	3.55	0.271	0.97	2.4	2.6	0.875	14	0.1	1.61		8
													6.72	

Substituents H, Me, NMe_2, OMe, Cl, Br, $CONH_2$, and NO_2. [a] 95% Confidence intervals of the regression coefficients. [b] *F* Value, used in testing the significance of regression when expanding the one-component model to a two-component model. [c] Number of points.

Conclusions.—If the point of an n.m.r. s.c.s. study is to derive substituent scales for a given system or framework, which later will be correlated to other external scales, such as reactivities, activities, *etc*., the multivariate data analysis method presented here will fulfill most demands. This is especially true if the aim of the study is more centred on predictive usefulness, as a criteria for a given signal assignment, for classification purposes, *etc*.

For the class of 2-substituted indenes, having 14 substituents ranging from dimethylamino to nitro, it was found that a two-component model was statistically significant in interpreting ^{13}C n.m.r. s.c.s. of the remote aromatic positions C-4--C-7.

Considering individual positions, the first component was the only significant effect explaining the chemical shift variation at C-7 and this component also dominated the C-4 and -6 ^{13}C s.c.s. These substituent constants were strongly correlated to the σ_R° values. The second component in the present model was correlated to σ_I and this component dominated the C-5 shielding variation. The extension of the study to other aromatic frameworks showed clearly that the usefulness of the d.s.p. equations to predict ^{13}C n.mr. s.c.s. is of a limited generality.

However, most reports in this field attempt to obtain a deeper physical understanding of the underlying principles of the transmission of substituent effects. In aromatic or conjugated systems the need for a general model is thus claimed. From our study the σ_R°–σ_I model seems to have an acceptable relevance as a general model in many cases, but in order to satisfy the statistical demands on the use of these fixed single or dual substituent models we would suggest the following procedure.

(1) Choose substituents representative for the whole substituent domain. The number should preferably exceed 6—7. A minimum basis set of substituents has been suggested.[1b] (2) Fit the n.m.r. s.c.s. to a single σ scale by least squares and calculate the residual standard deviation and the confidence interval for the slope ρ. (3) Plot the residuals against the observed s.c.s. Especially, if the examination of the plot indicates systematic patterns of residuals, test for d.s.p. and check significance by *F* tests. (4) Test for significance using any of the well accepted criteria for goodness of fit.[2b] If a dual substituent parameter equation is needed report the multiple parameter correlation coefficient and the calculated confidence intervals of the regression coefficients.

We thank the Research Group for Chemometrics, Umeå University, for clarifying discussions and for giving us access to the pattern recognition program SIMCA. Financial support from the Swedish Natural Science Research Council is gratefully acknowledged.

[0/918 *Received*, 16*th June*, 1980]

REFERENCES

[1] (*a*) D. F. Ewing in 'Correlation Analysis in Chemistry: Recent Advances,' eds. J. Shorter and N. B. Chapman, Plenum Press, New York, 1978, ch. 8, p. 357; (*b*) S. Ehrenson, R. T. C. Brownlee, and R. W. Taft, *Progr. Phys. Org. Chem.*, 1973, **10**, 1; (*c*) R. D. Topsom, *ibid.*, 1976, **12**, 1; (*d*) G. L. Nelson and E. A. Williams, *ibid.*, p. 229; (*e*) O. Exner in 'Advances in Linear Free Energy Relationships, eds. N. B. Chapman and J. Shorter, Plenum Press, London, 1972, ch. 1, p. 1.

[2] (*a*) J. Shorter in 'Correlation Analysis in Chemistry: Recent Advances,' eds. J. Shorter and N. B. Chapman, Plenum Press, New York, 1978, ch. 4, p. 119; (*b*) S. Wold and M. Sjöström, *ibid.*, ch. 1, p. 1.

[3] C. G. Swain and E. C. Lupton, *J. Amer. Chem. Soc.*, 1968, **90**, 4328.

[4] Y. Yukawa and Y. Tsuno, *Bull. Chem. Soc. Japan*, 1959, **32**, 971.

[5] (*a*) S. Clementi, F. Fringuelli, P. Linda, and G. Savelli, *Gazzetta*, 1975, **105**, 281; (*b*) S. Clementi and F. Fringuelli, *Analyt. Chim. Acta*, 1978, **103**, 477; (*c*) D. A. R. Happer, S. M. McKerrow, and A. L. Wilkinson, *Austral. J. Chem.*, 1977, **30**, 1715; (*d*) M. Sjöström and S. Wold, *Chemica Scripta*, 1974, **6**, 114; (*e*) G. R. Wiley and S. I. Miller, *J. Org. Chem.*, 1972, **37**, 767.

[6] W. B. Smith and T. W. Proulx, *Org. Magnetic Resonance*, 1976, **8**, 567.

[7] F. Hruska, H. M. Hutton, and T. Schaefer, *Canad. J. Chem.*, 1965, **43**, 2392.

[8] U. Edlund and S. Wold, *J. Magnetic Resonance*, 1980, **37**, 183.

[9] M. Sjöström and U. Edlund, *J. Magnetic Resonance*, 1977, **25**, 285.

[10] C. Eaborn, R. Eastmond, and D. R. M. Walton, *J. Chem. Soc.* (*B*), 1971, 127.

[11] L. A. Yanovskaya, G. V. Kryshtal, I. P. Yakovlev, V. F. Kucherov, B. Ya. Simkin, V. A. Bren, V. I. Minkin, O. A. Osipov, and I. A. Tumakova, *Tetrahedron*, 1973, **29**, 2053.

[12] U. Edlund, *Org. Magnetic Resonance*, 1978, **11**, 516.

[13] W. E. Parham and C. D. Wright, *J. Org. Chem.*, 1957, **22**, 1473.

[14] C. M. Suter and G. A. Lutz, *J. Amer. Chem. Soc.*, 1938, **60**, 1360.

[15] H. D. Porter and C. M. Suter, *J. Amer. Chem. Soc.*, 1935, **57**, 2022.

[16] R. Graf, *Annalen*, 1963, **661**, 111.

[17] D. Tabak and P. A. M. Oliveira, *Org. Prep. Proced. Internat.*, 1976, **8**, 243.

[18] (*a*) P. Doyle, R. H. B. Galt, and R. J. Pearce, *Tetrahedron Letters*, 1973, 2903; (*b*) J. N. Marx, D. C. Cringle, *Synth. Comm.*, 1973, **3**, 95.

[19] H. Shechter, J. J. Gardikes, T. S. Cantrell, and G. D. V. Tiers, *J. Amer. Chem. Soc.*, 1967, **89**, 3005.

[20] H. O. House and C. B. Hudson, *J. Org. Chem.*, 1970, **35**, 647.

[21] A.-M. Weidler and G. Bergson, *Acta Chem. Scand.*, 1964, **18**, (*a*) 1487; (*b*) 1484.

[22] (*a*) R. Faure, J. Elguero, E. J. Vincent, and R. Lazaro, *Org. Magnetic Resonance*, 1978, **11**, 617; (*b*) J. Llinares, J.-P. Galy, R. Faure, E.-J. Vincent, and J. Elguero, *Canad. J. Chem.*, 1979, **57**, 937.

[23] (*a*) S. Wold, *Pattern Recognition*, 1976, **8**, 127; (*b*) S. Wold and M. Sjöström in 'Chemometrics: Theory and Practice,' Amer. Chem. Soc. Symp. Ser., 52, ed. B. R. Kowalski, Amer. Chem. Soc. Washington, 1977, p. 243.

[24] S. Wold, *Technometrics*, 1978, **20**, 397.

[25] For a compilation see O. Exner in 'Correlation Analysis in Chemistry: Recent Advances,' eds. J. Shorter and N. B. Chapman, Plenum Press, New York, 1978, ch. 10, p. 439.

[26] G. C. Levy and G. L. Nelson, 'Carbon-13 Nuclear Magnetic Resonance for Organic Chemists,' Wiley, New York, 1972.