JOURNAL OF PHYSICAL ORGANIC CHEMISTRY
J. Phys. Org. Chem. 2003; **16:** 715–720

On the nature of modified composite electrical effect parameters

Marvin Charton[1]* and Barbara I. Charton[2]

[1]Chemistry Department, Pratt Institute, Brooklyn, New York 11205, USA
[2]Coles Science Library, New York University, New York, 10012, USA

Received 27 November 2002; revised 4 March 2003; accepted 5 March 2003

ABSTRACT: It is shown that a correlation of some property, reactivity or biological activity with pure parameters can also be carried out with composite parameters to produce a model with comparable statistics. Modified composite electrical effect parameters can be obtained from known composite electrical effect parameters by means of an algorithm such as

$$\sigma_X^{\blacklozenge} = (\sigma_X + c)^m$$

in which $\sigma_X^{\blacklozenge}$ is the modified composite electrical effect parameter (MCEEP), σ_X the original composite electrical effect parameter, c a constant and m an exponent. MCEEPs were calculated from this equation with c equal to 2 and m ranging from -4 to 4 when σ_X is σ_m or σ_p, and from -2 to 2 when it is σ_p^+. They were used to model 10 sets of chemical reactivities that had previously been correlated with the pure parameters σ_l, σ_d and σ_e, which represent the localized (field), intrinsic delocalized (resonance) and electronic demand sensitivity electrical effects, respectively. In most cases both the pure and the modified composite parameters give comparable fits to the data as measured by the values of $100R^2$, in accord with the prediction. The composition of the MCEEPs was a linear function of m. The advantage in using pure parameters is that they are readily interpretable. Correlations with composite parameters can be interpreted only if their composition has been determined. These results provide an understanding of the way in which topological parameters work. Copyright © 2003 John Wiley & Sons, Ltd.

KEYWORDS: pure parameters; composite parameters; modified composite parameters; topological parameters; electrical effect parameters

INTRODUCTION

The modeling of chemical reactivities, physical and chemical properties and biological activities by statistical methods requires the use of parameters. These parameters are of three types:

1. Pure parameters: these parameters represent a single effect. The σ_l parameter and its equivalent σ_F and σ_I parameters are examples of pure parameters as they all represent the same type of single electrical effect. Another example is a polarizability parameter of the type frequently used in bioactivity modeling.
2. Composite parameters: these parameters represent two or more effects. A composite parameter is characterized by its composition. Thus, the Hammett σ_p constant is an example of a composite electrical effect parameter. The log *P* parameter, which is frequently used in modeling biological activities, is another example.
3. Modified composite parameters: these are composite parameters whose composition has been altered by a mathematical operation.

Our objective in this work was to show that

1. the composition of composite parameters can be changed by mathematical operations on the parameters, and
2. a property which is modeled by some number of pure parameters may also be modeled by some number of composite parameters.

We have derived an equation which indicates that this is the case[1] and now wish to demonstrate its validity. If we are successful in establishing these points, then we can account for the way in which topological parameters[2–4] work. As topological parameters have been of considerable interest in the modeling of physical and chemical properties and of biological activities in the last decade, this is a problem of some importance. A modified form of the derivation based on the triparametric LDR

**Correspondence to:* M. Charton, Chemistry Department, Pratt Institute, Brooklyn, New York 11205, USA.
E-mail: mcharton@pratt.edu

model of the substituent electrical effect[5–7] is shown below. We make no claim for novelty in the mathematics involved in this derivation but in the application to correlation analysis where we believe it will be a powerful tool.

Consider the quantities Q_1 and Q_2, which are different functions of the electrical effect substituent constants σ_λ, representing the 'true' localized electrical effect, σ_δ representing the 'true' intrinsic delocalized electrical effect, and σ_ε representing the 'true' electronic demand sensitivity electrical effect. We may write

$$Q_{1X} = L_1\sigma_{\lambda X} + D_1\sigma_{\delta X} + R_1\sigma_{\varepsilon X} + h_1 \tag{1}$$

$$Q_{2X} = L_2\sigma_{\lambda X} + D_2\sigma_{\delta X} + R_2\sigma_{\varepsilon X} + h_2 \tag{2}$$

with $L_1 \neq L_2, D_1 \neq D_2$ and $R_1 \neq R_2$.

Let some quantity of interest Q_X be correlated with Q_{1X} and Q_{2X}. The correlation equation is

$$Q_X = a_1\ Q_{1X} + a_2Q_{2X} + a_0 \tag{3}$$

From Eqns (1) and (2) and the LDR equation,[5–7] Eqn (4), which represents electrical effects;

$$Q_X = L\sigma_{\lambda X} + D\sigma_{\delta X} + R\sigma_{\varepsilon X} + h \tag{4}$$

on equating coefficients,

$$\begin{array}{ll} L = a_1L_1 + a_2L_2, & D = a_1D_1 + a_2D_2, \\ R = a_1R_1 + a_2R_2, & h = a_1h_1 + a_2h_2 + a_0 \end{array} \tag{5}$$

This result is a special case of a more general relationship. Consider a set of composite parameters κ_i that are linear functions of the pure parameters ζ_j:

$$\kappa_i = \sum_{i=1}^{m}\sum_{j=1}^{n}(a_{ij}\zeta_j + a_{i0}) \tag{6}$$

Thus, for example, when $m = n = 3$, the following equations are obtained:

$$\kappa_1 = a_{11}\zeta_1 + a_{22}\zeta_2 + a_{13}\zeta_3 + a_{10} \tag{7}$$

$$\kappa_2 = a_{21}\zeta_1 + a_{22}\zeta_2 + a_{23}\zeta_3 + a_{20} \tag{8}$$

$$\kappa_3 = a_{31}\zeta_1 + a_{22}\zeta_2 + a_{33}\zeta_3 + a_{30} \tag{9}$$

If some quantity Q is a linear function of κ_i:

$$Q = \sum_{i,k=1}^{p} b_k\kappa_1 + b_0 \tag{10}$$

then it follows from the above that

$$Q = \sum_{l=1}^{q} c_1\zeta_1 + c_0 \tag{11}$$

In general, if some quantity is a linear function of 'pure' independent variables, it is also a linear function of composite independent variables which are themselves a linear function of the appropriate 'pure' independent variables. That is why it is not necessary to have 'pure' parameters, each representing only a single effect, in order to carry out correlation analysis for predictive purposes. Correlations based on 'pure' parameters are directly interpretable. Those based on composite parameters are interpretable only when the composition of these parameters has been determined.

METHOD

We chose electrical effect parameters for this investigation because composite electrical effect parameters are well understood. They are quantitatively described by the LDR equation [Eqn (4)], written in the form

$$\sigma_X = l\sigma_{lX} + d\sigma_{dX} + r\sigma_{eX} + h \tag{12}$$

where σ_X is a composite electrical effect parameter, σ_1 is the localized electrical effect parameter, identical with σ_{I}, σ_d is the delocalized electrical effect parameter and σ_e is the electronic demand sensitivity electrical effect parameter.

Substituent constant composition

The composition of composite electrical effect substituent constants can be described by the C_i value, which is the percentage contribution of the ith pure parameter. It is given by Eqn (3) where b_i are the coefficients of the independent variables in the LDR equation and x_i are the values of these variables for some reference substituent, X°. In this work we used a hypothetical reference substituent for which σ_l, σ_d and σ_e have the values 1, 1 and 0.1, respectively:

$$C_i = \frac{|b_ix_i| \times 100}{\sum_{i=1}^{n}|b_ix_i|} \tag{13}$$

Other measures of composite parameter composition are the per cent delocalized effect, P_D, given by

$$P_D = \frac{100|d|}{(|l| + |d|)} \tag{14}$$

and the electronic demand, η, given by

$$\eta \equiv \frac{r}{d} \tag{15}$$

Table 1. Substituents for which MCEEPs were calculated[a]

H, Me, Et, iPr, tBu, cPr, cHx, C_2H, Vi,[b] Ph, CF_3, CH_2Cl, C_6F_5, CH_2Ph, HCO, Ac, COEt, Bz, CO_2H, CO_2Me, CO_2Et, $CONH_2$, CN, $SiMe_3$, F, Cl, Br, I, OH, OMe, OEt, OCF_3, OAc, OPh, SH, SMe, SEt, SAc, SCF_3, SPh, SOMe, SO_2Me, SeMe, NO_2, *NH_2*, *NMe_2*, *NHPh*, NHAc, N_3, *PMe_2*, $POMe_2$, $PO(OMe)_2$

[a] Groups in italics were excluded from the correlations of the modified σ_p^+ constants.
[b] Vi = vinyl.

Composite parameter modification

Modified composite electrical effect parameters are calculated from composite electrical effect parameters by means of some algorithm. Topological modified composite parameters are often obtained by raising a topological parameter obtained from some algorithm to a power. Those used in this work, designated $\sigma^{\blacklozenge}$, were obtained from the expression

$$\sigma_X^{\blacklozenge} = (\sigma_X + c)^m \tag{16}$$

where $\sigma_X^{\blacklozenge}$ is the modified composite electrical effect parameter σ_X is the composite electrical effect parameter that is to be modified, c is a constant chosen so as to make the sum $(\sigma_X + c)$ positive for all of the groups in the data set, which avoids the loss of the negative sign that would otherwise occur if they were raised to a whole even integer, and m is the exponent.

Values of $\sigma_X^{\blacklozenge}$ were calculated from Eqn (16) for the σ_p^+, σ_p and σ_m constants of the substituents in Table 1. A value of 2 was chosen for c. Values of m ranged from −4 to 4 for σ_m and σ_p and from −2 to 2 for σ_p^+. Results of correlations of the $\sigma_X^{\blacklozenge}$ constants with Eqn (12) are given in Tables 2–4 and values of η, P_D, C_l, C_d and C_e in Tables 5–7.

Table 2. Results of correlations of modified σ_p^+ constants with the LDR equation[a,b]

m	$100R^2$	$A100R^2$	F	S_{est}	$S°$
−2	89.95	89.51	131.3	0.0492	0.331
−0.5	95.18	94.97	289.6	0.0201	0.229
−0.2	95.92	95.74	344.8	0.00889	0.211
0.2	96.78	96.64	440.7	0.0102	0.187
0.5	97.11	96.98	492.9	0.0292	0.178
2	97.21	97.09	511.7	0.321	0.174
6	84.95	84.28	82.80	0.509	0.405

[a] All sets in this table have 48 data points.
[b] The statistics reported are $100R^2$, the percentage variance of the data accounted for by the regression equation; $A100R^2$, $100R^2$ adjusted for the number of independent variables; F, formally, the probability that the coefficients of a regression are different from zero, used as a measure of the goodness of fit; S_{est}, the standard error of the estimate; $S°$, the standard error of the estimate divided by the root mean square of the data.

In order to determine the composition of the modified composite substituent parameters $\sigma_X^{\blacklozenge}$ as a function of m, the η, P_D, C_l, C_d and C_e values were correlated with the equation

$$\sigma_X^{\blacklozenge} = a_1 m + a_0 \tag{17}$$

Table 3. Results of correlations of modified σ_p constants with the LDR equation[a,b]

m	$100R^2$	$A100R^2$	F	S_{est}	$S°$
−4	76.81	75.86	53.00	0.0345	0.501
−2	88.25	87.77	120.1	0.0363	0.357
−0.5	94.84	94.63	294.0	0.0151	0.236
0.5	97.42	97.31	603.9	0.0209	0.167
2	98.48	98.42	1036	0.188	0.128
4	95.80	95.63	365.2	0.291	0.213

[a] All sets in this table have 52 data points.
[b] See footnote b, Table 2.

Table 4. Results of correlations of modified σ_m constants with the LDR equation[a,b]

m	$100R^2$	$A100R^2$	F	S_{est}	$S°$
−4	84.83	84.21	89.47	0.00781	0.405
−2	90.90	90.53	159.9	0.0134	0.314
−0.5	94.08	93.84	254.2	0.00850	0.253
0.5	95.40	95.22	332.2	0.0164	0.223
2	96.00	95.83	383.6	0.202	0.208
4	94.11	93.87	255.7	0.250	0.253

[a] All sets in this table have 52 data points.
[b] See footnote b, Table 2.

Table 5. Composition of $\sigma^{\blacklozenge}$ parameters derived from σ_p^+ constants

m	−2	−0.5	−0.2	0.2	0.5	2
η	2.48	2.12	2.06	1.97	1.88	1.48
P_D	69.3	64.5	63.2	61.8	61.2	57.8
C_l	26.2	31.3	32.6	34.1	34.8	38.9
C_d	59.1	56.7	55.9	55.1	54.9	53.2
C_e	14.7	12.0	11.5	10.8	10.3	7.89

Table 6. Composition of $\sigma^{\blacklozenge}$ parameters derived from σ_p constants

m	−4	−2	−0.5	0.5	2	4
η	3.37	2.18	1.41	0.970	0.416	−0.140
P_D	58.5	54.0	51.3	49.7	47.8	45.9
C_l	34.6	41.2	45.4	48.0	51.1	53.8
C_d	48.9	48.3	47.9	47.4	46.9	45.6
C_e	16.5	10.5	6.73	4.60	1.95	0.637

Table 7. Composition of $\sigma^{\blacklozenge}$ parameters derived from σ_m constants

m	−4	−2	−0.5	0.5	2	4
η	2.30	1.77	1.39	1.18	0.879	0.535
P_D	28.4	28.0	28.1	28.0	27.9	27.8
C_l	67.2	68.6	69.2	69.8	70.4	71.1
C_d	26.7	26.7	27.0	27.0	27.2	27.4
C_e	6.14	4.73	3.75	3.18	2.39	1.47

Table 8. Results of correlations with Eqn (17)[a,b]

Q	a_1	S_{a1}	a_0	S_{a0}	a_0[c]	$100R^2$	F	S_{est}	S°
σ_p^+									
η	−0.249	0.00622	2.00	0.00744	1.71	99.75	1605	0.0182	0.0611
P_D	−2.91	0.192	63.0	0.229	59.4	98.29	230.2	0.561	0.160
C_l	3.20	0.146	33.0	0.174	36.8	99.18	482.5	0.427	0.111
C_d	−1.50	0.120	55.8	0.143	53.9	97.52	157.2	0.350	0.193
C_r	−1.70	0.0286	11.2	0.0342	9.24	99.89	3537	0.0839	0.0412
σ_p									
η	−0.439	0.0347	1.37	0.0902	0.840	97.56	160.1	0.221	0.191
P_D	−1.66	0.104	51.1	0.270	49.2	98.45	254.2	0.662	0.152
C_l	2.42	0.207	45.7	0.538	48.7	97.14	136.1	1.32	0.207
C_d	−0.401	0.0368	47.5	0.0956	47.3	96.74	118.8	0.234	0.221
C_r	−2.02	0.244	6.82	0.634	4.00	94.46	68.17	1.55	0.208
σ_m									
η	−0.221	0.0107	1.34	0.0278	1.72	99.07	425.0	0.0682	0.118
P_D	−0.0654	0.0157	28.0	0.0409	27.4	81.20	17.28	0.100	0.531
C_l	0.481	0.0332	69.4	0.0863	69.3	98.13	210.0	0.211	0.167
C_d	0.0938	0.0120	27.0	0.0311	26.2	93.86	61.10	0.0764	0.304
C_r	−0.584	0.0276	3.61	0.0717	4.49	99.12	448.1	0.176	0.115

[a] All data sets have 6 data points.

[b] S_{a1} and S_{a0} are the standard errors of the coefficients. For the other statistics see footnote 2 of Table 3.

[c] From the correlation of σ_p^+, σ_p and σ_m with Eqn (12).

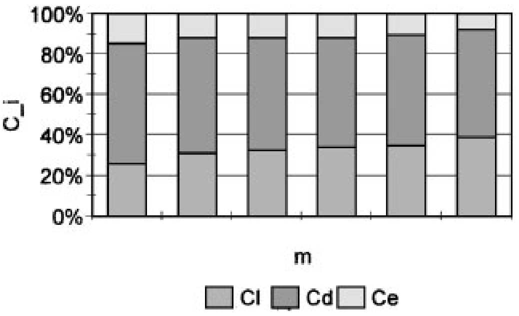

Figure 1. Composition of the modified composite substituent parameters obtained from the σ_p^+ constants; m ranges from −2 to 2

by means of simple linear regression analysis. The results of the correlation are reported in Table 8. They show a linear dependence of all of the measures of composition, η, P_D, C_l, C_d and C_e, on m. The a_0 values obtained in the correlation are equivalent to the values of η, P_D, C_l, C_d and C_e for the original composite substituent constant for which $m = 0$. Values of η, P_D, C_l, C_d and C_e for the σ_m, σ_p and σ_p^+ constants were calculated from the l, d and r values reported[8] for their correlation with Eqn (12). The values are reported in Table 8. They are in reasonable agreement with the a_0 values. The composition in terms of the C_i values is shown in Figs 1–3.

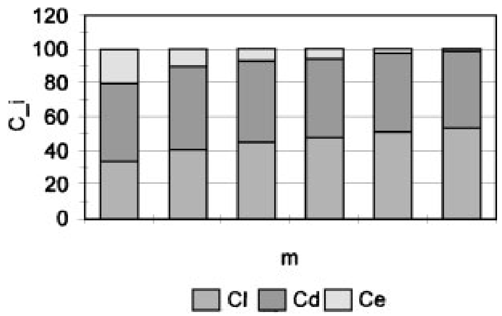

Figure 2. Composition of the modified composite substituent parameters obtained from the σ_p constants; m ranges from −4 to 4

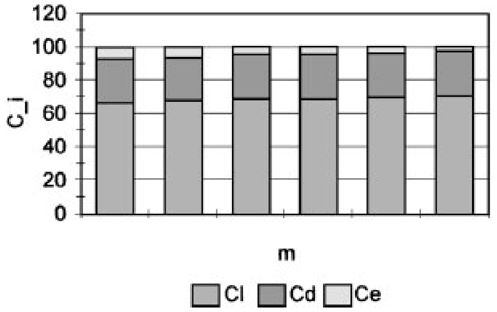

Figure 3. Composition of the modified composite substituent parameters obtained from the σ_m constants

It is of interest that as the magnitude of m increases, the value of $100R^2$ for the correlation of the MCEEPs with Eqn (4) decreases. This suggests that there may be a limit to the useful range of m in this algorithm.

RESULTS

Data sets taken from the literature[2] that had been correlated with Eqn (4) by means of multiple linear regression analysis were correlated with the modified composite parameter (MCP) equation:

$$Q_X = \sum_{i=l}^{n} \rho_i \sigma^{\blacklozenge}_{iX} + h \qquad (18)$$

Table 9. Values of $100R^2$ for correlations with the LDR and MCP equations[a]

Set	XGY, Rgt., *Q*, Sv	$100R^2$		Parameters used: type; *m*
		LDR	MCP	
P32	4-XPnOH,–,pK_a, H_2O	98.51	94.39	p^+; −2, −0.5
			94.73	p; −2, 2, 4
P120	(*E*)-2-XVnCO$_2$H,–, pK_a, H_2O	99.40	98.13	p^+; −2, −0.2, 0.2
P140	4-XPnCH$_2$OBz, OH^-, log *k*, aq. MeOAc	99.57	100	p^+; −2, −0.2, 2
P152	4-XPnOBz, OH^-, log *k* aq. EtOH	99.51	96.57	p^+; −2,−0.2
P248	4-XPnOCH$_2$Vi,–, log *k*, EtO(CH$_2$CH$_2$O)H	98.17	99.14	p^+; −2, −0.2, 0.2
P249	4-XPnCO$_2$Et, OH^-, log *k*, aq. EtOH	97.99	94.97	p^+; −2, 2
P268	4-XPy, EtI, log *k*, MeNO$_2$	98.47	98.41	p^+; −2, −0.5, 0.5, 2
P269	4-XPy, EtI, log *k*, CH$_2$Cl$_2$	97.84	97.86	p^+; −2, −0.5, 0.5, 2
P281	4-XPnNH$_3{}^+$,–, pK_a, H_2O	97.71	86.24	p^+; −2, −0.5, 0.5
			90.05	p; −4, −2
P304	4-XPyH$^+$,–, pK_a, H_2O	97.63	98.74	p^+; −2, −0.2, 0.2, 2

[a] The set numbers correspond to those in Ref 2, Tables 3 and 4. Modified composite parameters derived from σ_p^+ are labeled p^+, those derived from σ_p are labeled p. The *m* values of the $\sigma_X^\blacklozenge$ parameters follow the substituent type from which they were derived. Abbreviations: Rgt, reagent, Sv, solvent; 4-Pn, 1,4-phenylene; (*E*)-2-Vn, *trans*-1,2-vinylene; 4-Py, 1,4-pyridinylene; Vi, vinyl.

in the form

$$Q_X = \rho_1\sigma_{1X}^{\blacklozenge} + \rho_2\sigma_{2X}^{\blacklozenge} + \rho_3\sigma_{3X}^{\blacklozenge} + \rho_4\sigma_{4X}^{\blacklozenge} + h \qquad (19)$$

The data sets studied are given in Ref 2 and a comparison of the results is given in Table 9. Only modified parameters of the same type were used in any one correlation in order to prove a more rigorous test. Modified parameters from two or more different types of Hammett substituent constants would have much more variation in composition and should therefore give better results. The results obtained show that, the MCP equation generally gives results comparable to the LDR equation, in agreement with our predictions. The data sets that gave the poorest fit to the MCP equation were those involving the ionization constants of anilinium ions and phenols. Data sets such as these with a delocalized effect donor reaction site have a large negative value of η. The MCEEPs first used in the correlation of these data sets were derived from the σ_p^+ constants for which η is strongly positive. Correlation with MCEEPs derived from the σ_p constants gave better results in accord with their less positive η values. We believe that the use of MCEEPs calculated from the σ_p° constants which have negative values of η would probably give results with the MCP equation that are comparable to those obtained with the LDR equation. Alternatively, the use of MCEEPs based on both the σ_p and σ_m constants should give results comparable to those obtained with the LDR equation.

DISCUSSION

Our results show that modified composite parameters can be calculated from appropriate algorithms applied to ordinary composite parameters. They can be used to model chemical reactivities and by extension chemical and physical properties and biological activities. They can be useful in obtaining empirical predictive relationships. These can be interpreted only if the modified composite parameters can be related to the pure parameters of which they are composed. We believe that this explains the way in which topological parameters work. They are actually counts of the numbers of atoms of each kind, of bonds and of valence electrons. They also count structural features such as branches on the longest chain of a substituent. These counts can be related to steric effects, bond moments and polarizability, the fundamental parameters actually causing the relationship.

CONCLUSIONS

Modified composite electrical effect parameters can be generated by algorithms such as Eqn (16). We have limited this investigation to that algorithm, but there are many others that would work just as well.

The composition of the modified composite electrical effect parameters obtained from the algorithm defined by Eqn (16) varies linearly with *m*.

The modified composite electrical effect parameters obtained can be used in the MCP equation to correlate chemical reactivity data sets. They should be applicable to chemical and physical properties and also to biological activities.

Correlations with the MCP equation result in empirical relationships which are useful for prediction by interpolation but are not directly interpretable unless they have been resolved into their pure components.

REFERENCES

1. Charton M, Greenberg A, Stevenson TA. *J. Org. Chem.* 1985; **50**: 2643–2646.
2. Kier LB, Hall LH. *Molecular Connectivity in Chemistry and Drug Research*. Academic Press: New York, 1976.
3. Merrifield RE, Simmons HE. *Topological Methods in Chemistry.* Wiley–Interscience: New York, 1989.
4. Balaban A (ed). *From Chemical Topology to Three-dimensional Geometry.* Plenum Press: New York, 1997.
5. Charton M. *Prog. Phys. Org. Chem.* 1987; **16**: 287–315.
6. Charton M. In *The Chemistry of Dienes and Polyenes*, Rappoport Z (ed). Wiley: Chichester, 1997; 683–732.
7. Charton M. In *The Chemistry of Germanium, Tin and Lead Compounds*, Rappoport Z (ed). Wiley: Chichester, 2002; 537–578.
8. Charton M. In *The Chemistry of the Functional Groups. Supplement A, The Chemistry of Double Bonded Functional Groups*, vol. **2**, part 1, Patal S, Rappoport Z (eds). Wiley: Chichester, 1989; 239–298.