THE NON-ADDITIVE SUBSTITUENT EFFECTS
OF TWO SUBSTITUENTS ADJACENT TO THE REACTION CENTRE

O.EXNER

Institute of Organic Chemistry and Biochemistry,
Czechoslovak Academy of Sciences, 166 10 Prague 6

The cumulative effect of two substituents may be expressed either by the additive relationship, Eq. (*1*), or by correlation equations (*2*)—(*6*) of increasing complexity. Some properties of these equations are examined, in particular the degrees of freedom in relation to the necessary extent of experimental data, the allowed transformations of parameters, and their computation by iteration.

The available experimental data concerning effects of two substituents adjacent to the reaction centre do not allow more complex correlation equations than (*1*)—(*4*). Investigation of 15 reaction series including experimental properties as pK's, rate constants, IR frequencies, NMR shifts, and ionization potentials revealed that the simplest non-additive equation (*2*) yields frequently the best fit. The substituted effects are generally not additive; an approximately additive behaviour is observed only when the substituents are relatively distant and their effects weak, *e.g.* in the Hammett-type compounds.

The cumulative action of several substituents on one reaction centre represents one of the most general problems in structure-reactivity correlations. The common approach is based on the correlation with a fixed scale of empirical parameters which hold for the corresponding monoderivatives; a well-known example are Hammett's σ constants. By applying the same empirical constants to polyderivatives, one can decide whether the action of substituents is simply additive, or whether there is a saturation effect or its opposite[1,2]. This approach is necessarily confined to particular more complex structures, for which the respective correlation equation holds, *e.g.* in the case of Hammett equation to benzene *meta* and *para* derivatives; the effects of remote substituents are relatively small and a significant deviation from additivity is hardly detectable[2,3]. On the other hand, the effects of substituents directly adjacent to the reaction centre are more significant and theoretically more interesting but they cannot be correlated with known substituent constants in a simple manner[4-6], or the correlations described are quite restricted and imprecise[7].

Another possible approach does not refer to any fixed scale of constants but these are defined by the correlated data themselves. In this way the analysis of controlling factors is either abandoned or postponed, attention being focused to the type of correlation equation and to the interaction of substituents. The simplest possible type

of correlation equation is the additive scheme which in the case of two substituents takes the form:

$$y_{ij} = c_i + c_j + \varepsilon_{ij} . \tag{1}$$

Here ε_{ij} is a random variable which may be sometimes identified with the experimental error but more frequently expresses the imprecision of the correlation equation. The observed quantity y_{ij} is reproduced by the two empirical substituent constants c which are adjusted in order to minimize the sum of squares $\sum \varepsilon^2$. If Eq. (1) seems to be inadequate, more sophisticated equations, like Eqs $(2)-(6)$ in Table I, may be successively tested. From the mathematical point of view, Eqs. (1) is a simplified case of the general additive relationship[8] in which, on the other hand, the number of empirical constants is large with respect to the number of experimental data. It follows that the data matrix $Y \equiv (y_{ij})$ must be relatively extensive and reasonably filled. This applies the more to Eqs $(2)-(6)$.

The above approach has been used many times in studies of substituent effects[5,9−13] which, however, remained mostly restricted to the additive relationship as the only possibility[14−20]. The present paper deals generally with the cumulative effect of two substituents, but attention is focused to those immediately connected to a reaction centre represented by the simplest divalent groups: CH_2, CO, NH, S, SO_2, COO. As the experimental quantity, chemical as well as physical properties are investigated; the selection of reaction series depends mainly on the availability of suitable data. These data will be tentatively processed using the additive relationship, Eq. (1), as well as non-additive relationships, Eqs $(2), (3)$ and (4). Before proceeding to the results, some general properties of these correlation equations are to explored with particular respect to the data needed for a respective equation.
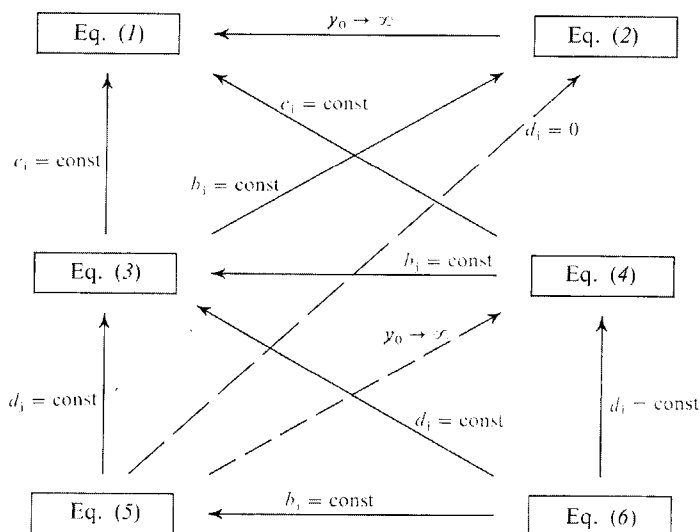
Form of the correlation equation. If the simplest form of the correlation, Eq.(1), has been found inadequate, there are two possibilities to improve the fit. Either the range of validity may be restricted by excluding some data, arranging them into subgroups *etc.*, this approach is usually based on chemical arguments and will be mentioned in connection with some practical examples. Alternatively, a more complex correlation equation is to be searched for in which each substituent may be given even more than one parameter. For practical purposes it is desirable to have a sequence of equations of increasing complexity, *i.e.* increasing number of constants; each should be a logical extension of the previous one and reducible to it by giving some parameters a particular constant value. This should guarantee that the fit with any equation cannot be significantly worse than with the preceding one. An attempt of such classification is given in Table I, the mutual relations of Eqs $(1)-(6)$ are visualized in Scheme 1.

Theoretical forms of the correlation equations have been derived by Wold[21] who assumed that each substituent (*i.e.* macroscopic variable) is characterized

by several numerical properties (*i.e.* microscopic variables). According to the number of these microscopic variables and according to the degree of approximation of the Taylor expansion either one of the equations (*1*), (*2*), (*3*) is produced or the equation of the type

$$y_{ij} = \sum_{a=1}^{A} c_{ia}c_{aj} + \varepsilon_{ij} .$$ (7)

This type actually embraces our Eqs (*4*)—(*6*), *e.g.* for $A = 3$ and $c_{1j} = $ const Eq. (*6*) is produced, or for $A = 3$, $c_{1j} = $ const and $c_{i2} = $ const Eq. (*4*). We preferred the more detailed classification of Table I since the individual equations have been used here and there in the literature. Thus, Eq. (*2*) was applied several times to substituent effects[6,10,11]. Eq. (*3*) represents the well-known type of the Hammett equation if this is treated as the problem of determining simultaneously all the constants σ, ϱ, and $\log k^0$ from the whole body of data[22]. A similar treatment of the Taft two-parameter equation[23,24] corresponds to Eq. (*6*). Otherwise the more complex Eqs (*4*) and (*6*) are known merely from treatment of solvent effects[25,26]. In some cases even equations with more terms were applied[27,28]; the pertinent statistical procedures are called factor analysis[27] and principal-component analysis[28].

SCHEME 1

Relations between the Correlation Equations (*1*)—(*6*)

In the field of substituent effect attention is to be focussed to the simple types of correlation equations as follows particularly from the subsequent analysis of degrees of freedom. These simplest types, Eqs (*1*)—(*3*), may be represented graphically within two dimensions, if the

observed quantity y_{ij} is plotted against the empirical parameter c_j (Fig. 1). One obtains a family of straight lines, one for each value of i. The lines are parallel in the case of Eq. (*1*), intersect in one point in the case of Eq. (*2*), and have no common point in the case of Eq. (*3*). In this way the three cases may be distinguished by a graphical test: One straight line is arbitrarily given the unit slope and the observed values of one selected row (column) of the matrix are plotted on it as ordinates; the abscissae of these points define the parameters c_j. The observed values from other rows (columns) may be then plotted against them on the ordinate, yielding one of the patterns in Fig. 1. Eq. (*4*) may be tested in this way after transforming the original matrix $Y \equiv (y_{ij})$ into the difference matrix $Y \equiv (y_{ij} - y_{1j} - y_{i1})$, *i.e.* subtracting one selected row and one selected column from the others. This procedure reduces Eq. (*4*) into Eq. (*2*). Correlations according to more complex equations cannot be represented by a two-dimensional graph and the matrix itself does not allow their easy identification either. An evidence is presented by the examples in Table II: Only from the first three the appurtenance to the respective equation is evident.

Before proceeding to further properties of the correlation equations (*1*)−(*6*), the symmetry of the data matrix is to be considered which is reflected in the symmetry of the pertinent equations. A symmetrical matrix arises if the indices i, j are the same for rows and columns $(i = 1, 2, ..., p, \ j = 1, 2, ..., p)$ and the equality $y_{ij} = y_{ji}$ holds. In the case of substituent effects this is fulfilled always when substitution occurs in equivalent positions (*e.g.* functional groups $CH_2, CO, NH, NH_2^{(+)}$, S, and SO_2). The number of rows and columns is evidently equal (p) and the number of (independent) entries in a full matrix is $N = (p^2 + p)/2$.

An inherently unsymmetrical matrix arises if the two substituents are in non-equivalent positions (*e.g.* the functional group COO), or if there are two variables, different in character (solvent–solute, substituent–reagent, *etc.*) The indices for rows $(i = 1, 2, ..., p)$ and for columns $(j = p + 1, \ p + 2, ..., p + q)$ are different and the number of rows and columns is generally not equal. A full matrix has pq entries. In addition an essentially symmetrical matrix may become effectively unsymmetrical if no index is common to both a row and a column (*i.e.* the same condition as above; $i = 1, 2, ..., p; \ j = p + 1, \ p + 2, ..., p + q)$, the reaction centre itself may be symmetrical. An example is given by the carbonyl compounds X^1COX^2 restricted to derivatives of carboxylic acids, *i.e.* with $X^1 =$ alkyl or aryl, $X^2 = OR, NR_2$, Cl *etc.* The symmetry of the matrix has obvious consequences for the number of empirical constants, degrees of freedom, *etc.* Note that equations (*3*) and (*6*) are inherently unsymmetrical due to different number of constants for variables i and j, respectively.

Computation of empirical parameters. The "best" values of the parameters can be defined in different way. We preferred the simple postulate

$$S = \sum_{ij} w_{ij} \varepsilon_{ij}^2 = \min, \tag{8}$$

where the weights w_{ij} equal 1 or 0, respectively, according to whether the pertinent value is known or not. This postulate allows easily to compare results obtained with

individual equations $(1)-(6)$ and gives all observed data equal weights. Another possibility is to give all the rows and columns of the matrix equal weights and to require that sums of squared residuals in each of them should be simultaneously

TABLE I

The Simplest Types of Correlation Equations with Two Variable Factors

Equation	Symmetry[a]	M[b]
$(1)\quad y_{ij} =$ $= c_i + c_j + \varepsilon_{ij}$	U	$p + q - 1$
	S	p
$(2)\quad y_{ij} =$ $= y_0 + c_i c_j + \varepsilon_{ij}$	U	$p + q$
	S	$p + 1$
$(3)\quad y_{ij} =$ $= b_i + c_i c_j + \varepsilon_{ij}$	U	$2p + q - 2$
$(4)\quad y_{ij} =$ $= b_i + b_j + c_i c_j + \varepsilon_{ij}$	U	$2p + 2q - 4$
	S	$2p - 1$
$(5)\quad y_{ij} =$ $= y_0 + c_i c_j + d_i d_j + \varepsilon_{ij}$	U	$2p + 2q - 3$
	S	$2p$
$(6)\quad y_{ij} =$ $= b_i + c_i c_j + d_i d_j + \varepsilon_{ij}$	U	$3p + 2q - 6$

[a] S symmetrical matrix of data, U unsymmetrical matrix; [b] number of independent parameters; [c] the new parameters after the transformation are marked with an asterisk; [d] recommended transformations yielding the "normalized" values of parameters; the normalization in parentheses

as small as possible[21,22]; this reasoning is connected with the so-called principal-
-component model.

TABLE I
(*Continued*)

Possible transformations of parameters[c]	Normalization[d]
$c_i = c_i^* + A$ $c_j = c_j^* - A$	$c_x^* = 0$
—	—
$c_i = Ac_i^*$ $c_j = c_j^*/A$	$c_x^* = 1$
—	—
$c_i = Ac_i^*$ $c_j = c_j^*/A + B$ $b_i = b_i^* - ABc^*$	$c_x^* = 1$ $c_y^* = 0$
$c_i = A(c_i^* + B)$ $c_j = (c_j^* + C)/A$ $b_i = b_i^* - Cc_i^* + D$ $b_j = b_j^* - Bc_{j_1}^* - BC - D$	$c_x^* = 0$ $c_y^* = 0$ $(c_z^* = 1)$ $b_x^* = 0$
$c_i = c_i^* + A$ $b_i = b_i^* - Ac_i - A^2/2$	$c_x^* = 0$
$c_i = A(c_i^* \cos \varphi + d_i^* B \sin \varphi)$ $c_j = (c_j^* \cos \varphi + d_j^* \sin \varphi/B)/A$ $d_i = C(d_i^* \cos \varphi - c_i^* \sin \varphi/B)$ $d_j = (d_j^* \cos \varphi - c_j^* B \sin \varphi)/C$	$c_x^* = 0$ $c_y^* = 0$ $(c_z^* = 1)$ $d_x^* = 1$
$c_i = c_i^* \cos \varphi + d_i^* \sin \varphi$ $d_i = d_i^* \cos \varphi - c_i^* \sin \varphi$	$c_x^* = 0$
$c_i = A(c_i^* \cos \varphi + d_i^* B \sin \varphi)$ $c_j = (c_j^* \cos \varphi + d_j^* \sin \varphi/B + D)/A$ $d_i = C(d_i^* \cos \varphi - c_i^* \sin \varphi/B)$ $d_j = (d_j^* \cos \varphi - c_j^* B \sin \varphi + E)/C$ $b_i = b_i^* + (E \sin \varphi/B - D \cos \varphi) c_i^* -$ $\quad - (BD \sin \varphi + E \cos \varphi) d_i^*$	$c_x^* = 0$ $c_y^* = 0$ $(c_z^* = 1)$ $d_y^* = 0$ $(d_z^* = 1)$ $b_x^* = 0$

need not be always useful; the indices x, y, z refer to certain standard rows or columns (*i.e.*
standard substituents, reactions *etc.*)

Eq. (8) implies the normal equations which are different in number and complexity in the cases of models $(1)-(6)$; the searched parameters could be in principle obtained by their solution. In the simplest case, Eq. (1) and the unsymmetrical matrix, the normal equations read

$$c_i \sum_j w_{ij} + \sum_j w_{ij}c_j = \sum_j y_{ij} \,,$$

$$(9)$$

$$c_j \sum_i w_{ij} + \sum_i w_{ij}c_i = \sum_i y_{ij} \,.$$

For the full matrix $(w_{ij} = 1)$ Eqs (9) have the solution

$$c_i = \frac{1}{q} \sum_j y_{ij} + A \,,$$

$$(10)$$

$$c_j = \frac{1}{p} \sum_i y_{ij} - \frac{1}{pq} \sum_{ij} y_{ij} - A \,,$$

where A is an arbitrary constant. The residual sum of squares is given by

$$S = \sum_{ij} y_{ij}^2 - \frac{1}{q} \sum_i (\sum_j y_{ij})^2 - \frac{1}{p} \sum_j (\sum_i y_{ij})^2 + \frac{1}{pq} (\sum_{ij} y_{ij})^2 \,.$$

$$(11)$$

In case of Eq. (2) the normal equations read (unsymmetrical matrix):

$$c_i \sum_j w_{ij}c_j^2 + y_0 \sum_j w_{ij}c_j = \sum_j c_j w_{ij} y_{ij} \,,$$

$$c_j \sum_i w_{ij}c_i^2 + y_0 \sum_i w_{ij}c_i = \sum_i c_i w_{ij} y_{ij} \,,$$

$$(12)$$

$$\sum_{ij} w_{ij}c_ic_j + y_0 \sum_{ij} w_{ij} = \sum_{ij} w_{ij} y_{ij} \,.$$

For more complex models and/or for an incomplete data matrix the explicit solution seems to be not possible. The best values of the parameters were obtained by iteration procedures and the final residual S from the differences of calculated and observed values (see Appendix, and refs[22-27]).

Eqs (10) reveal that the constants c_i, c_j are not obtained unambiguously but one value can be arbitrarily adjusted. This is well-known from the Hammett equation[22], where the choice $\sigma = 0$ for hydrogen and $\varrho = 1$ for benzoic acids is usual. The number of adjustable parameters increases with the complexity of correlation equation since the parameters pertinent to one variable $(e.g.\ b_i,\ c_i,\ d_i)$ can substitute one another in various manner. This problem has not yet received proper attention[24] and some mentions are not quite correct[26]. In Table I all admissible transformations of para-

meters are summarized for each model, the resulting number of independent para-
meters (M) is also given. This number is also visualized in Table II, where the inde-
pendent entries necessary to generate the whole matrix are demarkated by a broken
line; their number is equal to M. For instance, in the case of Eq. (6) each additional
column is unambiguously described by two entries, each additional row by three.
Together, there would be $3p + 2q$ independent entries (*i.e.* in two rows and three
columns) but six entries are counted twice.

The allowed transformations must be respected in all attempts to give the para-
meters any physical meaning. It is well known that the zero point of the σ scale in the
Hammett equation is arbitrary but the situation is much more complex with the
more sophisticated relationships[24,26]. When the empirical constants are to be tabulat-
ed, it is desirable to transform them in a standard manner, in order to obtain com-
parable quantities in each case. According to the correlation equation used, one
or several constants can attain arbitrary values, sometimes even zero. The simplest
kind of normalization is to give the reference substituent (usually hydrogen), or
standard solvent, reagent, *etc.* the value 0, or − if this is not possible − 1. With more
complex equations still further standards are needed[26]. In Table I the recommended
normalization transformations are listed which give the constants the most con-
ceivable values.

TABLE II
Examples of Simple Data Matrices Corresponding Exeactly to Correlation Equations $(1)-(6)$*
*   The broken line demarcates entries which are sufficient to describe the whole table. Their
number M, equal to the number of independent empirical constants, is given at the bottom of each
matrix as function of the matrix dimensions p and q.

Eq. (1)

1	2	4	5
11	12	14	15
21	22	24	25
41	42	44	45

$M = p + q - 1$

Eq. (2)

1	2	4	5
11	14	20	23
21	26	36	41
41	50	68	77

$M = p + q$

Eq. (3)

1	2	4	5
11	14	20	23
21	20	18	17
41	38	32	29

$M = 2p + q - 2$

Eq. (4)

1	2	4	5
11	14	10	17
21	20	28	23
41	38	50	41

$M = 2p + 2q - 4$

Eq. (5)

1	2	4	5
11	14	10	17
21	20	53	38
41	38	102	71

$M = 2p + 2q - 3$

Eq. (6)

1	2	4	5
11	14	10	17
21	20	53	38
41	38	2	11

$M = 3p + 2q - 6$

When M is known, the number of degrees of freedom f is obtained as

$$f = N - M , \qquad (13)$$

where $N = \sum w_{ij}$ is the number of all observed values. From S and f the standard deviations $s_1 - s_6$ are obtained corresponding to equations $(1)-(6)$, respectively. They can be compared in an empirical manner, or by an approximate F-test as far as there are many degrees of freedom. Strictly the F-test cannot be applied since the equations $(2)-(6)$ are not linear.

The data required. The number of degrees of freedom alows to formulate the requirements as to the set of observed data. These are given in the form of a symmetrical or unsymmetrical matrix (see above) which is rarely full; more frequently some entries are not available or even cannot be obtained at all. The data matrices obtainable in organic chemistry may be often filled to some 60% or even less. A better filling may be achieved by excluding the row and/or columns with few entries; in this way the body of data attains a better consistency but the degrees of freedom are further reduced.

TABLE III

Necessary Size of Data Matrix for Treatment by Empirical Equations $(1)-(6)$*

Eq. and symmetry[b]	Parameters[c]	Full matrix[a]			50% Filling[a]		
		min[d]	$N \geq 2M^e$	$N \geq 3M^e$	min[d]	$N \geq 2M^e$	$N \geq 3M^e$
(1) U	2	2	4	6	4	8	12
(2) U	2	3	4	6	5	8	12
(3) U	3	3	6	9	6	12	18
(4) U	4	3	7	11	8	15	23
(5) U	4	4	8	12	8	16	24
(6) U	5	4	9	14	9	19	29
(1) S	2	2	3	5	4	7	11
(2) S	2	3	4	6	5	8	12
(4) S	4	3	7	11	7	15	23
(5) S	4	4	7	11	8	15	23

[a] The least possible size of the data matrix is given; in the case of the unsymmetrical matrix equal dimensions are assumed $(p = q)$; [b] S symmetrical, U unsymmetrical; [c] number of empirical constants necessary to express one observed datum, excluding the constants common for the whole set (y_0); [d] absolute minimum with degrees of freedom $f > 0$; [e] number of all observed data at least twice (three times) greater than the number of independent empirical constants M.

In order to obtain statistically significant results, the degrees of freedom f must be in relation to the number of empirical parameters (M). A reasonable postulate, with respect to the obtainable data, may be $f \geqq 2M$, i.e. $N \geqq 3M$ (three entries to a constant); as the utmost limit one can accept $N \geqq 2M$. The latter value corresponds roughly to the requirement[21] that the dimension of the observation matrix should be at least three times the number (A) of terms in Eq. (7). In Table III the minimum dimensions of the matrix are listed for the requirements $N \geqq 2M$ and $N \geqq 3M$, respectively. The absolute minimum which allows the constants to be computed (i.e. $f > 0$) is also given. It is seen that the necessary number of data rapidly increases with the complexity of the correlation equation. The required dimension of the matrix is roughly proportional to the number of parameters necessary to express one observed value (second column of Table III). The table also reveals the advance of a full or reasonably filled matrix.

Summarizing we may state that more complex correlation equations call for relatively numerous and systematic experimental data which are rarely available in the case of substituent effects. In particular the systematic combination of substituents is usually absent so that extending the matrix makes the situation only worse. Although in the literature even sets with less degrees of freedom and more constants were processed,* we insist on the condition of two entries to one constant as the limit. Therefore, our calculation must be restricted to Eqs $(1)-(4)$, most attention being focused to distinguishing between the additive behaviour. Eq. (1), and the simplest type of a non-additive relation, Eq. (2).

PROCEDURES

Selection of data. The quantities observed and the pertinent model compounds (Table IV) were chosen rather arbitrarily, the simplest functional groups and extensive sets of data being generally preferred. Certain less extensive series (e.g. $1-3$) were investigated, too, since the correlations are very close and/or the property followed very important. In each series care was paid to obtain a reasonably filled matrix but simultaneously to include as diverse substituents as possible. Since the task is to investigate effects of adjacent substituents, the large series of groups differing only on a remote site were restricted, e.g. from various substituted phenyls only one or two were usually retained.

Our attention was focused to the general pattern, hence the number of substituents was not so important as elimination of gross errors, since these can invalidate the results to a great extent. In but few cases it was possible to take the whole body

---

* For instance the treatment of solvent effects[26] according to Eq. (6) is based on 171 observed data and involves 105 independent constants. This ratio is unfavourable, although the set is one of the largest available in kinetics.

of data from one experimental work (series 2, 3, 5, 14), more frequently critical collections were exploited, but collecting of individual items from different sources without the possibility of adjusting was avoided.

Dissociation constants of C-acids $R^1CH_2R^2$ in water at 25°C (series *1*) were taken from two critical tabulations[4,29] and were not statistically corrected since the correction would be constants. The substituents included were: NO_2, CN, $COCH_3$, COC_6H_5, $COOC_2H_5$, SO_2CH_3. In the case of carbonyl compounds the dissociation constants of the pure keto form was preferred as far as it was available. As the data are scarce, some less reliable ones could not be avoided.

The two characteristic frequencies of the methylene group in dihalomethanes, twisting (series *2*) and rocking (series *3*) were collected and reassigned by a single author[30]. IR and Raman mea-

TABLE IV

Reaction Series Investigated as to the Cumulative Substituent Effects in a Molecule R^1ZR^2

No	Reaction centre Z^a	Experimental quantity^a	Matrix dimensions^b	N^c	Filling %	N/M^d	f^e	Excluded entries
1	CH_2	pK	$6 \times 6S$	12	57	2·0	6	—
2	CH_2	$t(CH_2)$	$4 \times 4S$	8	80	2·0	4	—
3	CH_2	$r(CH_2)$	$4 \times 4S$	9	90	2·2	5	—
4	CH_2	$\delta\,^1H$	$10 \times 10S$	43	78	4·3	33	—
5	CH_2	$\delta\,^{13}C$	$9 \times 9S$	28	62	3·1	19	—
6	CO	$\log k$	$11 \times 11S$	43	65	3·9	32	—
7	CO	IP	$8 \times 8S$	20	56	2·5	12	2
7a	CO	IP	$8 \times 8S$	22	61	2·8	14	—
8	CO	$\nu(C{=}O)^f$	$8 \times 13U$	61	59	3·2	41	—
8a	CO	$\nu(C{=}O)^f$	$8 \times 16U$	78	61	3·4	55	—
9	CO	$\nu(C{=}O)^g$	$9 \times 16U$	88	61	3·7	64	3
10	CO	E	$10 \times 10U$	58	58	3·0	39	—
11	NH	pK	$12 \times 12S$	37	47	3·1	25	2
12	$NH_2^{(+)}$	pK	$9 \times 9S$	30	67	3·3	21	2
13	S	IP	$5 \times 5S$	11	73	2·2	6	—
14	SO_2	$\nu(SO_2)$	$11 \times 11S$	39	59	3·5	28	—
15	COO	$\log k$	$5 \times 11U$	54	98	3·6	40	—
15a	COO	$\log k$	$10 \times 11U$	100	91	5·0	81	—

^{a} For a detailed description of the compounds and of the conditions of measurement, see Results; ^{b} S symmetrical matrix, N unsymmetrical matrix; ^{c} number of all entries; ^{d} entries per one empirical constant if calculated according to Eq. (*1*); ^{e} degrees of freedom if calculated according to Eq. (*1*); according to Eq. (*2*) f is by one less; ^{f} in dioxan; ^{g} in tetrachloromethane.

surements were combined. The substituents were only halogens, the data for other substituents being too scarce.

The ^{1}H-NMR chemical shifts of CH_2 protons, measured in tetrachloromethane (series *4*) were taken mainly from two collections[31,32] and were not obtained under strictly identical conditions, in particular the dependence on concentration was rarely followed. The substituents were: H, CH_3, C_6H_5, CH=CH_2, Cl, Br, I, OCH_3, CN, COOH. In the case of ^{13}C-NMR shifts (series *5*) not even the same solvent could be maintained and measurement from several laboratories were combined[19,33-36]. These shifts are, of course, less sensitive to solvent and differ more from each other than ^{1}H shifts. The set of substituents was similar: H, CH_3, C_6H_5, CH=CH_2, Cl, Br, I, OH, COOH.

Logarithms of the second-order rate constants (l mol^{-1} min^{-1}) for the reaction of aliphatic ketones with hydroxylamine in NaOH and pyridine solution at 20°C were all measured in a single laboratory[37] (series *6*). The substituents were: CH_3, C_2H_5, C_3H_7, C_4H_9, C_5H_{11}, C_6H_{13}, $CH(CH_3)_2$, $CH_2CH(CH_3)_2$, $CH(CH_3)C_2H_5$, $CH(CH_3)C_3H_7$, $CH_2CH(CH_3)C_2H_5$. The data are systematic but some substituents are very similar and the corresponding rate constants are equal.

Ionization potentials of aliphatic ketones (series *7*) and sulphides (series *13*) from various sources have been critically collected[38]. The substituents were: H, CH_3, C_2H_5, C_3H_7, C_4H_9 in the case of the sulphides; in case of the ketones in addition $CH(CH_3)_2$, $CH_2CH(CH_3)_2$, and $C(CH_3)_3$. Series *7a* contains two items (isopropyl tert-butyl ketone and di-tert-butyl ketone) which were excluded from series *7*.

The C=O stretching frequencies of carbonyl compounds have been processed in three sets differing in several respects. Series *8* includes derivatives of carboxylic acids (except the carboxylic acids themselves) with the simple as well as more complex substituents: R^1 = CH_3, any higher alkyl C_2—C_5, C_6H_5, $C_6H_4NO_2$-4, $C_6H_4NH_2$-2, 4-C_5H_4N, CH=CHC_6H_5, NH_2; R^2 = OCH_3, OC_2H_5, OC_6H_5, ONHCOR, ON=$C(CH_3)_2$, ON=CHC_6H_5, —O—, NH_2, $N(CH_3)_2$, NHC_6H_5, NHOH, NHOCOR, Cl. All the data have been measured in the same laboratory[10,39,40], in dioxan solution. In the series *8a* the range was broadened by adding aldehydes (R^2 = H) and ketones (R^2 = CH_3, C_6H_5) measured under the same conditions[40]. On the contrary, the measurements in tetrachloromethane (series *9*) have been collected[14] from several sources and aldehydes, ketones and carboxylic acids constitute an essential part. The data are of different reliability although three most deviating were excluded. The substituents were: R^1 = CH_3, C_2H_5, t-C_4H_9, CH_2Cl, $CHCl_2$, CCl_3, CH_2Br, CF_3, $CH_2C_6H_5$, C_6H_5, C_6H_4Cl-4, $C_6H_4OCH_3$-4, $C_6H_4NH_2$-3, $C_6H_4NH_2$-4, $C_6H_4NO_2$-4, CH=CH_2; R^2 = OH, OCH_3, OC_2H_5, NH_2, NHC_6H_5, Cl, H, CH_3, C_6H_5. The integrated carbonyl intensities in tetrachloromethane (series *10*) have been obtained[15] in a similar manner and refer to a similar set of compounds with substituents: R^1 = CH_3, C_2H_5, CH_2Cl, CCl_3, C_6H_5, C_6H_4Cl-3, C_6H_4Br-4, $C_6H_4CH_3$-4, Cl, OC_2H_5; R^2 = OH, OCH_3, OC_2H_5, $OCH_2CH(CH_3)_2$, OC_6H_5, $N(CH_3)_2$, Cl, H, CH_3, C_6H_5.

Dissociation constants of N-acids R^1NHR^2 in water at 25°C (series *11*) were collected[6] from the literature and two deviating less reliable values were excluded (for CH_3CONH_2 and C_6H_5. .$CONH_2$). The statistical corrections were applied. The substituents were: H, CH_3, C_6H_5, C_6H_4Cl-3, $CH_2C_6H_5$, $COCH_3$, COC_6H_5, SO_2CH_3, SO_2CF_3, $SO_2C_6H_5$, $SO_2C_6H_4CH_3$-4, $SO_2C_6H_4NO_2$-4.

Dissociation constants of secondary ammonium ions $R^1NH_2^{(+)}R^2$ in water at 25°C (series *12*) were taken from critical tables[41] and statistically corrected. Although the best values available were always preferred, it was necessary to include also some denoted as "uncertain". The substituents were: H, CH_3, C_2H_5, $CH(CH_3)_2$, cyclo-C_6H_{11}, C_6H_5, $C_6H_4NO_2$-4, $CH_2C_6H_5$, CH_2. .CH_2OH. Excluded was the substituent t-C_4H_9 and the ions $NH_4^{(+)}$ and (4-$NO_2C_6H_4)_2NH_2^{(+)}$ (ref.[42]) although 4-$NO_2C_6H_4NH_2^{(+)}C_6H_5$ was retained[42].

TABLE V

Results of Correlations of Cumulative Substituent Effects in the Reaction Series of Table IV

No	Units	Standard deviations[a]					v_0[b]	Fit[c] ψ	F-test[d]
		s_0	s_1	s_2	s_3	s_4			
1	pK u.	3·15	1·18	0·402	—	—	3·032	0·13	0·025
2	cm^{-1}	84·3	21·4	1·38[e]	—	—	1 274·6	0·016	<0·005
3	cm^{-1}	142	22·4	9·52[e]	—	—	606·9	0·067	0·10
4	p.p.m.	1·19	0·237	0·222	—	0·197	11·8	0·20	—
5	p.p.m.	26·7	6·23	4·26	—	2·85[e]	37·60	0·23	—
6	log u.	0·844	0·155	0·056	—	0·015	3·391	0·018	<0·005
7	eV	0·465	0·072	0·022	—	—	7·52	0·047	<0·005
7a	eV	0·502	0·069	0·050	—	—	6·22	0·14	—
8	cm^{-1}	45·0	3·96	2·70	2·26[f] 2·72[f]	—	1 526	0·060	0·01
8a	cm^{-1}	42·6	4·48	4·06	3·13[f] 4·03[f]	—	1 460	0·11	—
9	cm^{-1}	35·5	3·84	3·58	3·38[f] 3·44[f]	2·95	2 149	0·11	—

10	l mol⁻¹ cm⁻²	366	99·4	84·3	78.1^f / 82.8^f	—	101	0·27	—
11	pK u.	7·18	1·98	*0·30*	—	—	− 5·30	0·042	<0·005
12	pK u.	4·15	0·486	*0·187*	—	—	−12·6	0·045	<0·005
13	eV	0·671	0·097	*0·006*	—	—	6·245	0·009	<0·005
14	cm⁻¹	70·4	12·4	5·23	—	4·85	931	0·074	0·005
15	log u.	0·677	0·144	0·037	0.029^f / 0.035^f	*0·020*	1·79	0·029	0·05
15a	log u.	0·987	0·120	0·084	0.034^f / 0.087^f	0·032	4·75	0·034	<0·005

[a] The standard deviation of the measured values from their average is denoted s_0, the standard deviations between measured and calculated values $s_1 - s_4$, according to the correlation used, Eq. (1)—(4), respectively; the values of s_i are omitted if the number of data was insufficient, or the correlation was forbidden by the symmetry; the most effective, recommended correlation is indicated by *italics*; it is always preferable to all the previous ones according to the F-test and is not surpassed by the next one; [b] constant y_0 in Eq. (2); [c] $\psi_i = s_i/s_0$; the values listed refer to the recommended correlation (*italics*); $\psi < 0.02$ was called[8] "very good", $\psi < 0.1$ "good" fit; [d] the F-test relates to the recommended correlation (*italics*) compared to the preceding model in the sequence $s_1 - s_4$; [e] the requirement of two items of data to one parameter is not fulfilled; [f] the two values correspond to the interchange of rows and columns of the unsymmetrical matrix.

The characteristic vibrations of the SO_2 group in sulphonyl compounds (series *14*) were processed using the arithmetical mean of the symmetric and antisymmetric frequencies; the original data[11] were completed by but few newer measurements[43,44]. The conditions were not uniform and even some Raman data were used together with the prevailing IR measurements; this is necessary owing to the varying physical properties of the compounds investigated. In particular salts of sulphonic acids are hardly comparable to other compounds but they are important for the correlation[11]. The substituents were: CH_3, C_6H_5, $C_6H_4CH_3$-4, $CH_2C_6H_5$, OH, OCH_3, OC_6H_5, $O^{(-)}$, NH_2, F, Cl.

Logarithms of the second-order rate constants (l mol^{-1} min^{-1}) of the alkaline hydrolysis of esters at 15°C were obtained in a single laboratory[45-47] and represent the best set in Table IV as to the filling of the matrix and the number of data related to one parameter. However, the set is rather inhomogeneous as far as the structure is concerned, so that it had to be divided into two subgroups. In the closer assortment (series *15*) the substituents in the acyl moiety were $R^1 = CH_3$. C_2H_5, C_3H_7, i-C_3H_7, t-C_4H_9; in the whole assembly (series *15a*) in addition $R^1 = H$, CH_2OH, CH_2OCH_3, $CHOHCH_3$, $C(OH)(CH_3)_2$. In the alkyl moiety the substituents were the same in the two cases: $R^2 = CH_3$, C_2H_5, C_3H_7, i-C_3H_7, C_4H_9, $CH_2CH\!=\!CH_2$, $(CH_2)_2OCH_3$, $(CH_2)_2OC_4H_9$, $(CH_2)_2OC_4H_9$, $(CH_2)_3OCH_3$, $(CH_2)_3Cl$.

Treatment of data. Each set of data was initially examined by plotting according to Fig. 1. This graphical test allowed firstly to assess the best correlation equation, secondly and more important to exclude strongly deviating points suspected of a gross error. These would not fit any simple model or, in the case of more complex models, would require a special empirical parameter for each entry. According to Table IV only 9 entries were excluded from the whole number of 606, *i.e.* 1·5%, but the excluding was essential to obtain significant results in the respective series.

Subsequently, the correlation equations $(1)-(4)$ were applied in their turn as far as the requirement of two entries to one empirical constant allowed; Eq. (3) can be applied only to an unsymmetrical data matrix. The least-squares calculations by successive approximations are sketched in the Appendix. As the result we obtained the standard deviations $s_1 - s_4$ related to the corresponding models, Eqs $(1)-(4)$, which are listed in Table V, and the empirical parameters b_i, c_i, y_0. Except y_0 which is independent of possible transformations, the constants have not been tabulated. The fit of one series with various models was estimated by comparing the standard deviations, if necessary an approximate F-test was used. The fits of different series can be compared according to the characteristic[8] $\psi = s/s_0$ where s_0 is the standard deviation of observed values from their average.

RESULTS AND DISCUSSION

The results obtained (Table V) allow in principle to discuss the suitability of correlation models, Eqs $(1)-(4)$, on the one hand, and the values of parameters b_i, c_i, y_0 on the other. The latter possibility will be abandoned since it would require to correlate these parameters with other properties or empirical constants of substituents, separately in each case. We feel also that the experimental data are not sufficiently extensive for such an analysis and that further systematic work would be needed.

However, these data are sufficient for deciding between the simpler models $(1)-(4)$ and the result is unambiguous: Eq. (2) is the model of choice for correlating substituent effects of two adjacent substituents. In 11 case of 15 in Table V it is clearly superior to the additive relationship Eq. (1). Simultaneously the overall fit is "satisfactory" according to the characteristic[8] ψ; it is comparable *e.g.* to typical correlations using the Hammett equation[12]. The preference of Eq. (1) in some remaining cases is merely apparent: The series *4*, *5*, and *10* are not correlated by any equation satisfactorily, and in series *7a* and *8a* the differences are leveled due to either including two deviating values, or improper broadening of the validity range, respectively. In the series *9* the fit is improved steadily from Eq. (1) to Eq. (4) but in no step the improvement is significant. In two series, *6* and *15* the best correlation is obtained with Eq. (4), in the series *15a* possibly with Eq. (3). In the literature most quantities of Tables IV, V have been discussed in the framework of additive substi-
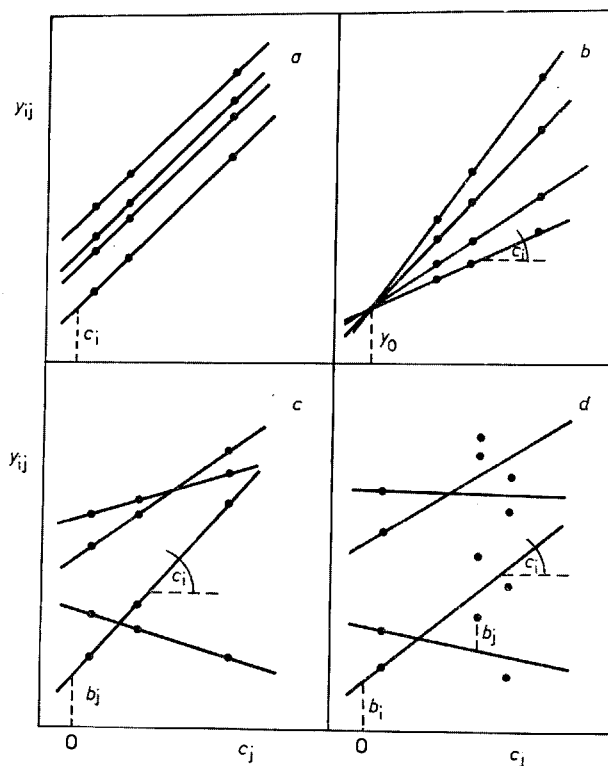

FIG. 1

Graphical Representation of Correlations According to Eq. $(1)-(4)$.
   Experimental values y_{ij} are plotted against the empirical constants c_j.

tuent effects. The resulting correlations and their limited success will be compared with our results and commented separately for each series:

The three series concerning acidities are correlated well by Eq. (2). The measurements on C-acids (series 1) are scarce and probably not very precise; in spite of some deviations the correlation is convincing and worth a graphical representation (Fig. 2). The striking feature is the near point of intersection which seems to be almost experimentally accessible and prompts to further experimental work. No satisfactory correlation of the acidity of C-acids has been presented hitherto. The limited correlations of nitroalkanes[4,48] or diketones[49] with substituent constants of various types assumed the additivity more or less explicitly, unless they were restricted to monosubstitution. Where striking deviations from additivity were observed[50,51], they were discussed qualitatively as exceptions but model of another type has not been taken into account. The only exception was an attempted correlation[52] with σ^* and σ^- constants using a cross term, but due to improper constants the correlation coefficient was as low as 0·69.
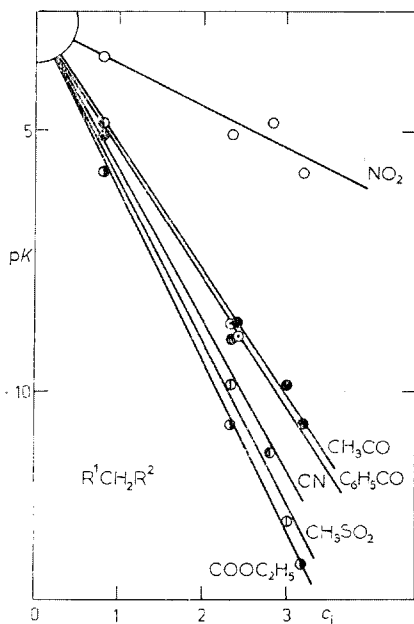

FIG. 2

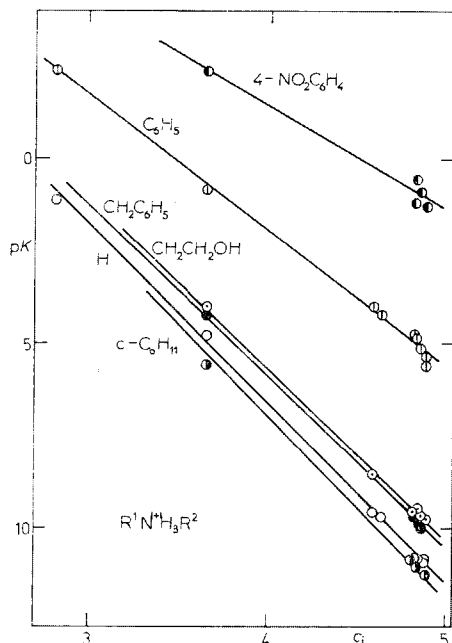Correlation of Dissociation Constants of Di-substituted Methanes According to Eq. (2)

FIG. 3

Correlation of Dissociation Constants of Secondary Ammonium Ions $R^1NH_2^+R^2$ According to Eq. (2)

Only part of the lines is shown.

The acidities of N-acids R^1NHR^2 (series *11*) were dealt with in a separate communication[6]; the non-additive behaviour is beyond any doubt. The acidities of corresponding ammonium ions (series *12*) were correlated several times on broad experimental material[5,20,53], but again additive substituent effects were only considered in spite of perceptible deviations of secondary amines. Our proof is somewhat less convincing since the experimental data are inhomogeneous. Since most substituents influence little the acidity, the success achieved depends strongly on substituents phenyl and 4-nitrophenyl while the important value for 4,4'-dinitrodiphenylamine lies already at pH = −6 and had to be excluded. In addition the point of intersection is relatively apart, see Fig. 3. Further experimental work would be desirable.

The correlations of vibrational frequencies did not yield completely uniform results. Those for twisting (t) and rocking (r) CH_2 frequencies in dihalogenated methanes (series *2* and *3*) are based on extremely restricted data but they were included into Tables IV, V since they are even so quite convincing, see Fig. 4. An extension to additional substituents seems possible. In the original paper a non-linear dependence on the sum of electronegativities was claimed[30] but it clearly arose only from an inefficient mathematical treatment. Interestingly enough the CH_2 deformation frequency[30] (δ) behaves similarly while the wagging frequency[30] (w) is really additive.

The much discussed carbonyl frequency is controlled by substitution in a complex manner. Our previous correlation[10] by Eq. (*2*) was based on our own measurements in dioxan[10,39,40]; the choice of solvent was dictated by the different polarity of compounds to be compared. The correlation[10] holds for derivatives of carboxylic acids as commonly defined; aldehydes and ketones are excluded. More complex derivatives with larger substituents are correlated particularly well. The set of compounds selected accordingly (series *8*) confirmed our previous conclusions[10]; although the value of y_0 is slightly different, $1\,526\ cm^{-1}$ compared to the original value[10] of $1\,514\ cm^{-1}$, the precision is much better than previously[11] for a less exactly defined set. (The values of y_0 are generally very inaccurate since they influence little the overall fit.) If some aldehydes and ketones are added to the set (series *8a*) the accuracy drops and the difference between the additive and non-additive model is obliterated. (The F-test would be significant only at the level $\alpha = 0.25$.) The other set of measurements (series *9*) differs in the solvent but more important in the selection of compounds where aldehydes and ketones are represented by 40%. Under these conditions the statement[14] is justified that the substituent effects are additive. The accuracy is, however, lower even when three most deviating compounds have been excluded. In the literature other additive correlations are found[54−57], which are both more restricted and less precise.

Selection of compounds is also decisive for the correlation of characteristic frequencies of the sulphonyl group (series *14*). The different physical properties of the

compounds investigated prevent measurements under identical conditions. Although the frequencies are not sensitive to solvent or state, the inconsistency of the data becomes evident and even their reproducibility is bad. Under these circumstances the only significant correlation was obtained with the arithmetical mean of the v_s and v_{as} frequencies, in which the errors are reduced[11]. The validity of Eq. (2) seems to be well documented by the data of Table V. The accuracy achieved is similar as in our previous treatment[11] and the value of y_0 differs little (931 cm^{-1} compared to 907 cm^{-1}). However, the result depends in a deciding manner on the presence of the substituent $O^{(-)}$, $i.e.$ on salts of sulphonic acids (see Fig. 1 in ref.[11]), which are measured under disparate conditions and actually do not contain the SO_2 group. Other correlations reported in the literature are based on the additivity principle whether they use the σ^* constants[44,58] or constants derived from the spectral data themselves[16]. They do never include sulphonate anions, more frequently they are even restricted to alkyl substituents, $i.e.$ to aliphatic sulphones[44,58]; the fit is then impaired due to the small value of s_0. Several other frequencies were correlated by the additive scheme with the relative success but special corrections were necessary in the presence of several identical substituents[59,60].

The main problem with IR intensities is the bad agreement when measurements from different laboratories are compared[15]; we believe that most of the scatter is due to this factor. Hence our results (series *10*) cannot decide between the two models,
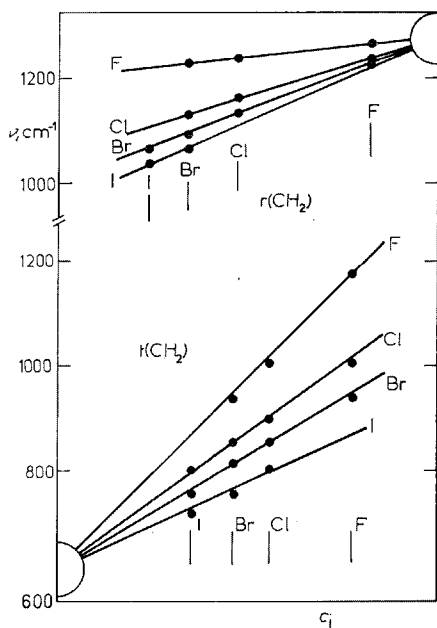

FIG. 4

Correlations of the Rocking and Twisting Vibrational Frequencies of the CH_2 Group According to Eq. (2)

Eq. (1) and (2); according to the F-test Eq. (2) is insignificantly better ($\alpha = 0\cdot25$) and according to the value of ψ none of the correlation is satisfactory by itself. There is another problem which quantity or function should be taken as y in the correlations. We choose simply E (in $1\,mol^{-1}\,cm^{-2}$) but the functions $\log E$ or $E^{1/2}$ were also used[15,61,62] although only the latter has a physical meaning[62]. Our correlation of E according to Eq. (2) with a small value of y_0 ($= 101\,1\,mol^{-1}\,cm^{-2}$) is almost equivalent to the correlation[15] of $\log E$ according to the additive equation (1). The latter model represents in fact a constraint of Eq. (2) by the *a priori* assumption that $y^0 = 0$.

The two correlations of NMR chemical shifts (series 4 and 5) suffer from shortage of strictly comparable data. Even so it seems that substituent effects are more complex in this case and cannot be ad equately represented by one constant for each substituent[27]; this applies particularly to ^{13}C shifts[36]. In fact no satisfactory correlation was obtained with Eqs (1)−(4) and more complex equations would require more systematic data. With the data available (Table IV) Eq. (1) is still the best choice for ^1H shifts although the fit is bad ($\psi = 0\cdot20$). Hence, the original Shoolery's rule[63] is substantiated as a very approximate relation. A more sophisticated approach[9], the "pairwise additivity", may be used only for trisubstituted methanes while for disubstituted it explicitly assumes the non-additive behaviour. Our results for ^{13}C shifts are somewhat more promising but even here the improvement by Eq. (4) does not offset the number of parameters required. In the literature additive relations are reported[19,35,64] but they are valid only for restricted classes of compounds: quite exactly for alkanes[64], well for other hydrocarbons[19]. The principle of pairwise additivity was also claimed[65] in this case.

The ionization potentials were treated in two series of Table IV, 7 and 13. These showed the superiority of Eq. (2) over Eq. (1) in a most convincing manner, although the number of derivatives is small. Nevertheless, in the series of ketones two entries had to be excluded (for the most sterically crowded compounds) and in the series of sulphides the data are inhomogeneous, the substituent R = H being of deciding importance. Hence further data would be still needed. In the literature the dependence on σ^* constants was claimed[38,66] for various classes of compounds including those in Tables IV, V. Even if the correction for the number of hydrogen atoms was made[38], the additive behaviour was always assumed and no other possibility considered[38,66]. In a more realistic approach the non-additivity was revealed in the case of dialkyl ethers, sulphides[13] and amines[67], but expressed formally by different contributions of the first, second, and third alkyl substituent.

The rate constants for the reaction of ketones with hydroxylamine (series 6) obey clearly Eq. (4) although even Eq. (2) yields a sufficiently good fit. The data have, however, a shortcoming that longer alkyls beginning with C_3H_7 have equal effects and the substituent CH_3 is substantial for the correlation. In the original literature[37] no rationalization of the results was attempted. The rate constants for the alkaline

hydrolysis of esters (series *15*) represent the unique case among the data of Tables IV, V. Since the nucleophilic attack proceeds on the carbonyl carbon, the substituent on oxygen (R^2) is actually no more directly adjacent. We included, however, this series as the only example of an intrinsically asymmetrical functional group. The greater distance of the substituent R^2 is reflected in the smaller variance of the constants $b_j (c_j)$ compared with $b_i (c_i)$. The complete set (series *15a*) yields almost the same fit as the selection, series *15*; the lower accuracy is counterbalanced by the greater value of s_0. Eq. (*4*) is the best choice in the series *15*, although even Eq. (*2*) is not much worse. In the series *15a* Eq. (*3*) is most economical, provided that the substituents in the acyl moiety are given two parameters (b_i, c_i).

CONCLUSIONS

The foregoing detailed analysis confirmed the statement that Eq. (*2*) is the most appropriate relationship correlating effects of two adjacent substituents. While it commonly gives a much better fit than the additive equation (*1*), it is virtually not more complex: There is only one parameter more (y_0), *i.e.* only one degree of freedom is lost (Table I); the more difficult calculation is of no relevance nowadays. On the other hand, the effect of more sophisticated equations, *e.g.* (*3*) or (*4*) is always controversial, the better fit being redeemed by the need of much more parameters.

Notwithstanding, in almost all examples from the literature only additive relationships have been used and the deviations discussed as anomalies. One cause could be deficient or improperly selected data; a few gross errors, or several substituents with special effects can rise the standard deviation in such a manner that the differences between the models are not perceptible (see *e.g.* series *7* and *7a*, *8* and *8a*, or the abnormally behaved substituents in ref.[6]). Another cause with the same final effect may be use of fixed empirical constants not corresponding closely to the substituent effects in a given case; for this reason we have preferred a treatment without predefined constants. However, the main reason of preferring additive relationships is in our opinion the mere tradition, most frequently no other model has been even tested.* Hence, the present analysis represents the first more general proof of non-additive effects of two substituents. Further development is possible by introducing a more sophisticated relationship on the one hand, or by extending to more substituents on the other hand. In both cases more comprehensive and more systematic experimental data will be needed than available for the present.

---

* "Wie denn von menschlichen Geiste keine andere Form, unter welcher die gemeinschaftliche Wirkung verschiedener Elemente stattfinden kann, vorstellbar ist, als die der Summe", W. Ostwald[68] (compare Hückel[69]: "On the Additive Thinking in Structural Chemistry"). However, the properties of the non-additive relationship Eq. (*2*) were anticipated in a qualitative fashion by Skrabal[70] as early as 1916, and discussed later[47].

APPENDIX

The least-squares calculations according to Eq. $(1)-(4)$ were done by successive approximations and programmed for the Hewlett-Packard calculator 9820 A. The simplest programs were used and the convergence was only tested on practical examples instead of mathematically proven. Two variants of each program were written: for an unsymmetrical and symmetrical data matrix. In the latter case only the entries above the main diagonal are given in the input and completed to a square matrix by transferring into symmetrical positions. Then the matrix is treated as an unsymmetrical one, the diagonal elements being given the twofold weight. The standard deviation is calculated from all doubled data with the doubled degrees of freedom actually present in the original symmetrical matrix (Table I). If a symmetrical matrix is completed to the square form and processed according to the program for the unsymmetrical matrices, an error arises which is, however, not important in practical examples.

In the input of the programs one gives the matrix dimensions $(p \times q)$ and the experimental data (y_{ij}). In an incomplete matrix the empty entries are designated by the value $y_{ij} = 0$, hence any actual value of y_{ij} must not equal zero exactly. In the output one obtains: Number of all data (N), degrees of freedom (f), standard deviation s_0, and a sequence of standard deviations reached by successive approximations (s_i). The approximations are stopped by the operator usually when s_i is constant to five significant figures since the values of the parameters (b_i, c_i, y_0) are more sensitive than s_i. Finally these parameters are printed after being normalized as recommended in Table I. All programs were tested on synthetic data obeying exactly the pertinent model and also on those corresponding to all simpler models.

Equation (1). In case of a full data matrix an algebraic solution is possible, see Eq. (10), (11). For an incomplete matrix the solution is obtained by iteration: After the s-th iteration the new values of $c_i^{(s+1)}$ are obtained from the old values of $c_j^{(s)}$ with respect to Eq. (9):

$$c_i^{(s+1)} \sum_j w_{ij} = \sum_j w_{ij} y_{ij} - \sum_j w_{ij} c_j^{(s)} .$$

Subsequently $c_j^{(s+1)}$ are obtained as

$$c_j^{(s+1)} \sum_i w_{ij} = \sum_i w_{ij} y_{ij} - \sum_i w_{ij} c_i^{(s+1)} .$$

The iterations converge rapidly and the starting values are immaterial (*e.g.* $c_j^{(1)} = 0$). In most of the examples of Table V a reasonable value of s_1 was obtained after 5 cycles and sufficiently precise values of c_i and c_j after 10 cycles. The more filled the matrix the faster proceeds the iteration; this holds even for the further models Eqs $(2)-(4)$. If a complete matrix is processed using the iterative program, the exact values are obtained in the first cycle, of course.

Equation (2). This model is no more linear and its main difficulty lies in the estimation of y_0. The following computational procedure was used by us but is not completely satisfactory: If y_0 is given an approximate value $(y_0^{(r)})$, the remaining constants c_i, c_j are obtained by the iterative solution of the normal equations (12):

$$c_i^{(s+1)} \sum_j w_{ij} c_j^{(s)} c_j^{(s)} = \sum_j w_{ij} y_{ij} c_j^{(s)} - y_0^{(r)} \sum_j w_{ij} c_j^{(s)} ,$$

$$c_j^{(s+1)} \sum_i w_{ij} c_i^{(s+1)} c_i^{(s+1)} = \sum_i w_{ij} y_{ij} c_i^{(s+1)} - y_0^{(r)} \sum_i w_{ij} c_i^{(s+1)} .$$

The new value of $y_0^{(r+1)}$ may be either given by the operator or computed from the last values

$c_j^{(s)}$ by the solution of equations (*12*). (The first equation is solved according to c_i, multiplied by $w_{ij}c_j$, the sum $\sum w_{ij}c_ic_j$ is computed and substituted into the third equation.)

$$y_0^{(r+1)}\left[\sum_i \frac{(\sum_j w_{ij}c_j^{(s)})^2}{\sum_j w_{ij}c_j^{(s)}c_j^{(s)}} - N\right] = \sum_i \frac{\sum_j w_{ij}c_j^{(s)} \sum_j w_{ij}c_j^{(s)}y_{ij}}{\sum_j w_{ij}c_j^{(s)}c_j^{(s)}} - \sum_{ij} w_{ij}y_{ij}.$$

The iterations converge well for synthetic examples but less satisfactorily in cases where Eq. (*2*) actually does not hold. In particular if the computed value of y_0 is very apart from the experimental values, it is both determined with a low precision and the convergence is slow; sometimes even y_0 changes so irregularly that one value is reached twice. In such cases it is important to start with a reasonable estimate of y_0 and to fasten the convergence by substituting a new y_0 value whenever needed. A given value of y_0 is always to be kept till a certain degree of convergence of c_i, c_j is reached. The examples of Table V required in the average 20−50 different values of y_0 with usually 4 cycles of computing c_i, c_j for each y_0. As a rule s_2 is obtained much more reliably and more quickly than y_0.

Equation (3). This case was solved[22] on the basis of the principal component model. Our program was only slightly different. From the s-th values $c_j^{(s)}$ new values $c_i^{(s+1)}$ and $b_i^{(s+1)}$ are obtained by linear regressions in the coordinates $c_j^{(s)}$ and y_{ij} (one regression for each i):

$$y_{ij} = b_i^{(s+1)} + c_i^{(s+1)}c_j^{(s)}.$$

Hence

$$c_i^{(s+1)} = (\sum_j w_{ij}\sum_j w_{ij}c_j^{(s)}y_{ij} - \sum_j w_{ij}c_j^{(s)}\sum_j w_{ij}y_{ij})/[\sum_j w_{ij}\sum_j w_{ij}c_j^{(s)}c_j^{(s)} - (\sum_j w_{ij}c_j^{(s)})^2],$$

$$b_i^{(s+1)} = (\sum_j w_{ij}c_j^{(s)}c_j^{(s)}\sum_j w_{ij}y_{ij} - \sum_j w_{ij}c_j^{(s)}\sum_j w_{ij}c_j^{(s)}y_{ij}) / [\sum_j w_{ij}\sum_j w_{ij}c_j^{(s)}c_j^{(s)} - (\sum_j w_{ij}c_j^{(s)})^2].$$

New values of $c_j^{(s+1)}$ are obtained by linear regression with the fixed origin in the coordinates $c_i^{(s+1)}$ and $y_{ij} - b_i^{(s+1)}$ (one for each j):

$$c_j^{(s+1)} = [\sum_i w_{ij}c_i^{(s+1)}(y_{ij} - b_i^{(s+1)})]/\sum_i w_{ij}c_i^{(s+1)}c_i^{(s+1)}.$$

The iterations converge very quickly; starting with the values $c_i^{(1)} = 1$ they were mostly finished in 10 cycles. For this model it has been proven[71] that the iterations yield the absolute minimum. Hence this applies also to the preceding models Eqs (*1*)−(*2*).

Equation (4). The problem is rather similar to the preceding one; due to the symmetry of the model equal regressions are made in both steps. The pertinent regression equations are

$$y_{ij} = b_i^{(s+1)} + c_i^{(s+1)}c_j^{(s)}$$
$$y_{ij} = b_j^{(s+1)} + c_j^{(s+1)}c_i^{(s+1)}.$$

The starting values $c_j^{(1)}$ must not be equal, one can choose *e.g.* $c_j^{(1)} = j$. For this reason the convergence depends on the sequence of rows and columns which should be reasonably arranged as far as possible. Usually the iterations converge satisfactorily although the difference between

two successive values does not always diminish quite regularly. The sets of Table V required usually $20-30$ cycles to determine s_4 to five significant figures. For this model there is no proof that the absolute minimum has been actually found.

REFERENCES

1. Miller S. I.: J. Amer. Chem. Soc. *81*, 101 (1959).
2. Exner O. in the book: *Advances in Linear Free Energy Relationships* (N. B. Chapman, J. Shorter, Eds), p. 41. Plenum Press, London 1972.
3. Exner O.: This Journal *25*, 1044 (1960).
4. Talvik A. J., Palm V. A.: Org. Reactivity *11*, 287 (1974).
5. Condon F. E.: J. Amer. Chem. Soc. *87*, 4485 (1965).
6. Exner O., Janák P.: This Journal *40*, 2510 (1975).
7. Hammond P. R.: J. Chem. Soc. *1964*, 471.
8. Exner O.: This Journal *31*, 3222 (1966).
9. Vladimiroff T., Malinowski E. R.: J. Phys. Chem. *46*, 1830 (1967).
10. Exner O., Horák M.: This Journal *24*, 968 (1959).
11. Exner O.: This Journal *28*, 935 (1963).
12. Kalfus K., Kroupa J., Večeřa M., Exner O.: This Journal *40*, 3009 (1975).
13. Kaufman J. J.: J. Phys. Chem. *66*, 2269 (1962).
14. Seth-Paul W. A., Van Duyse A.: Spectrochim. Acta *28A*, 211 (1972).
15. Seth-Paul W. A.: Spectrochim. Acta *30A*, 1817 (1974).
16. Butcher F. K., Charalambous J., Frazer M. J., Gerrard W.: Spectrochim. Acta *23A*, 2399 (1967).
17. Pascual C., Meier J., Simon W.: Helv. Chim. Acta *49*, 164 (1966).
18. Shorter J., Stubbs F. J.: J. Chem. Soc. *1949*, 1180.
19. Savitsky G. B., Namikawa K.: J. Phys. Chem. *68*, 1956 (1964).
20. Chremos G. N., Zimmerman H. K.: Chimia *18*, 265 (1964).
21. Wold S.: Chem. Scripta *5*, 97 (1974).
22. Wold S., Sjöström M.: Chem. Scripta *2*, 49 (1972).
23. Ehrenson S., Brownlee R. T. C., Taft R. W.: Progr. Phys. Org. Chem. *10*, 1 (1973).
24. Ehrenson S.: Tetrahedron Lett. *1964*, 351.
25. Swain C. G., Dittmer D. C., Kaiser L. A.: J. Amer. Chem. Soc. *77*, 3737 (1955).
26. Swain C. G., Mosely R. B., Bown D. E.: J. Amer. Chem. Soc. *77*, 3731 (1955).
27. Weiner P. H., Malinowski E. R., Levinstone A. R.: J. Phys. Chem. *74*, 4537 (1970).
28. Wold S., Andersson K.: J. Chromatogr. *80*, 43 (1973).
29. Ebel H. F.: *Die Acidität der CH-Säuren.* Thieme, Stuttgart 1969.
30. Wilmshurst J. K.: Can. J. Chem. *35*, 937 (1957).
31. Brügel W.: *Kernresonanz-Spektrum und Chemische Konstitution.* Steinkopff, Darmstadt 1967.
32. *The Sadtler Standard Spectra.* Sadtler Research Laboratories, Philadelphia $1966-1973$.
33. Spiesecke H., Schneider W. G.: J. Chem. Phys. *35*, 722 (1961).
34. Savitsky G. B., Namikawa K.: J. Phys. Chem. *67*, 2430 (1963);
35. Savitsky G. B., Pearson R. M., Namikawa K.: J. Phys. Chem. *69*, 1425 (1965).
36. Zetta L., Gatti G.: Org. Magn. Resonance *4*, 585 (1972).
37. Kletzke P. G.: J. Org. Chem. *29*, 1363 (1964).
38. Pôldoja P. K., Palm V. A.: Reakts. Sposobnost Org. Soedin. *4*, 786 (1967).

39. Exner O., Horák M.: This Journal *24*, 2992 (1959).
40. Tabačík V.: *Thesis*. Czechoslovak Academy of Sciences, Prague 1963.
41. Perrin D. D.: *Dissociation Constants of Organic Bases in Aqueous Solution*. Butterworths, London 1965; Supplement 1972.
42. Dolman D., Stewart R.: Can. J. Chem. *45*, 903 (1967).
43. Ghersetti S.: Boll. Sci. Fac. Chim. Ind. Bologna *21*, 237 (1963).
44. Doerffel K., Brunn J.: J. Prakt. Chem. *312*, 701 (1970).
45. Salmi E. J., Leino E.: Suom. Kemistilehti *17B*, 19 (1944).
46. Leimu R., Korte R., Laaksonen E., Lehmuskoski U.: Suom. Kemistilehti *19B*, 93 (1946).
47. Salmi E. J., Leimu R.: Suom. Kemistilehti *20B*, 43 (1947).
48. Talvik A. I.: Reakts. Sposobnost Org. Soedin. *9*, 233 (1972).
49. Calmon J.-P., Maroni P.: Bull. Soc. Chim. Fr. *1965*, 2532.
50. Taft R. W.: J. Amer. Chem. Soc. *79*, 5075 (1957).
51. Adolph H. G., Kamlet M. J.: J. Amer. Chem. Soc. *88*, 4761 (1966).
52. Talvik A. I.: Reakts. Sposobnost Org. Soedin. *2*, No. 1, 35 (1965).
53. Clark J., Perrin D. D.: Quart. Rev. Chem. Soc. *18*, 295 (1964).
54. Kagarise R. E.: J. Amer. Chem. Soc. *77*, 1371 (1955).
55. Morgan K. J., Unwin N.: J. Chem. Soc. (B) *1967*, 1336; *1968*, 880.
56. de Roos A. M.: Rec. Trav. Chim. Pays-Bas *87*, 1359 (1958).
57. Adelman R. L.: J. Org. Chem. *29*, 1837 (1964).
58. Ghersetti S., Zauli C.: Ann. Chim. (Rome) *53*, 710 (1963).
59. Smith A. L., Angelotti N. C.: Spectrochim. Acta *15*, 412 (1959).
60. Thomas L. C., Chittenden R. A.: Spectrochim. Acta *20*, 467 (1964).
61. Rao C. N. R.: Chem. Ind. (London) *1958*, 891.
62. Brown T. L.: J. Phys. Chem. *64*, 1798 (1960).
63. Jackman L. M., Sternhell S.: *Applications of Nuclear Magnetic Resonance Spectroscopy in Organic Chemistry*, 2nd Ed., p. 181. Pergamon Press, Oxford 1969.
64. Paul E. G., Grant D. M.: J. Amer. Chem. Soc. *85*, 1701 (1963).
65. Malinowski E. R., Vladimiroff T., Tavares R. F.: J. Phys. Chem. *70*, 2046 (1966).
66. Levitt L. S., Levitt B. W.: Israel J. Chem. *9*, 711 (1971).
67. Kaufman J. J., Koski W. S.: J. Amer. Chem. Soc. *82*, 3262 (1960).
68. Ostwald W.: Z. Physik. Chem. *3*, 417 (1889).
69. Hückel W.: J. Prakt. Chem. *277*, 122 (1957).
70. Skrabal A.: Monatsh. Chem. *37*, 495 (1916).
71. Lyttkens E. in the book: *Multivariate Analysis* (P. Krishnaiah, Ed.). Academic Press, New York 1966.

Translated by the author.