

ADDITIVE-MULTIPLICATIVE DESCRIPTION OF EXTRATHERMO-DYNAMIC RELATIONS: APPLICATION TO SOLVENT EFFECT

Oldřich PYTELA

*Department of Organic Chemistry,
Institute of Chemical Technology, 532 10 Pardubice*

Received May 12, 1989

Accepted June 5, 1989

The paper is focused on evaluation of significance of the additive-multiplicative model of extrathermodynamic relations (linear free energy relationships) as compared with the additive model. Application of the method of conjugated deviations to a data matrix describing manifestations of solvent effects in 367 processes in solutions (6 334 data) has shown that introduction of cross-terms into the additive model is statistically significant for a model with two and particularly three parameters. At the same time the calculation has provided a new set of statistical parameters for description of solvent effect with application of the additive-multiplicative model. Compared with an analogous set designated for the additive model, the new parameters show a lower mutual correlation, retaining the same nature of the properties described, i.e. polarity-acidity (PAC parameter), polarity-basicity (PBC), and polarity-polarizability (PPC).

Extrathermodynamic relations belong among the most important tools for quantitative description of dependences in organic chemistry¹⁻¹⁰. This is true in spite of the fact that these relations do not represent a fundamental law¹¹ but, primarily, a mathematical expression of similarities of effects of changes (perturbances) related to a standard state. The extent of similarity is different, which is indicated by comparisons of substituent effects^{2-8,10-15}, solvent effects^{7-10,16-19}, nucleophilicity^{6-8,10,20}, and/or their interdependences^{6-8,10}. This similarity principle is well expressed by the approximative description by the Taylor expansion^{1,6,18} (1)

$$\Delta G = \Delta G_0 + \sum_{i=1}^p \left[\frac{\partial \Delta G}{\partial P_i} \right]_{\Delta G = \Delta G_0} \Delta P_i + \sum_{i=1}^p \sum_{j=1}^p \left[\frac{\partial^2 \Delta G}{\partial P_i \partial P_j} \right]_{\Delta G = \Delta G_0} \Delta P_i \Delta P_j + \dots + R_{p+1}, \quad (1)$$

where ΔG symbolizes the physical quantity measured (which most often is related to the Gibbs energy), ΔG_0 is the same quantity in the standard state, ΔP symbolizes – in parametrical way – the extent of perturbation related to the standard state, and R_{p+1} means the error of approximation. As the expressions given in square brackets are constants, the closeness of description within a certain field is given

by the choice of the standard state and parametrization, i.e. adjustment of the ΔP values. Since the difficulties with parametrization rapidly increase with increasing number of parameters and, at the same time, the demands on number of experiments in an application increase too, the description of a dependence is almost always restricted to the first two terms in Eq. (1) with the number of parameters most often varying from one to four. Nevertheless, the neglecting of the interaction terms expressed by the third term of Eq. (1) obviously does not correspond to reality, because various manifestations of a phenomenon followed (e.g. solvent effect) are interconnected and affect each other^{18,21}. A mathematical expression of this interaction will lead not only to a better description of the dependence but also to a better interpretation of reality in terms of the similarity relation. The product form proved to be the best e.g. in the description of effects of two substituents²¹. On the other hand, the application of quadratic terms¹¹ does not bring any new effects but rather indicates an unsuitable parametrization of the linear terms or exceeding of the validity range. In the sense of these considerations it is possible to derive (from Eq. (1)) the simplest relation including the cross-terms in the form of Eq. (2)

$$\Delta G = a_0 + \sum_{i=1}^p a_i \Delta P_i + \sum_{i=1}^p \sum_{j>i}^p a_{ij} \Delta P_i \Delta P_j + \varepsilon, \quad (2)$$

where a are the regression parameters and ε is the residual error given by the approximation used and by the experimental error of ΔG quantity.

One of the parametrization possibilities of various models of extrathermodynamic relations consists in the statistical analysis of an extensive representative set of experimental data^{13,14,17,19,20}. In this respect a large contribution was brought by the methods of analysis of latent variables (PCA (refs^{22,23}), FA (ref.²³), PLS (ref.²⁴), etc.) which invaded chemistry in recent years and whose development was largely initiated by chemistry. These methods are suitable for studies of parametrization of linear models, of course, without interaction terms. A more versatile method is that of conjugated deviations used earlier for parametrizations of nucleophilicity²⁰, solvent effects¹⁹, as well as for analysis of chemical data²⁵. This method enables the optimization of values of the independent variables ΔP in Eq. (2) with simultaneous minimization of differences between the calculated and experimental ΔG values in a set of equations type (2). This procedure has the advantage of being little sensitive to missing data and of respecting the statistical significance of the regression coefficients. The parameters ΔP obtained show a good application ability in the definition models^{19,20}.

The aim of this present communication is to evaluate the additive-multiplicative description of dependences in organic chemistry, to verify the applicability of the method of conjugated deviations to parametrizations of these relations, and, using a data set from the field of solvent effects, to evaluate the differences in physico-

-chemical interpretation of the results obtained from the additive and from the additive-multiplicative models.

CALCULATIONS

The matrix selected for the analysis has already been used earlier in studies of solvent effects^{18,19}; the columns containing small numbers of data were excluded (Nos 10, 49, 219–225, 266, 286 according to ref.¹⁸). The calculations were carried out for the model without interactions with one, two, and three parameters, and for model (2) with two and three parameters. For the purposes of optimization of the parameters in model (2), the described algorithm^{19,20,25} was modified for calculation of regressions with involvement of the interaction terms. The calculations were carried out according to our own program in the FORTRAN language using an EC 1033 computer.

RESULTS AND DISCUSSION

The statistical significance of the difference between the additive and additive-multiplicative models can be evaluated on the basis of the residual deviations minimized in the method of conjugated deviations²⁵. Mathematically, the residual deviations s are defined by Eq. (3)

$$s = \left[\sum_{i=1}^n \sum_{j=1}^m e_{ij}^2 / \left(\sum_{j=1}^m N_j - np - \sum_{j=1}^m \bar{P}_j \right) \right]^{1/2}, \quad (3)$$

where n and m mean the numbers of rows and columns, respectively, of the data matrix, e_{ij} are differences between the standardized experimental and calculated variables, N_j is the number of values in the j -th column of the data matrix (which need not be complete), and \bar{P}_j represents the number of statistically significant regression parameters (according to the t -test) in the regression with the data of the j -th column. Obviously, the denominator of the expression (3) represents the respective number of degrees of freedom. The residual standard deviations obtained together with other characteristics are given in Table I for both the models tested. The results show a peculiarity in that the number of degrees of freedom increases with the number of parameters involved in the additive model. This is a consequence of a greater increase of the interpretation ability of a model with a greater number of parameters as compared with the number of parameters (the number of combinations increases faster). At the same time this fact indicates a small number of parameters in the linear models for interpretation of the given experimental data or a possibility of cross-interactions in the sense of the additive-multiplicative model. This statement is supported by the data of Table I, since the number of the series interpreted by the additive-multiplicative model represented by the sum of the points

N is higher, increases more slowly, and approaches more closely the total number of points (6 334) in the data matrix.

The magnitude of residual standard deviations from the calculations according to the models compared can be compared by the Fisher–Snedecor test (4)

$$F(v_A, v_{AM}) = (s_A^2/v_A)/(s_{AM}^2/v_{AM}), \quad (4)$$

whose values are given in Table I. A significant decrease of residual standard deviation was observed at the significance level $\alpha = 0.05$ for both the model with two independent parameters ($F_{0.95} = 1.048$) and particularly that with three independent parameters ($F_{0.95} = 1.048$). From another standpoint, when rejecting the agreement of the two models for $p = 2$ we make an error of 0.43%, whereas for $p = 3$ the error even will be less than $10^{-10}\%$.

The introduction of cross-terms can also be understood as an extension of a model containing simple independent variables only. Then the contribution of the extension term can be evaluated in relation to the unexplained variability according to the formula²⁵

$$F(v_A - v_{AM}, v_{AM}) = \frac{S_A - S_{AM}}{S_{AM}} \frac{v_{AM}}{v_A - v_{AM}}, \quad (5)$$

where S are the residual sums of squares (the numerator in the expression (3)) and v are the respective degrees of freedom. According to the relation (5) the contribution of cross-terms is univocally statistically significant for both the model with two ($F = 11.79$, $F_{0.95} = 1.40$) and three independent parameters ($F = 5.91$, $F_{0.95} = 1.14$).

Beside the statistical tests, the significance of the additive-multiplicative model (as compared with the additive model) can be demonstrated on the number of statistically

TABLE I

Residual standard deviations s , degrees of freedom v , sums of numbers of points in statistically significant series N , and values of the criterion (4), obtained by the calculation according to the additive-multiplicative models with p independent parameters

p	Additive model			Additive-multiplicative model			F
	s	v	N	s	v	N	
1	0.590	4 483	5 104	—	—	—	—
2	0.491	4 919	5 861	0.471	4 880	5 924	1.078
3	0.412	5 014	6 191	0.359	4 688	6 227	1.231

significant (according to the *t*-test) regression coefficients. Out of the total number of regressions, the regression coefficients for simple independent variables were significant in 53.5%, the regression coefficients for the cross-terms in 39.2% of cases. Hence it can be stated that the additive-multiplicative model has its significance in description of extrathermodynamic relations.

An application of a data set oriented to solvent effects makes it possible to evaluate the differences in parametrization of the additive and additive-multiplicative models (Table II, Figs 1, 2). Especially from the graphical representation it is obvious that the general pattern of distribution of solvents within the parameters space has not substantially changed. Hence the physico-chemical interpretation of the parameters has not changed either: the PAC parameter describes the solvent polarity and acidity, PBC refers to solvent polarity and basicity, and PPC to solvent polarity and polarizability¹⁹. It is important that even when the cross-terms are applied, the solvent polarity is not manifested as an independent factor but operates always in connection with other fundamental properties^{9,18} such as acidity, basicity, or polarizability. This phenomenon has already been declared earlier for the solvent acidity and basicity¹⁸ and was observed with all the properties mentioned for the additive model¹⁹. As the solvent polarity is not manifested as an independent factor in most processes in solution, it should not be used independently for interpretation and prediction of solvent effect, and other properties of solvent should also be respected. At the same time, on the basis of the observation mentioned it can be stated that by chemometrical means inaccessible are the internal fundamental parameters of solvents

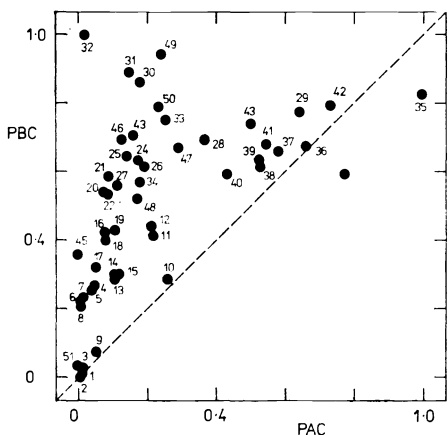


FIG. 1

Dislocation of solvents in the plane determined by the parameters PAC and PBC

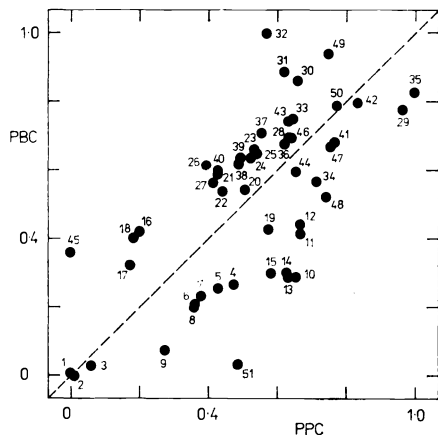


FIG. 2

Dislocation of solvents in the plane determined by the parameters PPC and PBC

TABLE II

The statistical parameters PAC, PBC, and PPC for description of solvent effect with involved cross-terms, their standard deviations s , and numbers of statistically significant corrections M used to adjust the individual parameters (for mathematical definitions of s and M see ref.¹⁹)

No.	Solvent	PAC s	M	PBC s	M	PPC s	M
1	Hexane	0.011 0.061	93	0.006 0.073	102	0.001 0.066	114
2	Heptane	0.007 0.061	43	0.000 0.089	42	0.014 0.059	49
3	Cyclohexane	0.014 0.073	105	0.028 0.074	139	0.060 0.080	135
4	Benzene	0.046 0.075	123	0.267 0.116	128	0.470 0.072	158
5	Toluene	0.041 0.085	79	0.252 0.047	79	0.426 0.069	98
6	<i>m</i> -Xylene	0.005 0.127	16	0.207 0.105	23	0.363 0.103	22
7	<i>p</i> -Xylene	0.012 0.071	32	0.231 0.066	45	0.377 0.086	51
8	Mesitylene	0.001 0.084	29	0.201 0.097	33	0.358 0.066	36
9	Tetrachloromethane	0.055 0.117	117	0.074 0.151	132	0.276 0.068	142
10	Chloroform	0.259 0.116	101	0.289 0.095	107	0.652 0.121	116
11	Dichloromethane	0.217 0.104	109	0.414 0.117	113	0.664 0.149	125
12	1,2-Dichloroethane	0.210 0.093	87	0.442 0.068	91	0.666 0.067	108
13	Chlorobenzene	0.106 0.064	78	0.287 0.068	72	0.633 0.281	89
14	Bromobenzene	0.104 0.106	51	0.297 0.090	58	0.628 0.113	64
15	Fluorobenzene	0.121 0.083	16	0.303 0.121	23	0.582 0.071	24
16	Diethyl ether	0.077 0.101	109	0.423 0.064	113	0.203 0.098	139

TABLE II
(Continued)

No.	Solvent	PAC <i>s</i>	<i>M</i>	PBC <i>s</i>	<i>M</i>	PPC <i>s</i>	<i>M</i>
17	Dibutyl ether	0.052 0.097	47	0.319 0.086	55	0.170 0.137	58
18	Diisopropyl ether	0.080 0.064	19	0.399 0.060	17	0.181 0.101	26
19	Anisol	0.105 0.095	37	0.427 0.090	43	0.575 0.089	43
20	Fenethol	0.076 0.122	14	0.539 0.250	16	0.508 0.222	20
21	Tetrahydrofuran	0.085 0.074	97	0.587 0.095	109	0.427 0.105	128
22	Dioxane	0.086 0.115	119	0.532 0.152	120	0.440 0.109	148
23	Acetone	0.159 0.141	128	0.707 0.108	124	0.551 0.125	150
24	Butanone	0.172 0.106	47	0.630 0.163	51	0.518 0.091	62
25	Cyclohexane	0.140 0.089	34	0.647 0.070	30	0.540 0.074	45
26	Methyl acetate	0.191 0.098	36	0.610 0.148	40	0.394 0.105	46
27	Ethyl acetate	0.112 0.105	106	0.562 0.093	113	0.411 0.102	127
28	Acetanhydride	0.364 0.271	19	0.692 0.121	17	0.632 0.162	21
29	Formamide	0.643 0.157	40	0.773 0.223	35	0.963 0.171	42
30	N,N-Dimethylformamide	0.177 0.068	115	0.859 0.085	121	0.658 0.079	140
31	N,N-Dimethylacetamide	0.145 0.136	64	0.889 0.110	57	0.620 0.088	67
32	Hexamethylphosphorus triamide	0.018 0.196	50	1.000 0.165	45	0.565 0.240	58
33	Acetonitrile	0.251 0.188	131	0.745 0.155	133	0.644 0.152	172

TABLE II
(Continued)

No.	Solvent	PAC <i>s</i>	<i>M</i>	PBC <i>s</i>	<i>M</i>	PPC <i>s</i>	<i>M</i>
34	Benzonitrile	0.181 0.163	43	0.569 0.084	45	0.710 0.174	59
35	Water	1.000 0.086	71	0.826 0.112	60	1.000 0.101	82
36	Methanol	0.663 0.130	126	0.672 0.127	114	0.623 0.104	147
37	Ethanol	0.582 0.097	116	0.660 0.071	98	0.533 0.097	182
38	1-Butanol	0.534 0.092	84	0.619 0.092	68	0.486 0.151	88
39	2-Propanol	0.527 0.080	99	0.633 0.082	80	0.493 0.099	106
40	2-Methyl-2-propanol	0.430 0.135	66	0.595 0.192	55	0.430 0.127	71
41	Benzyl alcohol	0.549 0.138	47	0.681 0.126	39	0.766 0.194	54
42	1,2-Ethanediol	0.733 0.117	57	0.791 0.099	51	0.831 0.122	66
43	2-Methoxyethanol	0.501 0.135	23	0.737 0.088	17	0.635 0.158	27
44	Acetic acid	0.774 0.181	46	0.591 0.154	39	0.654 0.109	42
45	Triethylamine	0.002 0.156	43	0.359 0.250	57	0.000 0.180	62
46	Pyridine	0.125 0.162	72	0.694 0.272	80	0.638 0.191	9
47	Nitromethane	0.295 0.159	72	0.668 0.244	81	0.753 0.131	95
48	Nitrobenzene	0.171 0.092	57	0.522 0.079	55	0.737 0.103	76
49	Dimethyl sulfoxide	0.243 0.201	115	0.941 0.150	120	0.749 0.060	135
50	Sulfolane	0.231 0.108	18	0.787 0.167	19	0.773 0.094	26
51	Carbon disulfide	0.000 0.127	45	0.163 0.036	55	0.489 0.159	58

(denoted as π in the previous paper¹⁹) whose combination represents the chemometrically determined empirical parameters in the respective models. Nevertheless it is meaningful to consider these fundamental parameters, because they can be defined by theoretical means and thus form a connection between the exactly theoretically defined quantities⁹ and experimental observation.

The correlation between the parameters observed in the additive model¹⁹ has decreased markedly by introduction of the cross-terms, viz. the dependence between the parameters of the type PA and PB from $r = 0.551$ to 0.306 , that of the type PA and PP from $r = 0.549$ to $r = 0.254$, and that of the type PB and PP from $r = 0.805$ to $r = 0.548$. The lowering of correlation is particularly significant for the parameters describing polarity-basicity (PB) and those describing polarity-polarizability (PP) represented in Fig. 2, since it makes more distinct the classification of solvents with predominant basicity or, on the contrary, predominant polarizability. Both figures show that it is impossible to mechanically transfer the ideas about acidity (acetic acid) or basicity (triethylamine) usually obtained from measurements in water to the properties of these molecules simultaneously acting as solvent.

When applying the parameters of Table II to the definition set we have found — for the PAC parameters — the absolute average regression coefficient of 11.3 , those for the PBC and PPC parameters being 4.8 and 5.1 , respectively. As the parameters are standardized in the interval $\langle 0, 1 \rangle$, the values given represent a measure of sensitivity of processes in solvents to the individual properties of solvent. In accordance with ref.¹⁸ this sensitivity is the highest for acidity in combination with polarity, of course in processes in which acidity is manifested (about 50% of cases according to the statistical significance of regression parameter).

REFERENCES

1. Wold S.: Chem. Scr. 5, 97 (1974).
2. Wold S., Sjöström M.: Chem. Scr. 2, 49 (1972).
3. Sjöström M., Wold S.: Chem. Scr. 6, 114 (1974).
4. Sjöström M., Wold S.: Acta Chem. Scand., B 30, 167 (1976).
5. Sjöström M., Wold S.: Chem. Scr. 9, 200 (1976).
6. Palm V. A.: *Osnovy kolichestvennoi teorii organicheskikh reaktsii*. Khimiya, Leningrad 1977.
7. Shorter J., Chapman N. B. (Eds): *Correlation Analysis in Chemistry: Recent Advances*. Plenum Press, New York 1978.
8. Shorter J.: *Correlation Analysis of Organic Reactivity*. Wiley, New York 1982.
9. Reichardt C.: *Solvent Effects in Organic Chemistry*, 2nd ed. Verlag Chemie, Weinheim 1988.
10. Exner O.: *Correlation Analysis of Chemical Data*. Plenum, New York; SNTL, Prague 1988.
11. Sjöström M., Wold S.: Acta Chem. Scand., B 35, 537 (1981).
12. Kramer C.-R.: Z. Phys. Chem. (Leipzig) 261, 745 (1980).
13. Swain G. G., Unger S. H., Rosenquist N. R., Swain M. S.: J. Am. Chem. Soc. 105, 492 (1983).
14. Spajner M. C., DeLigny C. L.: J. Chem. Res. (M) 1986, 1701.

15. Exner O., Buděšinský M.: *Magn. Reson. Chem.* **27**, 27 (1989).
16. Kamlet M. J., Abboud J.-L. M., Taft R. W.: *Prog. Phys. Org. Chem.* **13**, 485 (1981).
17. Swain C. G., Swain M. S., Powell A. L., Alunni S.: *J. Am. Chem. Soc.* **105**, 502 (1983).
18. Pytela O.: *Collect. Czech. Chem. Commun.* **53**, 1333 (1988).
19. Pytela O.: *Collect. Czech. Chem. Commun.* **54**, 136 (1989).
20. Pytela O., Zima V.: *Collect. Czech. Chem. Commun.* **54**, 117 (1989).
21. Exner O.: *Collect. Czech. Chem. Commun.* **41**, 1516 (1981).
22. Malinowski E. R., Howery D. G.: *Factor Analysis in Chemistry*. Wiley, New York 180.
23. Massart D. L., Vandeginste B. G. M., Deming S. N., Michotta Y., Kaufman L.: *Chemometrics*. Elsevier, Amsterdam 1988.
24. Wold S., Geladi P., Esbensen K., Öhman J.: *J. Chemometrics* **1**, 41 (1987).
25. Pytela O.: *Collect. Czech. Chem. Commun.* **55**, 42 (1990).

Translated by J. Panchartek.