Characteristic Vector Analysis of S_N1 Kinetic Data Influenced by Four Variables †

Romuald I. Zalewski * and Zdzisława Geltz
Department of General Chemistry, Academy of Economics, Poznań, Poland

Characteristic vector analysis has been used to explain the variability in the kinetic data of 1 260 solvolysis reactions caused by four variables. Two characteristic vectors V_1 and V_2 describe the variability with sufficient accuracy. V_1 and V_2 are linearly related to σ_0 substituent constants. S_1' and S_2' (scalar values after rotation of axes) are linearly dependent on the solvent polarity parameter Y and the reciprocal of the temperature, respectively.

An estimation of the degree of efficiency of various variables that simultaneously affect the reactivity of organic molecules has been based on the principle of additivity and independence.[1]

The additivity approach could be misleading because of the existence of perturbation among the various types of interaction. The logical consequence was to include cross-terms in the correlation analysis. Such mathematical models were designed to determine quantitatively the contribution of interacting effects in complicated physico-chemical data sets.[2]

There are few examples of the application of such models to the study of reactions influenced by a large number of independent factors. However, Pivovarov *et al.*[3] published a large set of kinetic data on solvolysis of β-dinitroalkylarylamines. The reaction takes place *via* an S_N1 mechanism, and variation of the reaction rates is influenced by a change in two substituents, temperature and polarity of the medium, and could be expressed by a complicated equation consisting of 16 variable terms and cross-terms of different complexity.

The practical application of such a procedure to explain experimental data or for the prediction of reaction rates for unknown compounds is very difficult, if not impossible.

This paper is an attempt at a more direct approach to study data sets that are influenced by a large number of independent variables. This approach is based on the characteristic vector analysis of a large data set for a solvolysis reaction. The advantage of this approach is that it is not based on a specific theory relating the experimental data to the variables, but on a general theory of the behaviour of data in similar processes.

Design of the Study.—In the present paper we studied the processes involved in the solvolysis reactions of β-dinitroalkyl-amines, of general formula $R^1C_6H_4NHCH_2C(NO_2)_2C_6H_4R^2$. The reaction rate is affected by four types of variables: R^1, R^2, temperature, and solvent composition. The reaction mechanism is S_N1 and is the same for all compounds. This is contrary to the recent results of Albano and Wold[4] who studied various S_N1 and S_N2 reactions.

In the present study 42 compounds (objects, i) were used. Six series of compounds for reaction in which $R^1 = p$-Me (1), H (2), p-Br (3), m-Br (4), p-CO$_2$Me (5), and m-NO$_2$ (6) were synthesized, each having seven new members with various substituents, $R^2 = p$-Me, p-OMe, H, p-Br, m-Cl, m-NO$_2$, and p-NO$_2$. The reaction rate for each object was measured under 30 different sets of conditions (variables, m). Reactions at five different temperatures, 10, 15, 25, 35, and 55 °C, were carried out in six different media containing 0, 30, 50, 70, 80, and 100% v/v methanol.

All the kinetic data (1 260 values) have been arranged in a

complete experimental data matrix of format 42 × 30. In each row, the kinetic data were arranged for 42 combinations of R^1 and R^2 (objects) under constant conditions (variables), *i.e.*, temperature and medium. In each column the kinetic data were arranged for one compound at 30 combinations of variables. The schematic arrangement is presented in the Scheme.

According to the specificity of characteristic vector analysis, by selecting processes of known types it is possible to use a mathematical model for: (i) classification of objects as having or not having data consistent with the model; (ii) classification of variables; and (iii) classification of processes.

We will deal with points (i) and (ii) only.

Mathematical Description of Characteristic Vector Analysis. —Characteristic vector analysis (c.v.a.) has been used in this laboratory for various purposes and has been described previously.[5,6] A brief description only will be given here. Reaction rates, log k for $i = 1,2 \ldots n$ objects and $m = 1,2 \ldots k$ variables, form a data matrix log $k_{i,m}$. The experimental data matrix could be approximated by the c.v.a. equa-

$$\log k_{m,i} = \log \bar{K}_{m,i} + V_{1,i}S_{m,1} + \ldots + V_{k,i}S_{m,k} + E_{i,m} \quad (1)$$

tion (1), where characteristic vectors V are specific to the variations of objects i and S describes their magnitudes which must be added to the mean data vector, log $\bar{K}$, in order to reconstitute the experimental data set.

The number of characteristic vectors will be equal to or less than n and in general is much less than n. The importance of the particular vector is measured by the percentage of the total variability explained by this vector, %TV, and by the residual variance of the difference between experimental and reconstituted data, W_{E^2}. The calculation is performed until %TV = 100 or just using the declared number of characteristic vectors.
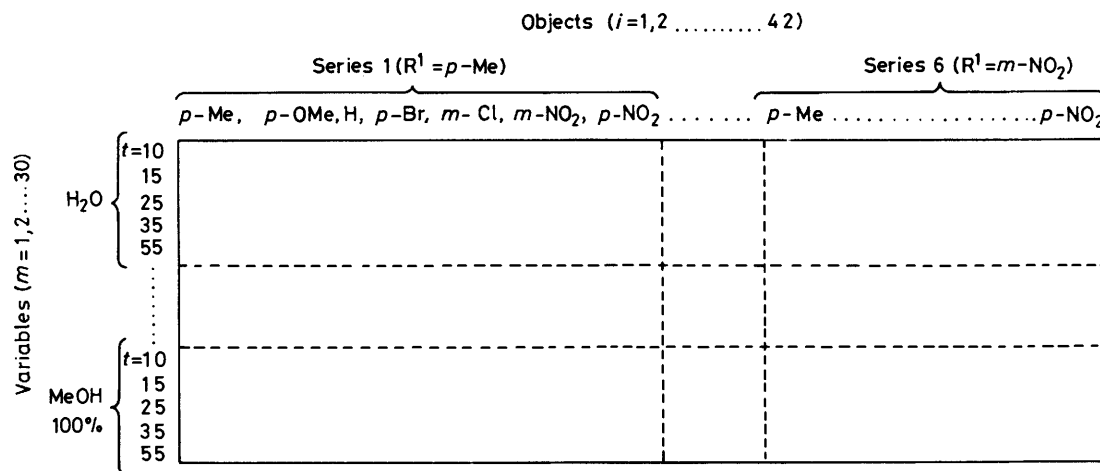
Results and Discussion

The experimental data matrix log $k_{m,i}$ can be approximated using the first characteristic vector $V_{1,i}$ with the total variability of 97.18%. The application of the second characteristic vector will improve the %TV to 99.25%.

The variance of unexplained residual $E_{i,m}$ after application of the first and second characteristic vectors calculated for all objects and variables is 0.001—0.035 and <0.01, respectively. Thus the standard deviation of reconstituted log k is acceptable when both characteristic vectors are used to explain the variability among all kinetic data, equation (2). $V_{1,i}$, $V_{2,i}$, and

$$\log k_{m,i} = \log \bar{K}_{m,i} + V_{1,i}S_{m,1} + V_{2,i}S_{m,2} + E_{m,i} \quad (2)$$

log $\bar{K}_{m,i}$ are the 42 set of numbers reported in Table 1. In the

---

† Presented at a Conference on Correlation Analysis in Organic Chemistry, Hull, 1982.

Objects (i =1,2 .... 42)

Series 1(R^1 =p-Me)   Series 6 (R^1=m-NO$_2$)

p-Me, p-OMe,H, p-Br, m- Cl, m-NO$_2$, p-NO$_2$ .... p-Me .... p-NO$_2$

Variables (m = 1,2....30)

H$_2$O { t=10 / 15 / 25 / 35 / 55 }

MeOH 100% { t=10 / 15 / 25 / 35 / 55 }

Scheme.

Table 1. Values of V_1, V_2, and log k

Series	Substituent R^1	Calculated values	Substituent R^2						
			p-Me	p-OMe	H	p-Br	m-Cl	m-NO$_2$	p-NO$_2$
1	p-Me	V_1	6.4714	6.3340	5.7885	4.6014	4.1016	2.6006	2.0479
		V_2	1.7690	1.7018	1.4369	0.8567	0.6127	−0.1157	−0.3966
		log k	−1.8170	−1.7740	−1.6021	−1.2302	−1.0724	−0.6007	−0.4297
2	H	V_1	6.4733	6.3516	5.8662	4.8163	4.3707	3.0375	2.5532
		V_2	1.1355	1.0927	0.9206	0.5522	0.3942	−0.0757	−0.2472
		log k	−2.2992	−2.2549	−2.0771	−1.6912	−1.5282	−1.0389	−0.8605
3	p-Br	V_1	6.4741	6.3750	5.9734	5.1079	4.7423	3.6424	3.2437
		V_2	0.2696	0.2590	0.2183	0.1296	0.0982	−0.0171	−0.0569
		log k	−2.9571	−2.9101	−2.7244	−2.3208	−2.1499	−1.6381	−1.4517
4	m-Br	V_1	6.4754	6.3896	6.0473	5.3098	4.9964	4.0582	3.7168
		V_2	−0.3239	−0.3114	−0.2625	−0.1564	−0.1114	0.0229	0.0724
		log k	−3.4087	−3.3601	−3.1677	−2.7522	−2.5763	−2.0487	−1.8572
5	p-CO$_2$Me	V_1	6.4783	6.4153	6.1644	5.6217	5.3873	4.6902	4.4439
		V_2	−1.2361	−1.1902	−1.0039	−0.5921	−0.4293	0.1041	0.2691
		log k	−4.1021	−4.0523	−3.8517	−3.4196	−3.2334	−2.6837	−2.4814
6	m-NO$_2$	V_1	6.4787	6.4175	6.1981	5.7322	5.5046	4.8898	4.6713
		V_2	−1.5119	−1.4581	−1.2269	−0.8098	−0.5267	0.1030	0.3434
		log k	−4.3122	−4.2609	−4.0583	−3.6273	−3.4321	−2.8731	−2.6685

first line values for seven R^2 substituents are reported, substituent R^1 being held constant (series 1). The other lines contain data for series 2—6. The linear regression between $V_{1,t}$ for series 1—6 yields a correlation coefficient of r >0.999, the data for series 2 being treated as the reference (R^1 = H). The same holds for $V_{2,t}$. In addition $V_{1,t}$ linearly relates to σ_0 substituent constants (Figure 1) with a very high correlation coefficient. $V_{2,t}$ behaves in a similar way; however, the parameters of the linear equations are different. There is no doubt that both vectors reflect the action of substituents R^1 and R^2 on reactivity.

The explanation of $S_{m,1}$ and $S_{m,2}$ values is more complicated. The plot of S_2 versus S_1 presented in Figure 2(a) indicates a complex situation. The points form straight lines. Six lines represent the data for different solvent composition (negative slope) and the other five lines represent the data for different temperatures (positive slope). However, the distribution of the data is asymmetrical. For example, S_2 values for 100% methanol overlap with the S_2 values for 80 and 70% methanol. The S_1 values behave in the same way. To avoid difficulties we rotated the S_1 and S_2 axis by 54°. The new plot $S_{m,1}'$ versus
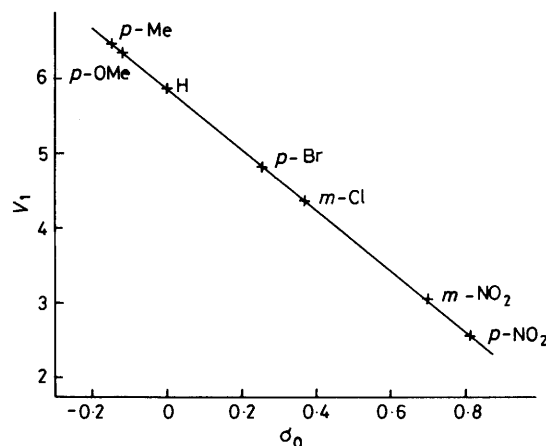

Figure 1. Values of $V_{1,t}$ for series 2 (R^1 = H) versus σ_0 substituent constants of R^2
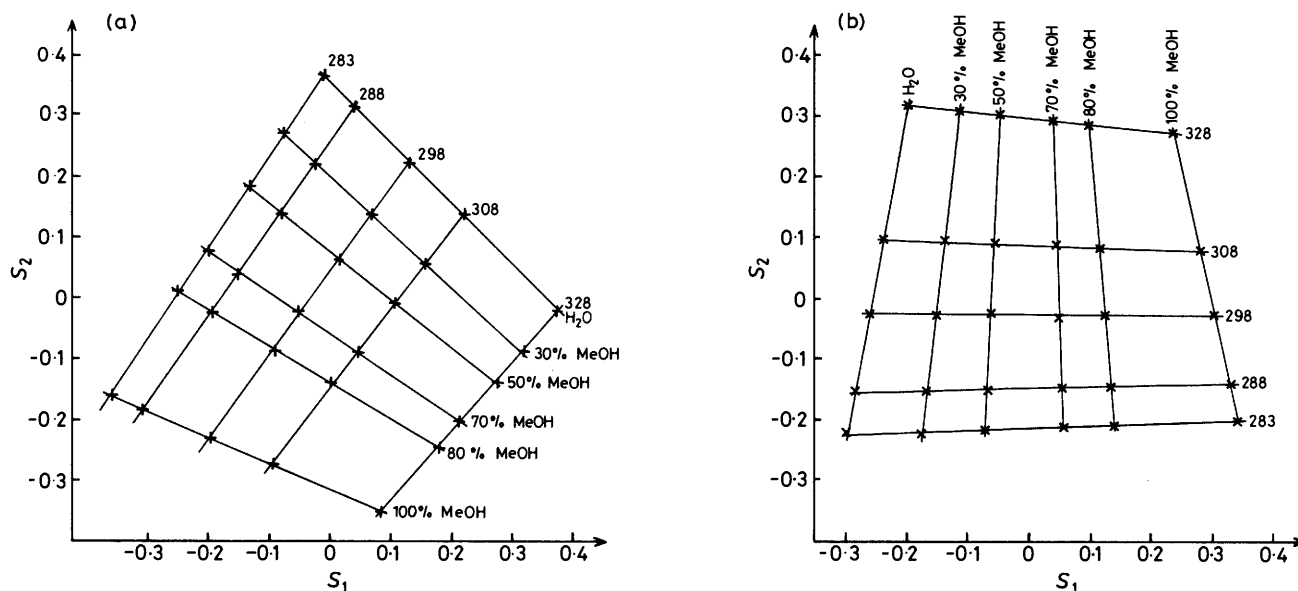
Figure 2. Plot of S_2 versus S_1: (a) indicates original data and (b) indicates data after rotation

Table 2. Values of S_1 and S_2 for all solvent–temperature combinations

Solvent	Calculated values	Temperature (K)				
		283	288	298	308	328
H₂O	S_1	−0.0106	0.0380	0.1311	0.2178	0.3758
	S_2	0.3701	0.3199	0.2259	0.1368	−0.0245
MeOH, 30%	S_1	−0.0802	−0.0286	0.0672	0.1571	0.3202
	S_2	0.2756	0.2223	0.1379	0.0571	−0.0879
MeOH, 50%	S_1	−0.1342	−0.0823	0.0160	0.1080	0.2752
	S_2	0.1845	0.1433	0.0658	<0.0066	−0.1376
MeOH, 70%	S_1	−0.2038	−0.1504	−0.0490	0.0459	0.2183
	S_2	0.0786	0.0432	−0.0243	−0.0875	−0.2022
MeOH, 80%	S_1	−0.2505	−0.1959	−0.0925	0.0043	0.1802
	S_2	0.0093	−0.0250	−0.0847	−0.1417	−0.2445
MeOH, 100%	S_1	−0.3637	−0.3067	−0.1981	−0.0966	0.0878
	S_2	−0.1629	−0.1862	−0.2314	−0.2732	−0.3503

Table 3. Parameters of regression for S_1' and S_2'

$T/°C$	$S_1' = aY + b$	r	MeOH (%)	$S_2' = c/T + d$	r
10	$0.1879 Y - 0.1295$	0.9959	0	$-1\ 121.81\ 1/T + 3.7404$	0.9999
15	$0.1853 Y - 0.1313$	0.9988	30	$-1\ 104.21\ 1/T + 3.6795$	0.9999
25	$0.1705 Y - 0.1208$	0.9987	50	$-1\ 073.41\ 1/T + 3.5769$	0.9999
35	$0.1339 Y - 0.1022$	0.9899	70	$-1\ 045.43\ 1/T + 3.4827$	0.9999
55	$0.1317 Y - 0.0926$	0.9987	80	$-1\ 026.48\ 1/T + 3.4186$	0.9999
			100	$-982.81\ 1/T + 3.2713$	0.9999

$S_{m,2}'$ is presented by Figure 2(b). The distribution of points along S_1' and S_2' is then symmetrical and their ordinates are expressed by equations (3) and (4).

$$S_1' = S_1 \cos\alpha + S_2 \sin\alpha \qquad (3)$$

$$S_2' = S_2 \cos\alpha - S_1 \sin\alpha \qquad (4)$$

$S_{m,1}'$ represents the effect of the solvent composition and, for a particular temperature, is linearly related to the polarity scale, Y.[7] At 298 K $S_{m,2}'$ is constant and is −0.025. $S_{m,2}'$ represents the effect of temperature and for a particular solvent composition is linearly related to the reciprocal of the temperature (Table 2). The parameters of regression are given in Table 3.

References

1 V. Palm, 'Fundamentals of Quantitative Theory of Organic Reactions,' Chimia, Moscow, 1967.
2 S. I. Miller, J. Am. Chem. Soc., 1959, **81**, 101.

3 S. A. Pivovarov, V. P. Selivanov, and B. V. Gidaspov, *Org. React.*, 1975, **12**, 305.

4 C. Albano and S. Wold, *J. Chem. Soc., Perkin Trans. 2*, 1980, 1447.

5 R. I. Zalewski, *J. Chem. Soc., Perkin Trans. 2*, 1979, 1637.

6 R. I. Zalewski, A. J. Sarkice, and Z. Geltz, *J. Chem. Soc., Perkin Trans. 2*, 1983, 1059.

7 A. H. Feinberg and S. Winstein, *J. Am. Chem. Soc.*, 1957, **79**, 1597, 1602, 1608; S. Winstein, H. Feinberg, and E. Grunwald, *ibid.*, p. 4146.