Journal of Chromatography, 92 (1974) 231–236

CHROM. 7340

CALCULATION OF ROHRSCHNEIDER CONSTANTS

P. SOUTER

School of Chemical Engineering, The University of New South Wales, P.O. Box 1, Kensington, N.S.W. 2033 (Australia)

(Received December 17th, 1973)

SUMMARY

Calculation of Rohrschneider-type constants characterising various solute phases has been investigated using three different methods—Rohrschneider's original procedure, a modified least-squares method incorporating his zero-sum-of-squares requirement, and the conventional least-squares method. While there is little difference in the predictive accuracy of the latter two methods when applied to Rohrschneider's data, in the interest of standardisation and compatability with other statistical procedures it is recommended that the conventional least-squares method be used in future work.

INTRODUCTION

The recent note by Leary *et al.*[1] discussed the method developed by Rohrschneider[2] for characterising the two phases in chromatographic separations. Rohrschneider constants (x, y, z, u, s) characterising the stationary solvent phase have indeed become widely used, the only areas of disagreement being how many are required and what solutes should best be used for their determination. The constants characterising the solute phase (a, b, c, d, e) appear to have found less application, partly due, no doubt, to the somewhat confused situation regarding their calculation. Leary's discussion of this question was timely.

Rohrschneider originally suggested that the retention index difference, $\Delta I_{i,j}$, for solute i on stationary solvent phase j (relative to its retention on squalane), could be expressed as a summation of a number of product terms, each term consisting of a solute-specific and a solvent-specific factor:

$$\Delta I_{i,j}\,(\text{calc.}) = a_i x_j + b_i y_j + c_i z_j + d_i u_j + e_i s_j \tag{1}$$

or more generally

$$\Delta I_{i,j}\,(\text{calc.}) = \sum_{k=1}^{n} a_{i,k} x_{k,j} \tag{2}$$

where the a factors are associated with the solute phase and the x factors with the solvent.

The usefulness of this suggestion lies in the fact that Rohrschneider was able to calculate retention times with adequate accuracy using only five product terms in eqn. 1 (*i.e.*, $n=5$). Indeed, for 22 stationary phases and 25 solutes the index differences showed a mean error of 6 index units, compared with an average experimental difference between duplicates of 3.5 index units. Considering the wide diversity of solute and solvent types used by Rohrschneider, this is an extremely good agreement and it is questionable whether attempts to reduce the errors by using more than five product terms can be statistically justified.

CALCULATION METHODS

The x factors in eqn. 2 are readily obtained experimentally from data on five selected solutes and are given by $x_{1,j} = (\Delta I/100)\text{benzene},j$;$x_{2,j} = (\Delta I/100)\text{ethanol},j$; and so on. Rohrschneider chose benzene, ethanol, butan-2-one, nitromethane and pyridine as his five standard solutes.

The a factors, on the other hand, were calculated by Rohrschneider with the intention of minimising the sum of the squared errors for each substance on all stationary phases, subject to the constraint that the sum of the errors should be zero. Putting

$$r_{i,j} = \Delta I_{i,j}\,(\text{meas.}) - \Delta I_{i,j}\,(\text{calc.}) \tag{3}$$

these conditions require that the $a_{i,k}$ values be assigned so that the expression

$$E_j = \sum_j r_{i,j}^2 \text{ is a minimum} \tag{4}$$

subject to

$$\sum_j r_{i\,j} = 0 \tag{5}$$

Later workers have used a number of different techniques in these calculations. McReynolds[3] applied regression analysis to his data (68 solutes, 25 solvents) and found that he needed seven product terms for an adequate representation. The improvement actually obtained from the two extra terms is, however, not quantitatively stated. Weiner and Howery[4] used the method of factor analysis with Rohrschneider's data and claimed that eight product terms were necessary to reproduce all these data to better than 1 % (an extremely stringent requirement in view of the experimental error of 3.5 index units between duplicates). The related method of principal-component analysis was applied by Wohl and Andersson[5] to retention data of McReynolds (10 solutes, 226 solvents) and indicated that after eliminating sixteen "abnormal" solvents, three product terms were sufficient to give a standard deviation in the error of 14.3 index units.

All these methods are related to the ordinary least-squares method in which the sum of the squared errors is minimised as in eqn. 4. The factor analysis and principal-component methods have the disadvantage that the easily measured x values are lost in the course of calculations. Rohrschneider himself, in later work[6], has chosen to evaluate solute polarity factors by the standard least-squares method instead of solving the five complicated equations he had derived earlier[2]. Indeed, in this later paper he presents the same set of equations as that given by Leary *et al.*[1].

Leary criticises Rohrschneider's original method of evaluating the solute-specific factors $a_{i,k}$ as being arbitrary and capable of giving either good or poor results. It must be pointed out that Rohrschneider was quite at liberty to choose a zero-sum constraint on his least-squares solution. Such a requirement, though novel, is essentially no less arbitrary than the usual choice of a least-squares criterion, and leads to a certain pleasing symmetry in the residual error terms. Nevertheless, any constraint of this sort necessarily causes the final mean square errors to be larger than they would otherwise be.

The difficulty in Rohrschneider's treatment lies not in his choice of original requirements, as embodied in eqns. 4 and 5, but in his subsequent manipulation of these equations to form a final working set of equations. That an error has been made in this process is apparent from an examination of his Table II, in which the errors for a given solute on all solvents do not sum to zero, as required by eqn. 5.

By taking partial derivatives of the expression E_j in eqn. 4 with respect to each of the $a_{i,k}$ factors and setting the resulting expressions to zero, a set of n equations is obtained which can be solved for the $a_{i,k}$ factors. This procedure leads to the conventional least-squares method and makes eqn. 5 redundant.

For eqn. 5 to be included in the procedure, it must be concatenated with eqn. 4 before partial derivatives are taken. The resulting set of equations will then satisfy both eqns. 4 and 5.

Zero-sum least-squares

Confining attention to a particular solute, subscript i can be dropped. It is then required to minimise the following expression:

$$\sum_{j} \left(\Delta I_j - \sum_{k=1}^{5} a_k \cdot x_{k,j}\right)^2 \tag{6}$$

subject to

$$\sum_{j} \left(\Delta I_j - \sum_{k=1}^{5} a_k \cdot x_{k,j}\right) = 0 \tag{7}$$

the ΔI_j being understood to be measured values.
Choosing a_5 arbitrarily, eqn. 7 gives

$$a_5 = \left(\sum_{j} \Delta I_j - \sum_{j} \sum_{k=1}^{4} a_k \cdot x_{k,j}\right) / \sum_{j} x_{5,j} = \left(\sum_{j} \Delta I_j - \sum_{k=1}^{4} a_k \cdot \sum_{j} x_{k,j}\right) / \sum_{j} x_{5,j} \tag{8}$$

which can be substituted into expression 6 to give

$$\sum_{j} \left\{ \Delta I_j - \frac{\left(\sum\limits_{j} \Delta I_j - \sum\limits_{k=1}^{4} a_k \cdot \sum\limits_{j} x_{k,j}\right)}{\sum\limits_{j} x_{5,j}} \cdot x_{5,j} - \sum_{k=1}^{4} a_k \cdot x_{k,j} \right\}^2$$

Collecting like terms gives

$$\sum_{j} \left\{ \Delta I_j - \frac{x_{5,j} \sum\limits_{j} \Delta I_j}{\sum\limits_{j} x_{5,j}} - \sum_{k=1}^{4} a_k \cdot x_{k,j} + \frac{\sum\limits_{k=1}^{4} a_k \cdot x_{5,j} \sum\limits_{j} x_{k,j}}{\sum\limits_{j} x_{5,j}} \right\}^2$$

which can be written

$$\sum_{j}\left(Y_j - \sum_{k=1}^{4} a_k \cdot X_{k,j}\right)^2 \tag{9}$$

where

$$Y_j = \Delta I_j - \frac{x_{5,j}\sum_{j}\Delta I_j}{\sum_{j} x_{5,j}} \tag{10a}$$

and

$$X_{k,j} = x_{k,j} - \frac{x_{5,j}\sum_{j} x_{k,j}}{\sum_{j} x_{5,j}} \tag{10b}$$

Thus, after a redefinition of variables according to eqns. 10a and 10b, an equivalent new expression, 9, containing $n-1$ product terms, is obtained. This expression can be minimised by standard least-squares methods to give the first four a values, and a_5 obtained by substitution into eqn. 8.

GEOMETRICAL INTERPRETATION

As an aid to visualising the procedures involved, Fig. 1 illustrates the relationships between the conventional least-squares method (LS), the zero-sum least-squares method (ZSLS) and Rohrschneider's procedure (RP), when applied to a situation in which only two product terms are involved. It shows the sum-of-squares contours plotted on the two-dimensional parameter space of the substance-specific factors a and b. The minimum point in the contour, corresponding to the least-squares solution for a and b, occurs at the intersection of the curves for which the partial derivatives with respect to a and b are zero. The zero-sum least-squares solu-

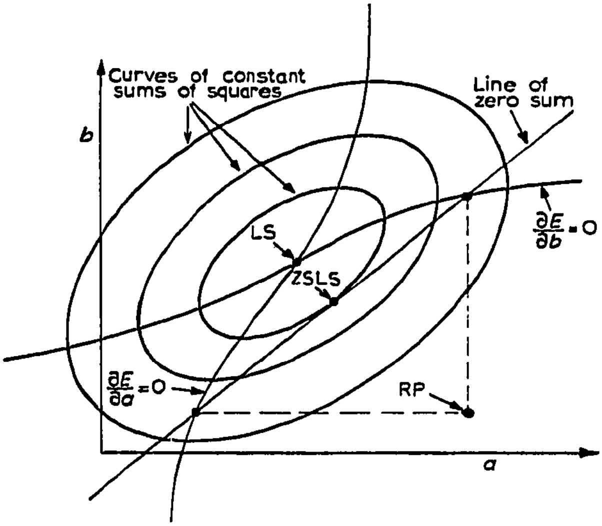

Fig. 1. Relationship between the least-squares method, the zero-sum least-squares method and Rohrschneider's procedure.

tion occurs on the line of zero sum at the point where the line is tangential to a contour line.

Rohrschneider's procedure utilises the intersection of the line of zero sum with the zero partial derivative curves, the intersection of the line with the curve for which $\partial E/\partial a = 0$, giving the value of b, and vice versa. For regularly shaped surfaces, such as that illustrated, the ZSLS solution will in general be closer to the minimum of the surface than the RP solution. Indeed, the situation may arise in which the line of zero sum fails to intersect one of the partial derivative curves. Under these circumstances there would be no possible RP solution, although the ZSLS and LS solutions would still exist.

RESULTS AND DISCUSSION

Rohrschneider's original data[2] have been used to calculate the solute-specific factors a, b, c, d, and e by means of three different procedures: (i) standard least-squares method, (ii) zero-sum least-squares method, (iii) Rohrschneider's procedure. The results for the ZSLS case are given in Table I, while in Table II the errors

TABLE I

SOLUTE-SPECIFIC FACTORS FOR ROHRSCHNEIDER'S DATA USING THE ZERO-SUM LEAST-SQUARES CRITERION

Solute	*Constant*					*Errors*		
	a	*b*	*c*	*d*	*e*	*RMS*	*Max.*	*Av. abs.*
2,4-Dimethylpentane	−20.12	0.42	12.38	−1.26	0.70	3.0	8.9	2.3
2-Ethyl-1-hexene	19.98	−0.36	5.43	−2.09	0.06	1.8	4.3	1.3
Cyclohexane	32.13	−19.97	−19.43	2.42	27.54	5.6	12.7	4.4
Toluene	103.28	0.29	5.95	−4.65	−2.18	2.1	4.8	1.5
Styrene	120.80	−2.92	−10.34	12.53	−3.19	4.8	9.0	3.7
Phenylacetylene	117.47	1.50	−75.33	58.52	5.80	5.6	10.4	4.6
Acetone	−7.00	−3.75	94.98	7.13	6.57	2.7	6.1	2.2
Propionaldehyde	10.71	−3.01	72.39	6.72	3.97	2.8	5.6	2.1
Crotonaldehyde	−0.80	−12.20	69.80	31.35	12.50	5.6	16.9	3.7
Butylacetate	−0.63	−11.18	59.09	11.63	17.83	7.1	16.3	5.4
Acetonitrile	2.12	−14.33	31.50	82.06	−0.01	5.1	8.7	4.4
Nitroethane	−2.43	−7.60	46.89	71.44	−3.04	3.7	7.0	2.9
Dioxane	50.40	−0.25	43.28	−11.04	37.12	5.0	11.4	3.8
Dibutyl ether	23.38	12.56	33.52	−17.44	−5.66	7.0	16.2	5.6
Thiophene	104.20	−3.56	−30.15	18.46	12.81	3.0	8.7	2.1
Chloroform	68.46	29.79	−70.51	50.31	−4.51	4.9	9.1	4.0
Carbon tetrachloride	69.00	−14.24	−50.18	21.02	27.39	6.2	15.4	4.3
Methyl iodide	75.62	−9.90	−37.92	25.71	18.12	7.6	14.8	6.0
Ethyl bromide	46.33	−6.45	0.24	16.49	9.13	4.4	8.3	3.5
$C_2F_2Cl_4$	14.28	−12.97	−32.36	27.07	30.95	9.3	20.2	7.7
Propanol	−5.63	101.79	−0.23	5.36	−4.23	3.1	8.2	2.3
Isopropanol	−16.25	91.99	13.94	−6.03	5.64	5.1	8.7	4.3
Allyl alcohol	21.60	113.11	−22.04	20.61	−19.33	3.5	8.2	2.9
tert.-Butanol	−12.68	72.64	30.45	−10.39	4.82	10.7	22.7	8.8
Cyclopentanol	9.94	78.85	−16.16	15.94	16.19	6.3	13.8	4.8
Overall errors (based on all 30 phases)						5.0	22.7	3.2

TABLE II

TYPICAL PREDICTION ERRORS FOR THE THREE PROCEDURES

Calculation procedure	*Errors for dioxane*				*Overall errors*			
	RMS	*Max.*	*Av. abs.*	*Av.*	*RMS*	*Max.*	*Av. abs.*	*Av.*
RP	5.9	12.3	4.6	−2.4	5.7	22.7	3.8	−1.0
ZSLS	5.0	11.4	3.8	0.0	5.0	22.7	3.3	0.0
LS	4.9	10.8	3.8	−0.5	5.0	22.7	3.3	−0.2

occurring when using these three methods for predicting retention indices for a typical solute and for the whole set of data are compared.

Results for the LS method are in complete agreement with those of Leary *et al.*[1], but results for RP do not agree so well with Rohrschneider's original results, differences of several units in the *a*, *b*, *c*, *d*, and *e* values being common. These differences might be due to calculations having been carried out with a different number of significant digits, or to the use of a different routine for solving simultaneous equations. In any case, the simultaneous equations that appear in Rohrschneider's method are poorly conditioned due to the presence of zeros throughout the leading diagonal; so differences of this order are not surprising.

The ZSLS solution gave very similar *a*, *b*, *c*, *d*, and *e* values to those obtained with the LS method, and gave prediction errors surprisingly close to those obtained with the LS method, the overall root mean square error being only marginally larger. This fact, together with the pleasing zero-sum property of the errors illustrated in Table II, might make the ZSLS solution seem an attractive alternative to the conventional LS approach, and justifies Rohrschneider's original specification of a zero-sum constraint on his procedure. However, in the present writer's opinion, the conventional least-squares method is to be preferred for polarity calculations for several reasons, *viz.* it is considerably simpler, it is closely related to the many alternative techniques available for analysing data of this sort, and it invariably leads to predictions having a smaller root mean square error.

ACKNOWLEDGEMENTS

The author is grateful to D. C. Dixon for helpful discussions on the subject, and to J. K. Haken for drawing the subject to his attention.

REFERENCES

1 J. J. Leary, S. Tsuge and T. L. Isenhour, *J. Chromatogr.*, 82 (1973) 366.
2 L. Rohrschneider, *J. Chromatogr.*, 22 (1966) 6.
3 W. O. McReynolds, *J. Chromatogr. Sci.*, 8 (1970) 685.
4 P. H. Weiner and D. G. Howery, *Can. J. Chem.*, 50 (1972) 448.
5 S. Wohl and K. Andersson, *J. Chromatogr.*, 80 (1973) 43.
6 L. Rohrschneider, *J. Chromatogr.*, 39 (1969) 383.