

## An Analysis of Errors in estimating Association Constants and Molar Extinction Coefficients from Spectrophotometric Data for 1 : 1 Molecular Complexes. Application to Literature Data

By Gino Carta\* and Guido Crisponi, Istituto Chimico Policattedra, Università di Cagliari, Via Ospedale 72, 09100 Cagliari, Italy

Association constants  $K$  and molar extinction coefficients  $\epsilon$  of 1 : 1 molecular complexes are calculated from literature data, and it is shown how the experimental error and the choice of the concentration of the reagents contributes to their respective standard deviations. Some indexes which are useful in pointing out aspects of the data are also presented. An empirical attempt is made to analyse current criteria for estimating the reliability of the parameters, and to see to what extent they are well grounded.

THE evaluation of association constant  $K$  for molecular complexes by spectrophotometry is a simple and rapid method, and one that can be applied to numerous chemical systems; furthermore, it is accurate if used carefully. This accounts for its wide employment.<sup>1-11</sup> However, in some cases literature results are characterized by large error limits and sometimes they are not reproducible when the experimental conditions are changed.<sup>12-14</sup>

In a previous paper<sup>15</sup> it was shown how variations in  $K$  and molar extinction coefficients  $\epsilon$  are functions of two different factors, (i) the experimental error affecting the data and (ii) the distribution of concentrations of the reagents involved in the complex formation. If the latter is improperly chosen, a poor determination of  $K$  and  $\epsilon$  results even with very low instrumental error.

Along these general lines we have re-examined literature data and drawn some conclusions about the statistical treatment of the data and its validity limits. To our knowledge literature data were re-evaluated by means of different non-linear least squares methods (n.l.l.s.m.) by Conrow<sup>16</sup> and Rosseinsky,<sup>17</sup> who generally found the same results but in some cases they were in clear disagreement with the original ones.

### CALCULATIONS

It has been shown<sup>15</sup> that the variations of  $K$  and  $\epsilon$  in the general case can be expressed as a function of  $G = K^{-1}[(a + b + K^{-1})^2 - 4ab]^{-1/2}$  (where  $a$  and  $b$  are the initial concentrations of the reagents) [equations (1)] where the sum

$$\begin{aligned} S_K^2 &= S^2(K_0/\epsilon_0)^2 \Sigma x^2 W / [\Sigma x^2 W \cdot \Sigma x^2 G^2 W - (\Sigma x^2 G W)^2] \\ S_\epsilon^2 &= S^2 \cdot \Sigma x^2 G^2 W / [\Sigma x^2 W \cdot \Sigma x^2 G^2 W - (\Sigma x^2 G W)^2] \end{aligned} \quad (1)$$

is over all the  $N$  experimental points.  $K_0$  and  $\epsilon_0$  are the values of  $K$  and  $\epsilon$  that minimize  $S^2 = \Sigma (D_e - \epsilon_0 x)^2 W / (N - 2)$ ,  $W$  is the weight of the single measure,  $x$  is the complex concentration calculated for  $K = K_0$ , and  $D_e$  is the experimental value of the absorbance.

If  $W = 1/D_e^2$  (which implies that the sum of square relative deviations of absorbance is minimized) and  $D_e = \epsilon_0 x$  (which neglects experimental errors) equations (1) can be arranged to (2) where  $\bar{G} = N^{-1} \Sigma G$  and  $\bar{G}^2 = N^{-1} \Sigma G^2$ .

$$\begin{aligned} S_K^2/K_0^2 &= S^2/\Sigma(G - \bar{G})^2 \\ S_\epsilon^2/\epsilon_0^2 &= S^2\bar{G}^2/\Sigma(G - \bar{G})^2 \end{aligned} \quad (2)$$

Equations (2) look like the variations of the parameters in a linear regression, with  $G$  as independent variable; therefore, the range ( $\Delta G$ ) and the distribution of this variable account for the good quality of the results. Contour lines of  $G$  ( $0 \leq G \leq 1$ ) are reported in Figure 1.

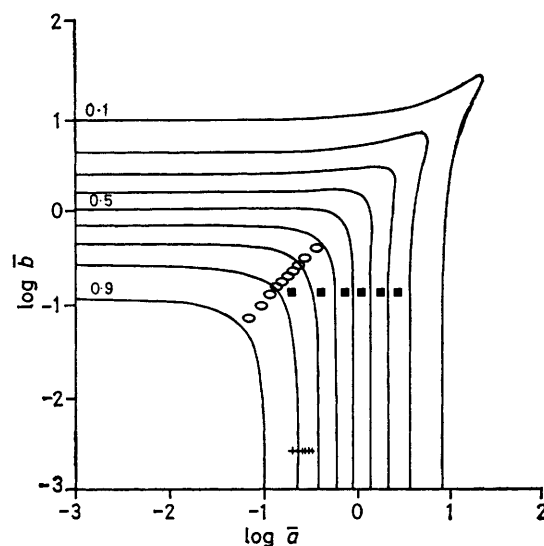


FIGURE 1  $G$  Values for refs. 21 (○), 31 (■), and 20 (+) in the Table are plotted in a ( $\log \bar{a}$ ,  $\log b$ ) plane.  $\bar{a}$  and  $b$  are the concentrations of the reagents in  $1/K$  units (i.e.  $\bar{a} = K \cdot a$  and  $b = K \cdot b$ ) and so the figures are independent of  $K$  values. Some contour lines of  $G$  are also shown with their respective values.  $G$  can be regarded as the independent variable of the system and hence its scattering controls the precision of the parameter, as can be seen for cases 23, 33, and 22 in the Table

The literature data were processed by a FORTRAN program based on two different non-linear least squares methods: (i) the Gauss-Newton linearization method<sup>18</sup> also used by Wentworth *et al.*<sup>19</sup> and by Rosseinsky;<sup>17</sup> (ii) the method proposed by Conrow *et al.*<sup>16</sup> with the difference that the complex concentrations have been calculated analytically because only 1 : 1 complexes were considered. The convergence of the first method in some difficult cases depended on the choice of the initial estimates, and sometimes there was no convergence at all; the introduction of a damping factor of 0.2 for the correction term obviates this

$$S_r = \sqrt{\Sigma (D_e - \epsilon_0 x)^2 \cdot x^2 / \Sigma x^2} \quad (3)$$

$$S_G = \sqrt{\Sigma x / \Sigma x^2 (G - \bar{G})^2} \quad (4)$$

difficulty, in spite of the greater calculation time. Using Conrow's method, the standard deviations of the parameters were obtained by relationship (1) where the  $G$  values are calculated with the estimated  $K_0$  and where  $W$  is given its proper value, generally 1. In all the examined cases there was no appreciable difference between the estimates of  $K_0$  and  $\epsilon_0$  calculated by both methods, or between their standard deviations.

The program gives as output some values that can be useful for analysing data and results critically. The relationships (3) and (4) are obtained by properly arranging the relationship (1) in the form  $S_K/K_0 = S_r \cdot S_G \cdot 1/\sqrt{N-2}$ ; the former is a measurement of the mean relative experimental error, while the latter is a function of the mean scattering of  $G$ , and hence of how well the points are chosen. A third indicator, RSM, is the ratio  $SM(K_0 + 0.5K_0)/SM(K_0)$ \* and, though it is statistically not very significant, it proves to be empirically helpful; it is equal to 1 for completely non-significant data (for which all  $G$  values are equal) and increases fast as  $S_G$  and/or  $S_r$  decreases.

#### DISCUSSION

The results shown in the Table generally agree, except in some cases, with those reported by Conrow<sup>16</sup> and Rosseinsky,<sup>17</sup> and this can probably be ascribed to some gross mistake. Sometimes our estimates differ significantly from the original ones. These differences are due to the fact that the authors fitted functions other than absorbance; therefore, automatically, a new weight is introduced for each measure, and it can be very different from the right one. This improper weighting, possibly added to difficult data (narrow  $\Delta G$ , low experimental precision, few experimental points), can lead to a false minimum; on this point the agreement with Derenleau<sup>7</sup> and Christian<sup>9</sup> is complete. The latter suggests the evaluation of the correct weight for each point and, if the linear approximation is too coarse, an iterative procedure to restore it gradually. This approach is basically faultless even though (particularly with data to linearize) it is not simpler than n.l.l.s.m. for which, generally, there is no weighting problem, and if there is, it is easier to manage.

Generally, the use of n.l.l.s.m. allows us more freedom in the choice of the experimental points, so we can explore and enlarge the  $G$  range, subject only to the restraints of the system and not to those deriving from the linearization conditions.

Discordant estimates of  $K$  and  $\epsilon$  for the same chemical system, when studied by different authors, or by different methods, or in different concentration conditions, are sometimes reported. Chemical explanations have been invoked, like interaction of the reagents with the solvent,<sup>12</sup> formation of complexes of higher order,<sup>13</sup> and variation of  $\epsilon$  of the complex with the concentration of the reagents<sup>14</sup> to account for these anomalies. Explanations of this kind are in every way legitimate, but, in our opinion, they should be invoked only after checking that discrepancies are statistically significant at a certain significance level  $\alpha$ . To avoid making incorrect

\* The term  $SM(K)$  denotes the minimum value of  $SS(K, \epsilon) = \Sigma(D_0 - \epsilon)^2 \cdot W$  for the indicated  $K$  value.

decisions (*i.e.* that the discrepancies are due to chance even when they are not)  $S_K$  should be as low as possible and this can be done by keeping the instrumental error a minimum, and  $\Delta G$  a maximum compatible with the actual restraints of the system (reagent solubilities, spectrophotometer model, available cells, *etc.*).

Several criteria have been proposed to estimate the reliability of the parameters and hence the consistency of the data. If the  $t$  test is used, a confidence interval for a parameter (*i.e.*  $K$ ) is given by relationship (5) where  $t_\alpha$

$$K_0 - t_\alpha S_K < K < K_0 + t_\alpha S_K \quad (5)$$

is the proper value of  $t$  distribution at a confidence level  $(1-2\alpha)$  for  $(N-2)$  degrees of freedom and  $S_K$  is the standard deviation  $K_0$ , estimated by relationship (1). This procedure implies both a linear model and independent parameters; the latter condition especially is rarely met in this kind of problem. The correlation between parameters can be accounted for by using the error ellipses,<sup>15,33,34</sup> which give, with respect to relationship (5), a different confidence region, leaving out points (pairs of  $K$  and  $\epsilon$ ) of lower probability. These different treatments are shown in Figure 2. The limitation introduced by approximating the true model with a linear one remains and it cannot be overcome by rigorous treatment. In any case, whatever the method of treating the data and whatever the experimental situation, it can be helpful to draw some contour lines  $SS_\alpha$  in a  $(K, \epsilon)$  plane<sup>34</sup> [relationship (6)] where  $F(2, N-2, 1-\alpha)$  is the value of

$$SS(K, \epsilon) = SS(K_0, \epsilon_0) [1 + 2/N - 2F(2, N-2, 1-\alpha)] = SS_\alpha \quad (6)$$

the  $F$  distribution with 2 and  $N-2$  degrees of freedom and a significance level of  $\alpha$  †]. The contour line  $SS(K_0, \epsilon_0)$  first of all allows us to verify whether the minimum found is the actual one; in fact, if so, the contour line should be reduced to a point. When, on the other hand, the minimum found is false, information should be obtained to locate it with better precision. The contour lines of relationship (6) give exact confidence regions, for which, however, a rigorous estimate of the right confidence level is not possible unless the model is linear. Moreover, the comparison between the ellipses and contour lines for the same confidence level shows clearly whether, and how much, the linear approximation is valid. Generally, we can say that the use of a statistical method is correct to the extent to which its confidence region overlaps the contour line for the same significance level; this is shown in Figure 2. Several cases among those reported in the Table are unsatisfactory from this point of view.

To sum up, from inspection of the Table we can see that the limits of significance of the results are above all imposed by the range of  $G$ ; therefore, once the approximate estimates of the parameters are obtained, the contour lines of  $G(a, b)$  (see Figure 1) might be helpful in

† The relationship  $\epsilon = \Sigma D_0 x / \Sigma x^2 \pm [(\Sigma D_0 x)^2 / (\Sigma x^2)^2 - \Sigma D_0^2 / \Sigma x^2 + SS_a / \Sigma x^2]^{\frac{1}{2}}$  allows us to calculate points, in a  $(K, \epsilon)$  plane, having the same  $SS_\alpha$  value when the value of  $K$  is given.



28	8	Hexamethylbenzene	CCl <sub>4</sub>	0.0703— 0.3515	2.794	0.091	3.2	2 301	45.5	1.9	0.68	12	57.8	0.51— 0.84	0.12 × 10 <sup>-5</sup>
29	9	2,6-Dichloro- <i>p</i> -benzoquinone Chloranil	CCl <sub>4</sub>	0.0007 0.001 148	(2.81) 2.811	0.081	2.9	(2 276) (1 527)	300	2.0	0.46	17	59.5	0.60— 0.80	0.55 × 10 <sup>-4</sup>
		Indole		0.0897— 0.2360	(2.86)			(1 510)							
26	5	Nickel(II)	0.5M-NaClO <sub>4</sub>	0.02—0.1	9.098	0.312	3.4	35.3	0.7	2.1	0.47	13	101	0.52— 0.82	0.16 × 10 <sup>-5</sup>
30	9	Selenocyanate Tetramethyl- <i>p</i> -phenylenedi- amine	Chloroform	0.005 0.1114— 0.3113	(8.9) 3.996	0.096	2.4	(35.5) 355.9	46.9	1.3	0.48	13	103	0.45— 0.69	0.28 × 10 <sup>-4</sup>
27	8	1,3,5-Tricyanobenzene Diphenylamine	Chloroform	2.364 × 10 <sup>-3</sup> 0.0413— 0.4127	(4.05) 0.5016	0.008	1.7	(354) 1 389	20.0	1.4	0.14	31	144	0.83— 0.96	0.18 × 10 <sup>-5</sup>
31	8	<i>s</i> -Trinitrobenzene Trimethylamine	Heptane	0.001 92 3.51 × 10 <sup>-5</sup> 5.28 × 10 <sup>-5</sup>	(0.51) 3 387	52.8	1.6	(1 391) 5 569	48.9	0.88	0.26	15	253	0.51— 0.80	0.22 × 10 <sup>-4</sup>
32	4	Sulphur dioxide Iodine	CCl <sub>4</sub>	4.55 × 10 <sup>-5</sup> 0.009 06— 0.099 79	(3 416) 69.32	1.22	1.8	(5 540) 1 510	9.00	0.60	0.48	7.9	288	0.13— 0.58	0.47 × 10 <sup>-3</sup>
		Dimethyl sulphide		0.005 44—	(71.4)			(1 490)							<i>c</i>
24	5	Carbon tetrachloride	n-Hexane	0.0998 0.1647— 1.2370	0.1096	0.002	1.7	8 232	124	1.5	0.08	36	297	0.88— 0.98	0.55 × 10 <sup>-6</sup> <i>c</i>
31	8	Mesitylene Trimethylamine	Heptane	0.0058 3.37 × 10 <sup>-5</sup> 89.7 × 10 <sup>-5</sup>	(0.112) 1 875	25.9	1.4	(1 174) 5 428	36.0	0.66	0.40	8.5	343	0.38— 0.89	0.37 × 10 <sup>-4</sup>
31	6	Sulphur dioxide Trimethylamine	Heptane	3.35 × 10 <sup>-5</sup> 6.87 × 10 <sup>-5</sup> 95.7 × 10 <sup>-5</sup>	(1 920) 1 011	15.7	1.6	(5 290) 5 205	46.0	0.88	0.27	11	365	0.51— 0.90	0.16 × 10 <sup>-4</sup>
32	4	Sulphur dioxide Iodine	CCl <sub>4</sub>	4.60 × 10 <sup>-5</sup> 0.4532 × 10 <sup>-3</sup> 9.064 × 10 <sup>-3</sup>	(1 035) 435.2	10.5	2.4	(5 000) 2 336	12.4	0.53	0.31	11	448	0.13— 0.75	0.91 × 10 <sup>-3</sup>
		Dimethyl selenide		0.4551 × 10 <sup>-3</sup> 0.182 × 10 <sup>-3</sup>	(471.7)			(2 310)							
26	4	Nickel(II)	0.5M-NaClO <sub>4</sub>	0.02—0.1	10.04	0.173	1.7	51.6	0.5	0.97	0.16	16	640	0.50— 0.69	0.16 × 10 <sup>-5</sup>
31	6	Selenocyanate Trimethylamine	Heptane	0.01 3.59 × 10 <sup>-5</sup> 49.28 × 10 <sup>-5</sup>	(10.0) 5 713	64.9	1.1	(51.6) 5 560	24.5	0.44	0.29	7.9	839	0.27— 0.77	0.13 × 10 <sup>-4</sup> <i>s</i>
		Sulphur dioxide		2.30 × 10 <sup>-5</sup>	(5 740)			(5 540)							

\* The values in parentheses are from the original papers. <sup>a</sup> These are the cases reported in Figure 2. <sup>b</sup> Frank and Oswalt proposed,<sup>15</sup> two years before Benesi and Hildebrand,<sup>1</sup> the linear approximation  $ab/D = (a + b)/\epsilon + 1/K\epsilon$  for the evaluation of  $K$  and  $\epsilon$ . <sup>c</sup> The concentrations for the fourth set of dimethyl sulphide-iodine data were taken as 10<sup>3</sup> S 99.79 and 10<sup>3</sup> B 5.438.<sup>16</sup>

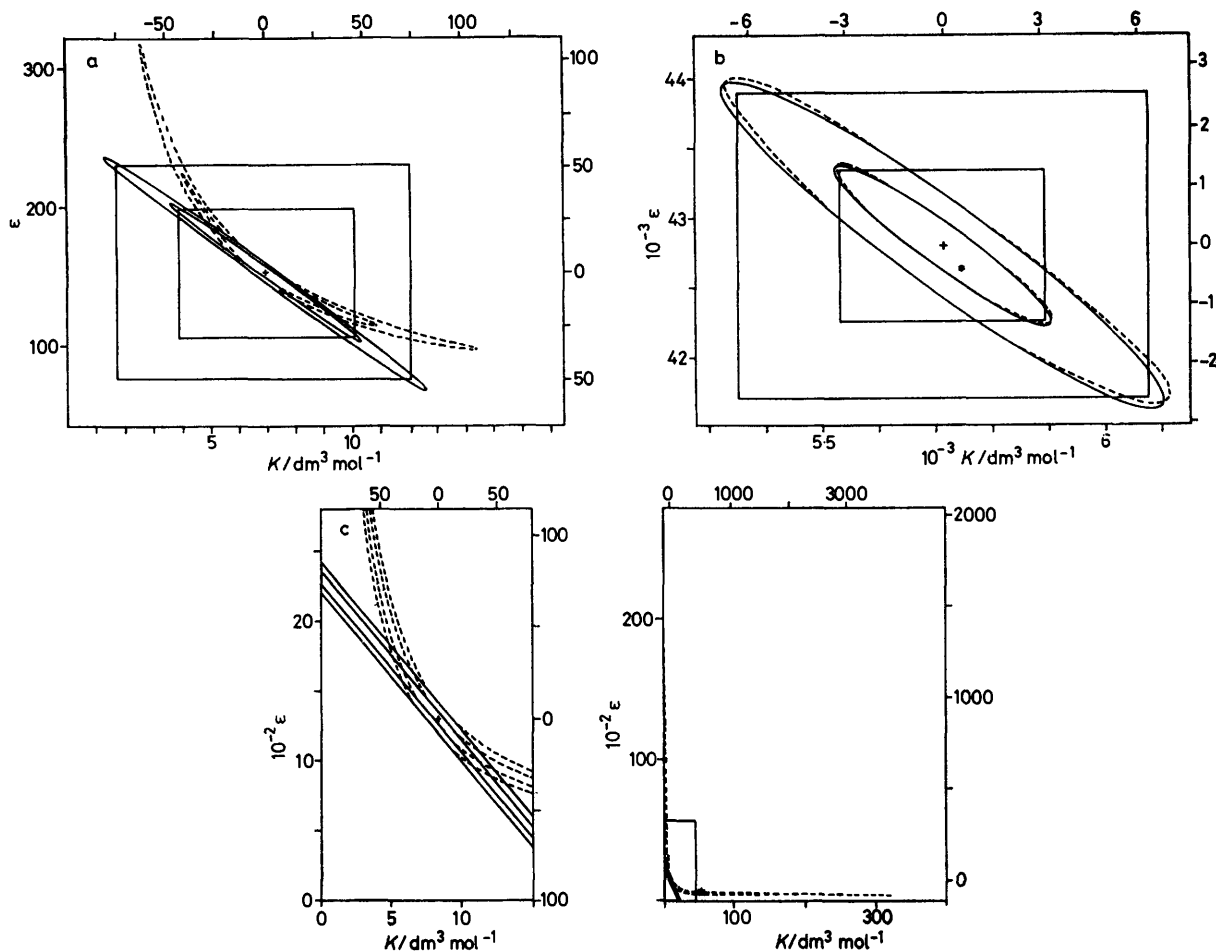


FIGURE 2 The contour lines of  $SS_{\alpha}$  according to relationship (6) (dotted lines), error ellipses (full lines), and the rectangle whose sides are given by (5) are reported for cases 21 (Figure 2a) and 31 (Figure 2b) of the Table at the 99% (outer lines) and 90% confidence levels (inner lines). In Figure 2c the curves at the 99% confidence level are presented in the right hand plot, while in the left hand plot only the central part of the curves for both confidence levels is presented, and this for a resolution problem. The upper and right scales give, respectively, the percentage deviation from  $K_0$  and  $\epsilon_0$ . The overlap percentage between the area enclosed within the contour line  $SS_{\alpha}$  and that within the considered statistics is a measure of how much this can be used correctly at a certain significance level  $\alpha$

choosing the experimental points in order to minimize  $S_G$  in a manner compatible with the system restraints. This can be done by gathering the experimental points at each end of the range of  $G$  if the 1 : 1 model is known by some physicochemical property; if not the minimization of  $S_G$  must be performed by picking out points equally distributed in the range of  $G$  (see also ref. 7). At any rate, the use of cells of different lengths makes it possible to enlarge the range of  $G$  allowing measurements of very low (or very high) absorbances, otherwise not measurable. Moreover, this device implies a better relative error of absorbance when the values of  $D_0$  are low. This experimental approach can improve the precision of the estimated parameters sometimes remarkably, and this can be very helpful when a decision has to be made in ascribing differences of results in different conditions either to physicochemical reasons or simply to statistical fluctuations.

[1/132 Received, 28th January, 1981]

#### REFERENCES

- <sup>1</sup> H. Benesi and J. H. Hildebrand, *J. Am. Chem. Soc.*, 1949, **71**, 2703.
- <sup>2</sup> J. A. A. Ketelaar, C. van der Stolpe, A. Goudsmit, and W. Dzcubas, *Recl. Trav. Chim. Pays-Bas*, 1952, **71**, 1104.
- <sup>3</sup> R. L. Scott, *Recl. Trav. Chim. Pays-Bas*, 1956, **75**, 787.
- <sup>4</sup> P. R. Hammond, *J. Chem. Soc.*, 1964, 479.
- <sup>5</sup> N. J. Rose and R. S. Drago, *J. Am. Chem. Soc.*, 1959, **81**, 6138.
- <sup>6</sup> W. B. Person, *J. Am. Chem. Soc.*, 1965, **87**, 167.
- <sup>7</sup> D. A. Derenleau, *J. Am. Chem. Soc.*, 1969, **91**, 4044.
- <sup>8</sup> R. A. La Budde and M. Tamres, *J. Phys. Chem.*, 1970, **74**, 4009.
- <sup>9</sup> S. D. Christian, E. H. Lane, and F. Gasland, *J. Phys. Chem.*, 1974, **78**, 557.
- <sup>10</sup> W. C. Coburn and E. Grunwald, *J. Am. Chem. Soc.*, 1958, **80**, 1318.
- <sup>11</sup> L. G. Sillen, *Acta Chem. Scand.*, 1962, **16**, 159.
- <sup>12</sup> M. Tamres, *J. Phys. Chem.*, 1960, **65**, 654.
- <sup>13</sup> S. D. Ross and M. M. Labes, *J. Am. Chem. Soc.*, 1957, **79**, 79.
- <sup>14</sup> P. H. Emslie, R. Foster, C. A. Fyfe, and I. Horman, *Tetrahedron*, 1965, **21**, 2843.
- <sup>15</sup> G. Carta, G. Crisponi, and V. Nurchi, *Tetrahedron*, 1981, **37**, 2115.

- <sup>16</sup> K. Conrow, G. D. Johnson, and R. E. Bowen, *J. Am. Chem. Soc.*, 1964, **86**, 1025.
- <sup>17</sup> D. R. Rosseinsky and H. Kellawi, *J. Chem. Soc. A*, 1969, 1207.
- <sup>18</sup> F. R. Bevington, 'Data Reduction and Error Analysis for the Physical Sciences,' McGraw-Hill, New York, 1969.
- <sup>19</sup> W. E. Wentworth, W. Hirsch, and E. Chen, *J. Phys. Chem.*, 1967, **71**, 218.
- <sup>20</sup> R. Bhattacharya and S. Basu, *Trans. Faraday Soc.*, 1958, **54**, 1286.
- <sup>21</sup> R. Foster, *J. Chem. Soc.*, 1957, 5098.
- <sup>22</sup> R. F. Weimer and J. M. Prausnitz, *J. Chem. Phys.*, 1965, **42**, 3643.
- <sup>23</sup> J. D. McCullough and D. Mulvey, *J. Am. Chem. Soc.*, 1959, **81**, 1291.
- <sup>24</sup> R. Anderson and J. M. Prausnitz, *J. Phys. Chem.*, 1963, **39**, 1225.
- <sup>25</sup> H. S. Frank and R. L. Oswalt, *J. Am. Chem. Soc.*, 1947, **69**, 1321.
- <sup>26</sup> L. Kullberg, *Acta Chem. Scand.*, 1974, **A28**, 897.
- <sup>27</sup> R. Foster, D. L. Hammick, and A. A. Wardley, *J. Chem. Soc.*, 1953, 3817.
- <sup>28</sup> R. Foster, D. L. Hammick, and P. J. Placito, *J. Chem. Soc.*, 1956, 3881.
- <sup>29</sup> R. Foster and P. Hanson, *Trans. Faraday Soc.*, 1964, **80**, 2189.
- <sup>30</sup> R. Foster and T. J. Thomson, *Trans. Faraday Soc.*, 1963, **59**, 2287.
- <sup>31</sup> J. Grundnes and S. D. Christian, *J. Am. Chem. Soc.*, 1968, **90**, 2239.
- <sup>32</sup> N. W. Tideswell and J. D. McCullough, *J. Am. Chem. Soc.*, 1957, **79**, 1031.
- <sup>33</sup> O. L. Davies and P. L. Goldsmith, 'Statistical Methods in Research and Production,' Oliver and Boyd, Edinburgh, 1972.
- <sup>34</sup> N. Draper and H. Smith, 'Applied Regression Analysis,' Wiley, New York, 1966.