Talanta, Vol. 32, No. 3, pp. 171–175, 1985
Printed in Great Britain. All rights reserved

0039-9140/85 $3.00 + 0.00
Copyright © 1985 Pergamon Press Ltd

MULTIPARAMETRIC CURVE FITTING—VII*

DETERMINATION OF THE NUMBER OF COMPLEX SPECIES BY FACTOR ANALYSIS OF POTENTIOMETRIC DATA

JOSEF HAVEL
Department of Analytical Chemistry, Purkyně University, Kotlářská 2, 611 37 Brno, Czechoslovakia

MILAN MELOUN†
Department of Analytical Chemistry, University of Chemical Technology, 532 10 Pardubice, Czechoslovakia

(*Received* 30 *May* 1984. *Accepted* 7 *November* 1984)

Summary—The number of complex species in solution may be determined by a computer-assisted factor analysis of a set of potentiometric titration curves, by finding the rank of the normalized data matrix. An application of the program SPECIES is demonstrated for some examples of titration data. The method is limited in that it can discriminate only between species with differing degrees of polymerization.

In the study of solution equilibria, it is not only necessary to determine the stoichiometry and stability constant of each complex species, but also to find out how many such species are present. This number can emerge from a careful least-squares treatment which seeks to find the best chemical model. One method is to use a species selector[1] in which the various complexes are tested and the best combination of them is determined. Such a treatment was first used for potentiometric data analysis by Sillén in the program LETAGROP,[1,2] and a new species selector was included recently in the program PSEQUAD.[3] The problem of species selection and determination is also discussed in a forthcoming publication.[4] With spectrophotometric data, the number of species in solution can be determined as the rank of the second moment of the absorbance matrix. This can be found by matrix-algebra operations[5] or by factor analysis.[6–8] Principal-component analysis[9,10] has been used to find the number of components from mass spectra,[11,12] fluorescence data,[13] *etc.* The use of the method in spectrophotometry is reviewed in a forthcoming monograph.[14]

This paper describes a factor-analysis method for finding the number of species present in a system studied by potentiometric titration. As with spectrophotometric data, the rank of the matrix constructed from normalized potentiometric data is shown to be less than or equal to the number of species in solution.

THEORY

Matrix analysis of normalized potentiometric data

Consider the complex-forming equilibria of components A, B and C; the general reaction equation and the corresponding overall stability constant are given by

$$p\mathrm{A} + q\mathrm{B} + r\mathrm{C} \rightleftharpoons \mathrm{A}_p\mathrm{B}_q\mathrm{C}_r \tag{1}$$

$$\beta_{pqr} = [\mathrm{A}_p\mathrm{B}_q\mathrm{C}_r]/([\mathrm{A}]^p[\mathrm{B}]^q[\mathrm{C}]^r) \tag{2}$$

The equation for the potentiometric titration curve must be transformed into the form of normalized variables $Z = f(\mathrm{pA})$ where the free equilibrium concentration of component A is usually measured potentiometrically as $\mathrm{pA} = -\log[\mathrm{A}]$ (*e.g.*, pH) and Z represents the average number of one component (for example A) bound per B or C.

$$Z = (c_\mathrm{A} - [\mathrm{A}])/c_\mathrm{B} = \left(\sum_{i=1}^{n_c} p_i[\mathrm{A}_{p_i}\mathrm{B}_{q_i}\mathrm{C}_{r_i}]\right)\Big/ c_\mathrm{B} = \left(\sum_{i=1}^{n_c} p_i c_i\right)\Big/ c_\mathrm{B} \tag{3}$$

where c_A, c_B and c_C, are the total analytical concentrations of A, B and C, [A] is the free equilibrium concentration of A, and n_c is the number of complex species in solution.

Each curve $Z = F(\mathrm{pA})$ has n_s points, and curves may be measured for n_B total concentrations of either B or C (or both), to give finally an $n_s \times n_\mathrm{B}$ matrix $\mathbf{Z}$ $[(Z_{k,j},\ k = 1,\ n_s),\ j = 1,\ n_\mathrm{B}]$. Equation (3) may be written as

$$Z = \sum_{i=1}^{n_c} e_i\, c_i \tag{4}$$

where $e_i = p_i/c_\mathrm{B}$, or in matrix notation

$$\mathbf{Z} = \mathbf{E}\ \mathbf{C} \tag{5}$$

where matrix $\mathbf{Z}$ is $(n_s \times n_\mathrm{B})$, $\mathbf{E}$ is $(n_\mathrm{B} \times n_c)$ and the concentration matrix $\mathbf{C}$ is $(n_c \times n_s)$.

Equation (5) is analogous to the generalized form of the Beer–Lambert law for unit path-length, and element e_i of $\mathbf{E}$ is analogous to the molar absorptivity.

*Part VI: *Talanta* 1984, **31**, 1083.
†Author for correspondence.

The rank of matrix **Z** may be obtained from the equation

$$\text{rank}(\mathbf{Z}) = \min[\text{rank}(\mathbf{E}), \text{rank}(\mathbf{C})] \leqq \min(n_B, n_c, n_s) \quad (6)$$

If we now arrange experimentally that $n_B > n_c$ and $n_s > n_c$ we obtain

$$\text{rank}(\mathbf{E}) \leqq n_c \; [\text{rank}(\mathbf{E}) = \min(n_B, n_c)] \quad (7)$$

$$\text{rank}(\mathbf{C}) \leqq n_c \; [\text{rank}(\mathbf{C}) = \min(n_c, n_s)] \quad (8)$$

so

$$\text{rank}(\mathbf{Z}) \leqq n_c \quad (9)$$

Thus, the rank of **Z** gives directly the number of complex species in solution. The treatment is analogous to the factor analysis of absorbance matrices by Kankare[6] and Wernimont.[15] The "spectra" of the solutions in this case are replaced by the functions $Z = f(c_B)$ or $Z = f(c_c)$.

In continuance of the analogy with absorbance matrix analysis,[6] the matrix **Z** may be subjected to factor analysis. Thus, the second moment matrix **M** is defined by

$$\mathbf{M} = (1/n_s) \cdot \mathbf{Z}\,\mathbf{Z}^T \quad (10)$$

where $\mathbf{Z}^T$ denotes the transpose of **Z**. From equation (9),

$$\text{rank}(\mathbf{M}) \leqq n_c \quad (11)$$

This implies that matrix **M** has at most n_c non-zero eigenvalues. Because of experimental errors, the actual number of non-zero eigenvalues is $\min(n_B, n_s)$. Let the eigenvalues of **M** be r_i and suppose that there are k independent components in the equilibrium system. Then the residual standard deviation of Z is given by

$$s_k(Z) = \left[\left(\text{tr}(\mathbf{M}) - \sum_{i=1}^{k} r_i\right) \Big/ (n_B - k)\right]^{1/2} \quad (12)$$

where tr(**M**) is the trace of **M**. Let the precision of estimation of Z be $s_{inst}(Z)$; then it may be concluded that if $s_k(Z) < s_{inst}(Z)$ it is probable that $n_c < k$. Values for the eigenvalues r_i, and the standard deviation of Z, $s_k(Z)$, and also for the relative variance (expressed as a percentage), RV, the cumulative percentage relative variance, CRV, and the Malinowski factor indicator function,[16] IND, are calculated and printed by the program SPECIES.

$$\text{RV}\ (\%) = [r_i/\text{tr}(\mathbf{M})] \times 100 \quad (13)$$

$$\text{CRV}\ (\%) = \sum_{i=1}^{k} r_i \times 100/\text{tr}(\mathbf{M}) \quad (14)$$

$$\text{IND} = s_k(Z)/(n_B - k)^2 \quad (15)$$

The higher the value of the variance RV, the higher the contribution of the corresponding eigenvalue to the trace, and thus the more significant the eigenvalue. Eigenvalues with variance < 0.01% are regarded as negligible. Similarly, consideration of the CRV shows how 100% variance is approached as more eigenvalues are included. If inclusion of a certain eigenvalue brings the CRV to 99.99%, the remaining eigenvalues should be assumed to be of negligible significance. The Malinowski factor indicator function IND should reach a minimum when the correct number of factors (*i.e.*, species) has been chosen.[16]

PROPOSED PROCEDURE

For various possible integer values of the rank of **Z** the standard deviation, $s_k(Z)$, is calculated from equation (12). The graph of $s_k(Z) = f(k)$ where k is an actual value of the rank, consists of two nearly linear parts which intersect on extrapolation. The point of intersection should lie near to the experimental standard deviation value, $s_{inst}(Z)$. The coordinate k on the rank-axis for this intersection is the value sought for the rank of the matrix, and is equal to the number of complex species in solution (Fig. 1). At the same value of k, $s_k(Z)$ should be $\leqq s_{inst}(Z)$, the CRV should reach almost 99.99%, the variance of any remaining eigenvalues should be negligible (< 0.1%) and IND should reach a minimum. However, these criteria may *not* all be met at the same value of k.

Limitation of the method

The situation for potentiometric data is not strictly analogous to that for spectrophotometric data, in that the method is applicable *only* when a change in the total analytical concentration of component B or C, c_B or c_c, leads to a set of different curves for $Z_B = f(\text{pA})$ or $Z_C = f(\text{pA})$. That is, the complexes that can be detected must not be mononuclear (in B or C). For mononuclear complexes the curves of $Z_B = f(\text{pA})$ are identical; thus there is just one curve, and the rank of **Z** is one. The method cannot be used, therefore, to find the number of differently protonated species of a weak acid H_nL. [In such cases the number of species may be indicated by the number of inflection points on the $Z = f(\text{pH})$ curve.]

In the case of metal–ligand equilibria, curves for varied metal or varied ligand concentrations should be measured. The matrix **Z** is then obtained by taking vertical sections at constant pH values, or by taking horizontal sections at constant Z values. Only species with differing p_i indices can be distinguished.

Computation

Computations were done with use of a mainframe computer EC 1033. The program SPECIES is available, with specimen data, on request.

DISCUSSION

The program SPECIES was validated by use of literature titration data and simulated data sets.

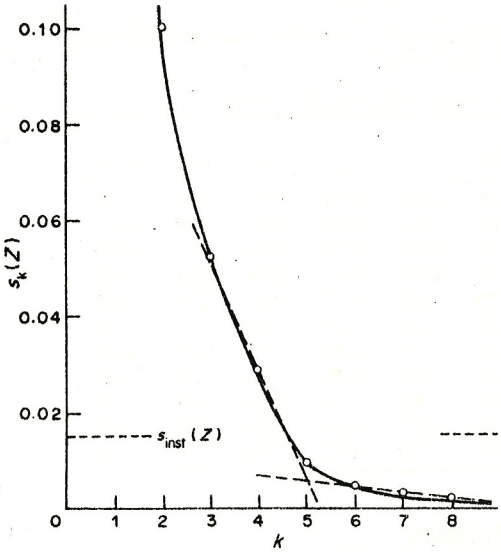

Fig. 1. Graphical determination of the number of complexes. The k-co-ordinate of the point of intersection of the two linear parts of $s_k(Z) = f(k)$ gives a good estimate of the matrix rank, and the $s_k(Z)$-co-ordinate should be close to the value of $s_{inst}(Z)$ (data from Table 2.).

Baldwin and Wiese[17] studied polymolybdate equilibria in 1M magnesium perchlorate solution and found three $H_p(MoO_4)_q$ species with (p, q) values (2, 1), (8, 6) and (9, 8). The normalized Z values for various pH values at different total concentrations of molybdate were taken from Fig. 1 in that paper and the matrix **Z** was constructed (Table 1). On the assumption that the precision of reading Z-values from this graph is $s_{inst}(Z) \approx 0.02$, the rank of the matrix should be three. It may be concluded that there are 3 polynuclear complexes in solution.

Sasaki and Sillén[18] studied the hydrolysis of molybdate in 3M sodium perchlorate and found eight $H_p(MoO_4)_q$ species with indices (1, 1), (2, 1), (8, 7), (9, 7), (10, 7) and (11,7). The Z values were read from Fig. 1 in that paper with precision $s_{inst}(Z) \approx 0.02$ for various pH values in the range 1.50–7.50. The resulting matrix **Z** was treated by program SPECIES (Table 2). The rank of the matrix is 5, in agreement with the five polynuclear species (with differing p_i indices) among those found in solution by the authors.

Biedermann and Ciavatta[19] found three complexes, $Y(OH)^{2+}$, $Y_2(OH)_2^{4+}$, and $Y_3(OH)_5^{4+}$ in hydrolysed solutions of Y(III) in 3M perchlorate. The matrix **Z** was constructed from the data of their Fig. 2 with precision $s_{inst}(Z) \approx 0.02$ (Table 3). The matrix rank is found to be 2, in accordance with the two polynuclear hydrolytic products formed in solution.

A data set was simulated for a metal-ion hydrolysis based on a study by Ahlberg[20] of Hg(II) in 3M perchlorate. Five species, $HgOH^+$, $Hg(OH)_2$, $Hg_2(OH)^{3+}$, $Hg_2(OH)_2^{2+}$ and $Hg_4(OH)_3^{5+}$ were found. The $Z = f(pH)$ curves were simulated by the program POLET,[21] then the Z values were loaded with a random error so that $s_{inst}(Z) = 0.001$. Program SPECIES showed the rank of **Z** to be 2, which

Table 1. Determination of the number of complexes in a polymolybdate system from the potentiometric data of Baldwin and Wiese[17]

Part 1. Matrix **Z** (dimensions $n_s = 10$, $n_B = 8$); $s_{inst}(Z) = 0.02$

pH \ $c_{MoO_4^{2-}}$, $10^{-3}M$	13.25	11.07	9.86	8.603	7.299	5.947	4.545	3.088
5.25	0.25	0.14	0.11	0.06	0.04	0.03	0.02	0.02
5.00	0.55	0.46	0.38	0.32	0.26	0.20	0.10	0.05
4.75	0.82	0.77	0.74	0.70	0.65	0.58	0.40	0.25
4.50	1.02	0.99	0.97	0.90	0.88	0.86	0.79	0.65
4.25	1.15	1.14	1.13	1.11	1.10	1.09	1.04	0.98
4.00	1.24	1.23	1.23	1.23	1.23	1.23	1.22	1.21
3.75	1.35	1.35	1.35	1.33	1.32	1.31	1.30	1.30
3.50	1.46	1.45	1.41	1.40	1.38	1.36	1.35	1.35
3.25	1.60	1.50	1.46	1.44	1.42	1.41	1.39	1.38
3.00	1.70	1.60	1.55	1.50	1.48	1.48	1.42	1.41

Part 2. Values of eigenvalues, r, percentage relative variance, RV, percentage cumulative relative variance, CRV, residual standard deviation, $s_k(Z)$, and Malinowski indicator function, IND, for various postulated values of matrix rank, k; Trace = 9.722430

k	r	RV	CRV	$s_k(Z)$	IND
1	9.671622	99.477	99.477	0.085195	0.001739
2	0.046129	0.474	99.952	0.027924	0.000776
3	0.004098	0.042	99.994	0.010777	0.000431
4	0.000330	0.003	99.997	0.007922	0.000495
5	0.000187	0.002	99.999	0.004616	0.000513
6	0.000034	0.000	100.000	0.003865	0.000966
Estimated rank		4	3	3	3

Table 2. Determination of the number of complexes from a potentiometric study of molybdate hydrolysis by Sasaki and Sillén[18]

Part 1. Matrix **Z** (dimensions $n_s = 11$, $n_B = 10$); $s_{inst}(\mathbf{Z}) = 0.02$

pH \ $c_{MoO_4^{2-}}$, $10^{-3}M$	160	80	40	20	10	5	2.5	1.25	0.62	0.31
6.5	0.60	0.27	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6.0	1.01	0.89	0.68	0.34	0.06	0.01	0.00	0.00	0.00	0.00
5.5	1.13	1.09	1.02	0.91	0.73	0.44	0.12	0.04	0.03	0.03
5.0	1.19	1.18	1.17	1.12	1.08	0.96	0.81	0.54	0.20	0.10
4.5	1.23	1.22	1.21	1.21	1.20	1.18	1.12	1.03	0.78	0.66
4.0	1.30	1.30	1.29	1.29	1.29	1.29	1.29	1.26	1.23	1.15
3.5	1.38	1.38	1.375	1.375	1.37	1.37	1.37	1.365	1.36	1.36
3.0	1.45	1.45	1.45	1.45	1.45	1.45	1.45	1.45	1.45	1.45
2.5	1.50	1.50	1.50	1.50	1.50	1.50	1.50	1.50	1.50	1.50
2.0	1.68	1.62	1.60	1.57	1.56	1.55	1.55	1.55	1.55	1.55
1.5	1.80	1.78	1.75	1.73	1.72	1.72	1.72	1.71	1.71	1.70

Part 2. Trace = 14.724625

k	r	RV	CRV	$s_k(Z)$	IND
1	14.255160	96.812	96.812	0.228392	0.002820
2	0.386226	2.623	99.435	0.102004	0.001594
3	0.063890	0.434	99.869	0.052574	0.001073
4	0.014082	0.096	99.964	0.029626	0.000823
5	0.004849	0.033	99.997	0.009131	0.000365
6	0.000321	0.002	99.999	0.004904	0.000306
7	0.000076	0.001	100.000	0.002582	0.000287
Estimated rank		6	5	5	(7)

Table 3. Determination of the number of complexes in the hydrolysis of Y(III), from normalized potentiometric data of Biedermann and Ciavatta[19]

Part 1. Matrix **pH** (dimensions $n_s = 8$, $n_y = 7$); $s_{inst}(\mathbf{Z}) = 0.002$

Z \ c_y, $10^{-3}M$	0.98	0.50	0.30	0.20	0.10	0.05	0.025
0.10	5.45	5.55	5.65	5.75	5.83	5.91	6.00
0.20	5.61	5.74	5.82	5.89	6.00	6.12	6.21
0.40	5.76	5.90	6.00	6.08	6.21	6.32	6.42
0.60	5.86	5.99	6.10	6.16	6.29	6.41	6.53
0.80	5.92	6.06	6.15	6.22	6.36	6.50	6.60
1.00	5.99	6.10	6.20	6.28	6.41	6.53	6.66
1.20	6.02	6.13	6.24	6.31	6.45	6.58	6.70
1.40	6.06	6.16	6.27	6.34	6.49	6.61	6.72

Part 2. Trace = 265.732100

k	r	RV	CRV	s_k(pH)	IND
1	265.731066	99.999	99.999	0.013126	0.000365
2	0.000710	0.000	100.000	0.008050	0.000322
3	0.000162	0.000	100.000	0.006362	0.000398
Estimated rank		1	1	2	2

corresponds to the two polynuclear species with different p_i indices (Table 4).

CONCLUSIONS

The graphical procedure for examination of $s_k(Z) = f(k)$ appears to yield results in agreement with the literature,[17-19] as shown in Tables 1–4. The commonly used methods for estimation of matrix rank in factor analysis, *viz.* IND, the Malinowski factor indicator function, and CRV, the cumulative relative variance, do not appear to be so reliable as criteria, at least for this particular application of factor analysis.

Table 4. Determination of the number of complexes from simulated potentiometric data for hydrolysis of Hg(II), based on the stability constants determined by Ahlberg[20]

Part 1. Matrix **Z** (dimensions $n_s = 9$, $n_B = 10$); imposed random error, $s_{inst}(Z) = 0.001$

pH \ c_{Hg}, 10^{-3}, M	2000	1300	1000	750	500	200	100	50	5	2.5
0.75	0.027	0.018	0.014	0.011	0.008	0.004	0.003	0.002	0.002	0.001
1.00	0.053	0.034	0.026	0.020	0.014	0.008	0.005	0.004	0.003	0.003
1.10	0.071	0.044	0.034	0.026	0.019	0.010	0.007	0.005	0.004	0.004
1.20	0.095	0.059	0.045	0.034	0.024	0.013	0.009	0.007	0.005	0.005
1.30	0.127	0.079	0.059	0.045	0.031	0.017	0.012	0.009	0.007	0.007
1.40	0.166	0.105	0.079	0.059	0.041	0.022	0.015	0.012	0.009	0.009
1.50	0.212	0.139	0.106	0.079	0.055	0.029	0.020	0.016	0.012	0.012
1.60	0.262	0.181	0.140	0.106	0.073	0.038	0.027	0.021	0.016	0.016
1.75	0.341	0.255	0.206	0.160	0.112	0.058	0.041	0.032	0.025	0.025

Part 2. Trace = 0.067746

k	r	RV	CRV	$s_k(Z)$	IND × 10^5
1	0.067600	99.783	99.783	0.004039	4.9864
2	0.000145	0.214	99.997	0.000473	0.7391
3	0.000002	0.002	100.000	0.000196	0.4000
4	0.000000	0.000	100.000	0.000164	0.4555
5	0.000000	0.000	100.000	0.000117	0.4680
Estimated rank		3	2	2	3

REFERENCES

1. L. G. Sillén and B. Warnqvist, *Ark. Kemi*, 1969, **31**, 341.
2. N. Ingri and L. G. Sillén, *ibid.*, 1969, **23**, 97.
3. I. Nagypál and L. Zékány, in D. J. Leggett (ed.), *Computational Methods for the Determination of Stability Constants*, Plenum Press, New York, 1984.
4. M. Meloun and J. Havel, *Computation of Solution Equilibria, Part 2, Potentiometry, Folia Fac. Sci. Nat. Univ. Purk. Brunensis*, Brno, 1985, in the press.
5. R. M. Wallace, *J. Phys. Chem.*, 1960, **64**, 899.
6. J. J. Kankare, *Anal. Chem.*, 1970, **42**, 1322.
7. M. Takatsuki and K. Yamaoka, *J. Sci. Hiroshima Univ., Ser. A.*, 1976, **40**, 387.
8. N. Ohta, *Anal. Chem.*, 1973, **45**, 553.
9. T. W. Anderson, *Ann. Math. Statist.*, 1963, **34**, 122.
10. D. Katakis, *Anal. Chem.*, 1965, **37**, 876.
11. G. L. Ritter, S. R. Lowry, T. L. Isenhour and C. L. Wilkins, *ibid.*, 1976, **48**, 591.
12. F. J. Knorr and J. H. Futrell, *ibid.*, 1979, **51**, 1236.
13. I. M. Warner, E. R. Davidson and G. D. Christian, *ibid.*, 1977, **49**, 2155.
14. M. Meloun and J. Havel, *Computation of Solution Equilibria, Part 1, Spectrophotometry, Folia Fac. Sci. Nat. Univ. Purk. Brunensis*, Brno, 1984, in the press.
15. G. Wernimont, *Anal. Chem.*, 1967, **39**, 554
16. E. R. Malinowski and D. G. Howery, *Factor Analysis in Chemistry*, Wiley, New York, 1980.
17. W. G. Baldwin and G. Wiese, *Ark. Kemi*, 1969, **31**, 419.
18. Y. Sasaki and L. G. Sillén, *ibid.*, 1968, **29**, 253.
19. G. Biedermann and L. Ciavatta, *ibid.*, 1964, **22**, 253.
20. I. Ahlberg, *Acta Chem. Scand.*, 1962, **16**, 887.
21. J. Havel and M. Meloun, *Talanta*, 1985, **32**, in the press.