Talanta, Vol. 31, No. 11, pp. 947–954, 1984
Printed in Great Britain.

0039-9140/84 $3.00 + 0.00

MULTIPARAMETRIC CURVE FITTING—V
THE GENERAL PROGRAM 'ABLET', A SYSTEM FOR REGRESSION ANALYSIS IN STUDIES OF SOLUTION EQUILIBRIA

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

JOSEF ČERMÁK
Computing Centre, College of Chemical Technology, 532 10 Pardubice, Czechoslovakia

(*Received* 22 *July* 1983. *Revised* 12 *May* 1984. *Accepted* 1 *June* 1984)

Summary—The general program ABLET is a system of subprograms for non-linear regression analysis of experimental data to find an appropriate model. The structure of ABLET provides a suitable organizational framework in which just two specific subroutines have to be supplied by the user. The resulting program can estimate non-linear model parameters with their standard deviations, test the agreement between experimental data and a mathematical model, test the accuracy and reliability of the parameters found, and simulate synthetic data for preselected parametric values. Heuristic, and/or algorithmic minimization strategies aid examination of the local and overall minima. The method of construction of the program for a particular system is discussed.

The original LETAGROP VRID of Sillén and Ingri[1,2] has been rewritten in the form of the subroutine LETAG in autocode MOST F 13 for the ODRA 1013 computer,[3] in Fortran for the Hewlett-Packard 2116B,[4] and in Fortran IV for the EC 1040 or EC 1033 computers.[5] Various programs[6–15] have been based on LETAG and applied to studies of protonation and complex-formation equilibria and determination of stability constants.[7,16–27]

More than ten years of experience with the use of LETAG, and critical comparison with other up-to-date minimization algorithms, have led to the improved version, ABLET, and revision of all our previous programs based on it. Adaptation of the ABLET system to a particular equilibrium problem is now much easier. The ABLET system also makes a statistical test of the reliability of the estimated parameters, plots a curve-fitting graph by a printer routine, and provides three minimization strategies for finding not only the global minimum but also any selected one. Minimization may be done algorithmically or by trial-and-error or by a combination of the two.

The seven programs of the ABLET family[28,29] are (1) DHLET for determination of thermodynamic dissociation constants and parameters of the extended Debye–Hückel equation, (2) DCLET for determination of dissociation constants and molar absorptivities,[30] (3) NCLET for determination of stability constants from competitive titration equilibria,[11] (4) MRLET for determination of ligand purity and the stability constant of a predominant complex from mole-ratio data,[31] (5) SPLET for determination of stability constants and parameters of a chemical model from absorbance data,[28] (6) EXLET for analysis of extraction/photometric data,[28] and (7) POLET for determination of stability constants and complex stoichiometry from potentiometric titration.[28]

The ABLET system allows the minimization subroutine to be changed from LETAG to another subroutine from the library.

THEORY

Modus operandi *of the ABLET system*

In studies of chemical equilibria it is often necessary to fit a function $f(x;\vec{\beta})$ to a set of experimental data. Unknown parameters are estimated by minimizing the difference between the experimental and calculated data:

$$U = \sum_{i=1}^{n} w_i (y_{\text{exp},i} - y_{\text{calc},i})^2 = \text{minimum} \qquad (1)$$

where w_i represents the statistical weight, $y_{\text{exp},i}$ is a single observation made for x_i, and $y_{\text{calc},i} = f(x_i; \beta_1, \beta_2, \ldots, \beta_m)$ is the functional relationship assumed to exist. Each observed y_i for a given x_i is related to the calculated value of y_i by the equation $y_i = f_i + \epsilon_i$ where ϵ_i represents a random error. Random (or observational) errors are assumed to follow a Gaussian (normal) distribution expressed as

$$f(\epsilon_i) = \frac{\exp[-\epsilon_i^2/2s^2(y)]}{\pi\sqrt{2}\,s(y)} \qquad (2)$$

Part IV—*Talanta* 1979, **26**, 569.

where $s^2(y)$ is the variance of y. Theoretically, the assumption that the model giving minimum U is the best fit is justified only if (*i*) the correct form of the equation for y_{calc} is known and used, (*ii*) there are no errors other than random errors in y, (*iii*) the random errors in y have a Gaussian distribution, and (*iv*) w_i is an exact indication of the inherent accuracy of y_i. None of these conditions is usually strictly fulfilled, but U = minimum is still widely used as a criterion because there is no better alternative.

A computer program for analysing data from equilibrium studies may usefully be constructed from logical units, each consisting of one or more subroutines. The division of the program into logical units aids the understanding of the functional structure of the whole program, Fig. 1.

The MASTER unit contains the main part of the program. The INPUT unit reads and checks data, and makes some preliminary calculations. The RESIDUAL-SQUARE SUM unit formulates the sum of squared residuals, *i.e.*, the squares of the differences between experimental and calculated values of the dependent variable. The relevant mathematical model in the form of an explicit or implicit functional relationship must be available.

Example. The dissociation equilibrium of HL^z, $HL^z \rightleftharpoons L^{z-1} + H^+$, is characterized by the thermodynamic (activity) dissociation constant $K_a^T = a_H a_L / a_{HL}$. The dependence of the mixed dissociation constant $K_a = a_H[L]/[HL]$ on ionic strength, assuming that both ions HL^z and L^{z-1} have roughly the same ion-size parameter $\mathring{a}$ (10^{-10} m) and that the overall salting-out coefficient is given by $C = C_{HL} - C_L$, is expressed by the Debye–Hückel equation in the form

$$pK_a = pK_a^T - AI^{1/2}(1 - 2z)/(1 + B\mathring{a}I^{1/2}) + CI \quad (3)$$

where

$$A = 0.5112 \text{ mole}^{-1/2}.\text{l}^{1/2}.\text{K}^{3/2}$$

and

$$B = 0.3291 \text{ mole}^{-1/2}.\text{m}^{-1}.\text{l}^{1/2}.\text{K}^{1/2}.10^{10}$$

for aqueous solutions and 25°. The mixed dissociation constant pK_a represents a dependent variable and the ionic strength I is taken as the independent variable, because it can be adjusted precisely so that its random error is less than that of the dependent variable pK_a. The three unknown parameters pK_a^T, $\mathring{a}$ and C are to be estimated by minimization of U:

$$U = \sum_{i=1}^{n} w_i(pK_{a,exp,i} - pK_{a,calc,i})^2 \quad (4)$$

The program used is DHLET.[8]

The MINIMIZATION unit contains the least-squares curve-fitting algorithm LETAG[3–5] which estimates unknown parameters by the minimization of a residual-square sum function. LETAG is based on the approach developed by Sillén and Ingri[1,2] which approximates the residual-square sum as a possible or proposed. The statistical analysis of re-

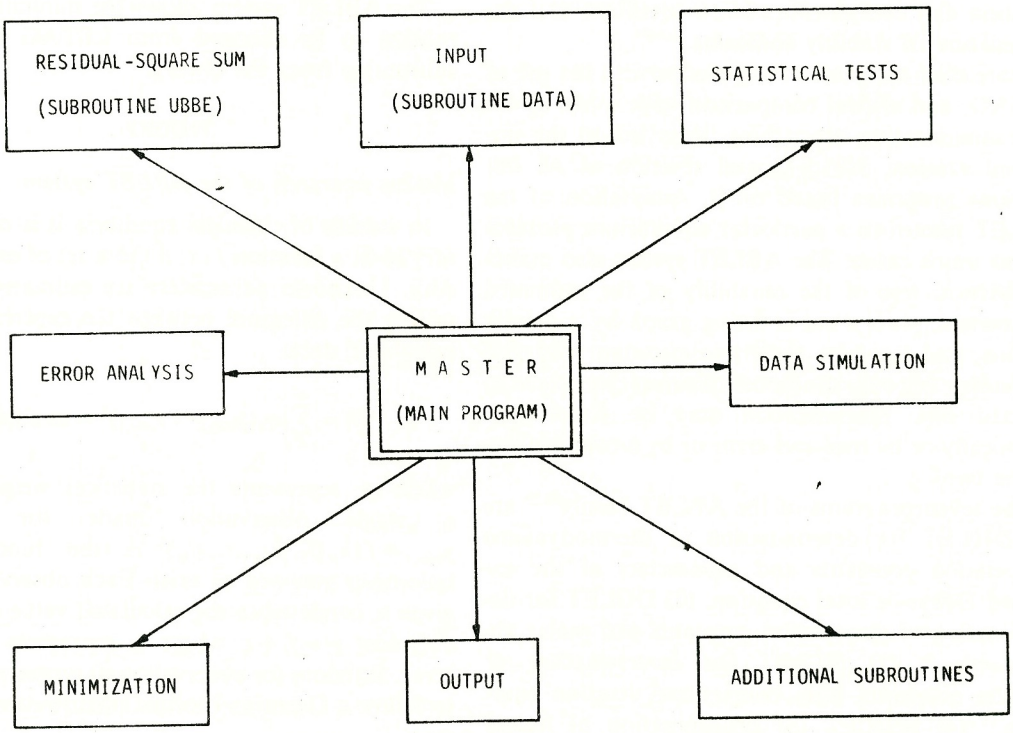

Fig. 1. The functional units of the general ABLET program. The three units to be written by the user are SUBROUTINE UBBE, SUBROUTINE DATA, MAIN PROGRAM. The rest of the ABLET system is identical for all programs.

Table 1. Shortened form of LETAG minimization process for input data listed in Table 3

DHLET PROGRAM

DHLET: SIMULATED DATA SET

xxxxxx RURIK = 7 xxxxxx LASK (INITIAL GUESS):
NUMBER OF ESTIMATED PARAMETERS = 3
NUMBER OF POSITIVE PARAMETERS = 3
NUMBER OF TWIST MATRIX ELEMENTS = 0
INITIAL GUESS OF THE FIRST PARAMETER (PKT) = 4.9
THE SECOND PARAMETER (A) = 5.0
THE THIRD PARAMETER (C) = 0.15

xxxxxx RURIK = 3 xxxxxx STEG (STEP OF PARAMETERS):
1 0.12 2 0.45 3 0.15

xxxxxx RURIK = 2 xxxxxx UTTAG (RESIDUAL-SQUARE SUM):
U = 7.67215E-01 4.90000 5.00000 0.15000

xxxxxx RURIK = 5 xxxxxx SKOTT (SHOT):
U = 7.67215E-01 4.90000 5.00000 0.15000
U = 3.16076E-01 5.02000 5.00000 0.15000
U = 1.79436E 00 4.78000 5.00000 0.15000
U = 6.36359E-01 4.90000 5.45000 0.15000
U = 9.31824E-01 4.90000 4.55000 0.15000
U = 1.38683E-01 4.90000 5.00000 0.30000
U = 2.35673E 00 4.90000 5.00000 0.00000
U = 2.46029E-01 5.02000 5.45000 0.15000
U = 3.62017E-02 5.02000 5.00000 0.30000
U = 1.14367E-01 4.90000 5.45000 0.30000
MINUSGROP (MINUS PIT)

KBOM (PARAMETERS)

NUMBER	VALUE	DARR1	DARR2
1	4.98758E 00	−1.00000	−1.00000
2	5.17657E 00	−1.00000	−1.00000
3	2.76862E-01	−1.00000	−1.00000

PROVA (TESTING)

U = 1.85920E-03	4.98758E 00	5.17657E 00	2.76862E-01	
1 ITERATION	U = 1.85920E-03	4.98758E 00	5.17657E 00	2.76862E-01
2 ITERATION	U = 1.55934E-03	4.99138E 00	5.04157E 00	2.76862E-01
3 ITERATION	U = 3.07154E-04	4.99590E 00	4.55719E 00	2.98522E-01
4 ITERATION	U = 3.03922E-04	4.99606E 00	4.55890E 00	2.98029E-01
5 ITERATION	U = 3.03919E-04	4.99605E 00	4.55925E 00	2.98007E-01
6 ITERATION	U = 3.03919E-04	4.99605E 00	4.55925E 00	2.98007E-01

second-degree surface of an elliptical paraboloid for one of the two parameters (*i.e.*, $m = 2$) in $(m + 1) = 3$-dimensional space. For more parameters than two ($m > 2$), it deals with a hyperparaboloid in $(m + 1)$-dimensional space.

Table 1 gives an example of a DHLET output which estimates three parameters, pK_a^T, $\hat{a}$ and C [from 20 data points (pK_a, I)] from Table 2. Key 7 calls the block LASK, which reads the initial guessed values for the parameters. Key 3 calls block STEG which reads the size of the initial minimization steps. Key 2 calls block UTTAG which calculates the residual-square sum function for an actual parameter value. Then a systematic search of parameters by block LETA, called by key 5, begins.

The minimization process starts with a "central" set $\vec{\beta}_c$ equal to initial guessed values for the parameters, supplied by the user, and calculates U for $\vec{\beta}_c$ and for sets where one or two elements in $\vec{\beta}_c$ have been changed by steps h_i. In the heuristic (trial-and-error) strategy, these steps are always supplied by the user, but in the algorithmic strategy they are supplied by the user only for the first iteration; for subsequent iterations they are calculated by the program.

From the U values for $(m + 1)(m + 2)/2$ systematically chosen points, the coefficients of the equation for a second-degree surface through these points are calculated, and hence the position $\vec{\beta}_0$ of the minimum on that surface. For brevity, the procedure described so far will be referred to as a "shot". The $\vec{\beta}_0$ obtained from the first shot may be used as the central value for the next shot, and so on. The co-ordinates of the pit, the lowest point of function U, are computed from the coefficients of the equations describing the approximate surface.

This pit-mapping method can also be employed when the parameters are inhomogeneous (*e.g.*, stability constants, molar absorptivities, concentrations, *etc.*). The U surface is very often distorted and not symmetrical. With such a cleft-like surface (a "skew pit"), or if there is a high degree of correlation between the β_i parameters, difficulties arise in the calculations. When the steps chosen are too large and the pit is deep and skew, the points are high up on the wall of the "cleft", and terms of third and higher degree become important. On the other hand, if the steps are too small, rounding errors in the computer can cause rounding errors in the calculation of $\vec{\beta}_0$.

Table 2 Shortened form of the last part of the output from DHLET, printed after termination of the minimization process from Table 1

7 ITERATION U = 3.03919E-04 4.99605E 00 4.55925E 00 2.98007E-01

xxxxxx SKRIK (OUTPUT) xxxxxx
PARAMETERS AND THEIR STANDARD DEVIATIONS:

FIRST (PKT):	4.99605	±0.00112
SECOND (A):	4.55925	±0.03784
THIRD (C):	0.29801	±0.00212

I	I(EXP)	PK(EXP)	PK(CALC)	RESIDUAL
1	0.0100	4.8646	4.8656	−0.0010
2	0.0400	4.7752	4.7719	0.0033
3	0.0900	4.7019	4.7054	−0.0035
4	0.1600	4.6661	4.6602	0.0059
5	0.2500	4.6407	4.6322	0.0085
6	0.3600	4.6145	4.6188	−0.0043
7	0.4900	4.6084	4.6182	−0.0098
8	0.6400	4.6318	4.6289	0.0029
9	0.8100	4.6484	4.6499	−0.0015
10	1.0000	4.6726	4.6804	−0.0078
11	1.2100	4.7179	4.7198	−0.0019
12	1.4400	4.7769	4.7677	0.0093
13	1.6900	4.8213	4.8236	−0.0023
14	1.9600	4.8896	4.8873	0.0023
15	2.2500	4.9522	4.9585	−0.0063
16	2.5600	5.0424	5.0370	0.0054
17	2.8900	5.1242	5.1226	0.0015
18	3.2400	5.2178	5.2152	0.0026
19	3.6100	5.3129	5.3147	−0.0019
20	4.0000	5.4196	5.4210	−0.0014

STATISTICAL ANALYSIS OF RESIDUALS:
RESIDUAL MEAN = 1.52E-06
MEAN RESIDUAL = 0.0041
STANDARD DEVIATION = 0.0050
SKEWNESS = 0.022
KURTOSIS = 2.390
PEARSON CHI SQUARE = 3.20 and SHOULD BE 12.60
(FOR 6 D.F. AND 0.95 PROBABILITY LEVEL)
R-FACTOR = 0.00103

The values of parameters are varied along the main axis of the pit instead of parallel to the original co-ordinate axes, in the second and subsequent refinement cycles.

For equilibrium constants and certain other types of parameters such as concentrations, a negative value has no physical meaning. LETAG[5] contains a number of safeguards to prevent such parameters from becoming negative during a shot, and at the same time to check that the calculation is not carried too far from the minimum.

If the calculated minimum $\vec{\beta}_0$ gives negative values for such parameters, they are set equal to zero, and the computer searches for the minimum of the "reduced pit". The block performing this operation, MIKO, searches the calculated second-degree surface systematically for the set $\vec{\beta}$ that gives the lowest value for U while no member of $\vec{\beta}$ is negative.

MIKO can also be used to find the set $\vec{\beta}$ that gives the lowest U value, subject to the restriction that each member β_i should have a value of either zero or not less than its standard deviation $s(\beta_i)$ multiplied by a selected factor F_s.

The ERROR ANALYSIS unit calculates the standard deviations of the estimated parameters and dependent variable. LETAG[5] is used for calculating the residual-square sum function for the pit found, U_{min}, and the standard deviation of y, $s(y)$, which is denoted by SIGY in the output in Table 1 and is the square root of U_{min} divided by the square root of the number of degrees of freedom [$(n - m)$ where n is the number of data points and m the number of parameters]:

$$s(y) = \sqrt{U_{min}/(n-m)}. \quad (5)$$

For defining the standard deviations $s(\beta_i)$ of the parameters, the term "D-boundary" was introduced by Sillén[1,2]. It is the curve or supercurve on which $U = U_{min} + s^2(y)$. The "D-boundary" is sometimes called the U contour. The standard deviation of each parameter $s(\beta_i)$ has a parabolic distribution and is calculated as the maximum difference between the value for β_i at any point on the "D-boundary" and the value for β_i at the minimum, $s(\beta_i) = D_i = [(\beta_D - \beta_0)_i]_{max}$. In the output of Table 1, $s(\beta_i)$ is printed under the heading DARR2 if it is available in the minimization process, otherwise −1.0 is printed.

The STATISTICAL TESTS unit identifies the "best" model when more than one hypothesis is

siduals involves examination of the differences between the experimental and calculated values of the dependent variable, $r_i = y_{\mathrm{exp},i} - y_{\mathrm{calc},i}$. The assumptions necessary for the least-squares treatment have already been mentioned. If the model represents the data adequately, the residuals should possess characteristics that satisfy or at least do not contradict the basic assumptions: thus the residuals should be randomly distributed about y_{calc}. Systematic departures from randomness indicate that the model is not satisfactory. The examination of plots of the residuals *vs.* x may assist numerical and/or graphical aids in the analysis of residuals. A study of the signs of the residuals (+ or −) and sums of signs can also be used.

For the analysis of residuals, graphical presentation is extremely helpful for the following tests: (*i*) detection of an outlier (an extreme residual), (*ii*) detection of a trend in the residuals, (*iii*) detection of sign changes, (*iv*) detection of an abrupt shift level of the experiment, (*v*) examination of residuals for normal distribution. A relative frequency plot should give approximately the familiar bell-shaped curve about a mean of zero.

For a more objective statistical analysis,[32] the set of residuals can be described by its four moments (*cf.* Table 2) and may be used for hypothesis testing.

(*i*) The first moment is the arithmetic mean of residuals $m_{r,1} = \bar{r}$, which should be equal to zero.

(*ii*) The second moment is the variance

$$m_{r,2} = \left(\sum_{i=1}^{n} r_i^2/n\right) - \left(\sum_{i=1}^{n} r_i/n\right)^2$$

and its square root is the standard deviation, which should have a value similar to the experimental error in the dependent variable y, $s_{\mathrm{inst}}(y)$. The same rule is applied to the mean of the residuals which is calculated as the arithmetic mean of the absolute values of the residuals.

(*iii*) The third moment, the coefficient of symmetry (skewness) gives information about the symmetry of the residual distribution curve;

$$m_{r,3} = \sum_{i=1}^{n} (r_i - \bar{r})^3/nm_{r,2}^{3/2}$$

is equal to zero for a Gaussian distribution.

(*iv*) The fourth moment, the coefficient of kurtosis, characterizes the "peakedness" of the residual distribution curve and is defined by

$$m_{r,4} = \sum_{i=1}^{n} (r_i - \bar{r})^4/nm_{r,2}^2.$$

For a normal Gaussian shape it has the value 3.

(*v*) A goodness-of-fit statistic χ^2 (Pearson test) is derived from the difference between the observed and calculated probability distribution. The residuals may be divided into eight classes, each of which should contain 12.5% of all residuals. These classes are defined by the limits $-\infty$, $-1.15s$, $-0.675s$, $-0.319s$, 0.0, $0.319s$, $0.675s$, $1.15s$, $+\infty$. Since the residuals standard deviation s is calculated from the residuals themselves, the total χ^2 has 6 degrees of freedom. A fit can be accepted at the appropriate confidence level if the experimental value χ^2 is less than the value expected.

(*vi*) The Hamilton R-factor is defined by the expression

$$R = \left[\sum_{i=1}^{n} w_i (y_{\mathrm{exp},i} - y_{\mathrm{calc},i})^2 \Big/ \sum_{i=1}^{n} w_i y_{\mathrm{exp},i}^2\right]^{1/2}.$$

This value is compared with the limiting value R_{lim} calculated by saying $r_i = y_{\mathrm{exp},i} - y_{\mathrm{calc},i}$ is the residual in the ith equation calculated from pessimistic estimates of the errors in all experimental quantities, by using the error-propagation rules. To test alternative hypotheses, the R-factor ratio test can be applied. If, for example, a particular hypothesis H_0 gives an R-factor of R_0 and an alternative hypothesis H_1 gives the value R_1, then H_1 can be rejected at the α-significance level if $R_1/R_0 > R_{m,n-m,\alpha}$ where m is the number of parameters that have been refined and $(n-m)$ is the number of degrees of freedom of the least-squares adjustment. The value of $R_{m,n-m,\alpha}$ is found from statistical tables.[33]

Table 2 gives an example of a goodness-of-fit test for the data from Table 3 and the minimization process of Table 1.

The DATA SIMULATION unit calculates a simulated curve by the addition of generated random errors to the calculated precise curve. To test the reliability of a written program and the validity of the parameters estimated from a given type of experimental curve, simulated data are often used initially.

The user supplies values for the parameters, their standard deviations if available, the values of parametric weights if the conditioning of a particular parameter in the model is bad, and the values of the standard deviation of the dependent variable $s_{\mathrm{inst}}(y)$, denoted by SINST. A set of n values of the independent variable x should also be given. The program then calculates precise values of "theoretical points" along the exact curve $y = f(x; \beta_1, \beta_2, \ldots, \beta_m)$. Each theoretical point is then converted into a simulated "experimental" one by the addition of a Gaussian error generated by a random-number generator. The four statistical moments, Pearson χ^2 test and Hamilton R-factor test are then applied to the "experimental" curve points, in order to check whether the errors are normally distributed.

The weight for each curve point, w_i, is calculated for m parameters

$$\vec{\beta} = (\beta_1, \beta_2, \ldots, \beta_m),$$

a vector of their standard deviations

$$\vec{s}(\vec{\beta}) = [s(\beta_1), s(\beta_2), \ldots, s(\beta_m)]$$

and a vector of their parametric weights

$$\vec{w}(\vec{\beta}) = [w(\beta_1), w(\beta_2), \ldots, w(\beta_m)]$$

Table 3. Simulation of pK_a–I curve points for pre-selected values of parameters, their standard deviations, their weights and errors s_{inst} (pK_a) for the dissociation $HL^- = L^{2-} + H^+$. Statistical analysis of the set of errors generated tests whether a distribution is Gaussian [$z = -1$, $(z-1)^2 - z^2 = 3$]

PRESELECTED VALUES:
INSTRUMENTAL ERROR = 0.0050

PARAMETER	STANDARD DEVIATION	WEIGHT
5.000	0.020	0.010
4.500	0.300	0.980
0.300	0.010	0.010

xxxxxx DATA xxxxxx
NUMBER OF POINTS = 20
THE CONSTANT A = 0.5115
THE CONSTANT B = 0.3291
THE CHARGE OF IONS (Z – 1)**2 – Z*2 = 3

SIMULATION OF PK–I CURVE

I	I(EXP)	PK(ACCUR)	ERROR	PK(LOADED)	WEIGHT
1	0.0100	4.869341	−0.004709	4.864632	8.371428
2	0.0400	4.775226	−0.000053	4.775172	2.681057
3	0.0900	4.708259	−0.006368	4.701891	1.488267
4	0.1600	4.662537	0.003544	4.666080	1.023263
5	0.2500	4.634170	0.006563	4.640733	0.786365
6	0.3600	4.620485	−0.005943	4.614542	0.646007
7	0.4900	4.619591	−0.011206	4.608385	0.554365
8	0.6400	4.630105	0.001697	4.631802	0.490324
9	0.8100	4.650997	−0.002646	4.648351	0.443288
10	1.0000	4.681484	−0.008908	4.672576	0.407440
11	1.2100	4.720958	−0.003016	4.717941	0.379244
12	1.4400	4.768940	0.007988	4.776928	0.356556
13	1.6900	4.825051	−0.003786	4.821265	0.337922
14	1.9600	4.888982	0.000621	4.889603	0.322366
15	2.2500	4.960482	−0.008307	4.952174	0.309191
16	2.5600	5.039344	0.003041	5.042385	0.297890
17	2.8900	5.125397	−0.001236	5.124161	0.288112
18	3.2400	5.218496	−0.000682	5.217814	0.279560
19	3.6100	5.318518	−0.005667	5.312851	0.272021
20	4.0000	5.425362	−0.005801	5.419560	0.265330

STATISTICAL ANALYSIS OF GENERATED ERRORS:
ERROR MEAN = −2.24E-03
MEAN ERROR = 0.0046
STANDARD DEVIATION = 0.0055
SKEWNESS = −0.869
KURTOSIS = 2.145
PEARSON CHI SQUARE = 6.40 AND SHOULD BE 12.60 (FOR 6 D.F. AND 0.95 PROBABILITY LEVEL)
R-FACTOR = 0.00113

according to the general scheme

$$y_{i,0} = f(x_i; \beta_1, \beta_2, \ldots, \beta_m) \tag{6}$$

$$y_{i,1} = f[x_i; \beta_1 + s(\beta_1), \beta_2, \ldots, \beta_m] \tag{7.1}$$

$$y_{i,2} = f[x_i; \beta_1, \beta_2 + s(\beta_2), \ldots, \beta_m] \tag{7.2}$$

$$\ldots \quad \ldots \quad \ldots$$

$$y_{i,m} = f[x_i; \beta_1, \beta_2, \ldots, \beta_m + s(\beta_m)] \tag{7.m}$$

$$P_j = 1 \Big/ \sum_{i=1}^{n} (y_{i,j} - y_{i,0})^2 \tag{8}$$

$$w_i = \sum_{j=1}^{m} P_j w(\beta_j) |y_{i,j} - y_{i,0}| \tag{9}$$

The value of the parametric weight $w(\beta_j)$ serves to increase the sensitivity of the jth parameter in the model. When all the parameters are well-conditioned in a model, their parametric weights are equal.

Table 3 demonstrates a data simulation for the program DHLET. The parametric values are $\beta_1 = 5.0$ ($= pK_a^T$), $\beta_2 = 4.5$ ($= \mathring{a}$), $\beta_3 = 0.3$ ($= C$). The standard deviations of the parameters can be set at any desired values; in this case the values were $s(\beta_1) = 0.02$, $s(\beta_2) = 0.3$, $s(\beta_3) = 0.01$. The statistical weight for each parameter can also be set to a suitable value to improve the sensitivity of the parameter in the model. Therefore $w(\beta_1) = 0.01$, $w(\beta_2) = 0.98$ and $w(\beta_3) = 0.01$. As the ion-size parameter $\mathring{a}$ is ill-conditioned in the model and its estimate is not so precise as the estimate of the other two parameters, its parametric weight was set at a value 98 times the parametric weights for the thermodynamic constant and salting-out coefficient. Statistical weights for each point on the pK_a–I curve are calculated according to the scheme above.

The OUTPUT unit displays the estimated results in various graphical ways: deconvolution of the experimental absorbance curve into the absorbance curves

for individual species, distribution diagrams for all the complex species in equilibrium, curves of molar absorptivities of requested species, curve fitting, fitness tests, *etc.* are possible.

The ADDITIONAL SUBROUTINES unit contains various mathematical subprograms, format-free reading subroutines, subroutines of matrix calculus, *etc.*

CONSTRUCTION OF AN ABLET PROGRAM

An ABLET program is constructed from (1) three specific units formulated by the user for the particular mathematical model of the equilibrium study and (2) six permanent units which are the same for all ABLET family programs.

Units formulated by the user

The MASTER unit contains the MAIN PROGRAM which reads in some of the organization data, keys, termination criteria, *etc.* SINST is set such that either synthetic data (SINST > 0) or experimental data (SINST = −1.0) are treated. The present version of ABLET allows up to N = 50 curve points and M = 8 unknown parameters. Enlargement of arrays for M in single precision may lead to instability in the least-squares algorithm LETAG, and is at the risk of the user. An array for the parametric vector XK(M), parametric standard deviations SIGXK(M) and parametric weights WEI(M) should be declared. Variable arrays are transferred by COMMON/FUNC/. This contains a vector for the independent variable XEXP(50), a vector for the dependent variable YEXP(50), a vector for the generated random error ERR(50), a vector for the calculated variable YCAL(50) and a vector of weight for each curve point W(50). The input channel number NI is transferred by COMMON/KANAL/ and four numerical constants for printer-plotting subroutine PLOTT and seven keys are transferred by COMMON/PLOT/ and COMMON/ISW/.

```
MAIN
DIMENSION XK(8),SIGXK(8),WEI(8)
COMMON/FUNC/XEXP(50),YEXP(50),ERR(50),
YCAL(50),W(50)/KANAL/NI
COMMON/PLOT/MY,MX,NLS,NCL,DUMMY
  (1930)/ISW/ISSW(7)
...
...
STOP
END
```

The INPUT unit should contain a version of SUBROUTINE DATA(IOU,NB) written for the particular equilibrium problem. The subroutine reads the experimental data and does some preliminary calculations. It has two arguments: IOU is an output channel number and NB the number of curve points. All variables are transferred by labelled COMMON blocks. The preliminary calculations must be specified by the user.

```
SUBROUTINE DATA(IOU,NB)
COMMON/FUNC/XEXP(50),YEXP(50),ERR(50),
YCAL(50),W(50)/KANAL/NI
CALL READI(NB,1)
...
...
RETURN
END
```

The RESIDUAL-SQUARE SUM unit should contain a subroutine UBBE(U,NK,XK,NB) which uses the parametric vector XK(M) and the independent variable vector XEXP(N) to calculate a dependent variable vector YCAL(N) and then the sum of squared residuals

$$\sum_{i=1}^{n} W(I)*[YEXP(I) - YCAL(I)]^2.$$

The arguments for UBBE are NB, the number of curve points; NK, the number of parameters; XK, the parametric vector; and U, the residual-square sum function. Other variables are transferred by a labelled COMMON.

```
SUBROUTINE UBBE(U,NK,XK,NB)
DIMENSION XK(NK)
COMMON/FUNC/XEXP(50),YEXP(50),ERR(50),
YCAL(50),W(50)
U = 0.0
DO 1 I = 1,NB
...
...
YCAL(I) = ...
1 U = U + W(I)*(YEXP(I) - YCAL(I))**2
RETURN
END
```

Permanent parts of the ABLET program

The MINIMIZATION unit contains SUBROUTINE LETAG(IOU,NAUT,NK,NB,XK,UMIN, ISSW,DARK2,DATA,UBBE,SKRIK),[5] which performs a search for parameters XK and their standard deviations SIGXK. The LETAG subroutine is divided into independent logical blocks, use of which is controlled by the value of IRUR. Its arguments are IOU, the output channel number; NK, the number of parameters; NB, the number of curve points; ISSW, the vector of keys; subroutines DATA, UBBE and SKRIK; NAUT, a key; XK, a parametric vector; DARK2, a vector of parametric standard deviations calculated in LETAG; and UMIN, the minimum of U achieved.

The STATISTICAL TEST unit contains SUBROUTINE STATS(IOU,NK,NB,X,Y,EPS) which is adapted from the program MINIQUAD.[32]

The DATA SIMULATION unit contains three subprograms, SIMUL, NORAND and FUNCTION RANDOL. In the SUBROUTINE SIMUL

(X,YC,YM,NB,ISTART) X is the vector of the independent variable; YC the vector of dependent variable calculated precisely; YM the vector of dependent variable loaded with an error generated by YM(I) = YC(I) + SINST*EPS(I) where EPS(I) is the random error generated by subroutine NORAND;[34] SINST is the standard deviation of the dependent variable; ISTART is a starting value for generation of random numbers. SUBROUTINE NORAND(D1,D2,IS) is the routine[34] for generation of random numbers and calls internal FUNCTION RANDOL(IS).

The OUTPUT unit contains SUBROUTINE SKRIK(NB,NK,IOU,IRUR,XK,SIGXK) and outputs the estimated parameters and their standard deviations, a statistical analysis of residuals, and a printer-plot of the curve-fitting done by subroutine PLOTT.[35]

The ADDITIONAL SUBROUTINES unit contains seven subroutines: WEIGHT, PLOTT, READI, READR, INVERT, PINUS and MULLE. SUBROUTINE WEIGHT(UBBE,NB,XK,SIGXK,WEI) calculates weights for simulated curve points. SUBROUTINE PLOTT(XX,YY,NDATA,NDMAX, ISYMBL,NF,XLINE,MX,YLINE,MY,NLS,NCL, MM,LL,AREA,YSCALE) makes a line-printer plot.[35] SUBROUTINE READI(I,N) and SUBROUTINE READR(A,N) are used for format-free input. I is the identifier of the variable or array to contain N integers, and A is the identifier of the variable or array to contain N reals. The input channel is defined by COMMON/KANAL/NI, *e.g.*, NI = 5. As separator between two numbers, blank(s) or comma(s) may be used, or any symbol(s) which cannot be interpreted as numbers by READI or READR (*e.g.*, 1E7 is interpreted by READR as 1.0×10^7 but by READI as two different integers 1 and 7). When *r* numbers of the same value are to be read in, the form *r** number may be used, *e.g.*, 5.52, 5.52, 5.52 may be written as 3*5.52. The form 5*b, where b is one blank, causes READI or READR to skip the next readings, leave the corresponding variables at their previous values.

Three internal subroutines of LETAG are SUBROUTINE INVERT for matrix inversion,[2] SUBROUTINE PINUS for multiplication of a vector by a matrix,[2] and SUBROUTINE MULLE for multiplication of a matrix by a matrix.[2]

CONCLUSION

Many problems in a wide variety of fields of analysis can be reduced to the problem of finding a correct mathematical model and its unknown parameters by minimizing the difference between experimental and calculated data. ABLET is a system of subprograms to solve such problems. Since in principle ABLET is designed to handle any function $y = f(x; \vec{\beta})$ it is quite general, and it may suit the needs of quite different users, not only in the study of solution equilibria.

A listing of the ABLET system and data input instructions are available on request.

REFERENCES

1. L. G. Sillén, *Acta Chem. Scand.*, 1964, **18**, 1085.
2. N. Ingri and L. G. Sillén, *Ark. Kemi*, 1964, **23**, 97.
3. J. Čermák and M. Meloun, *Sb. Věd. Pr., Vys. Šk. Chemickotechnol., Pardubice*, 1974, **32**, 37.
4. *Idem, 2nd Symposium on Algorithms in Computation Technique*, High Tatras 1974, *Knižnica algoritmov*, Vol. 2, SVTS Bratislava, 1974.
5. *Idem, Sb. Věd. Pr., Vys. Šk. Chemickotechnol., Pardubice*, 1978, **39**, 41.
6. *Idem, ibid.*, 1973, **30**, 31.
7. M. Meloun and S. Kotrlý, *Collection Czech. Chem. Commun.*, 1977, **42**, 2115.
8. M. Meloun and J. Čermák, *Sb. Věd. Pr., Vys. Šk. Chemickotechnol., Pardubice*, 1973, **30**, 49.
9. V. Říha, M. Meloun, M. Franz and J. Čermák, *ibid.*, 1975, **33**, 39.
10. M. Meloun and J. Čermák, *ibid.*, 1974, **32**, 57.
11. *Idem, Talanta*, 1976, **23**, 15.
12. J. Havel, M. Vrchlabský, J. Sekaninová and J. Komárek, *Scripta Fac. Sci. Nat. Univ. Purkynianae Brunensis*, 1979, **9**, 51.
13. L. Müllerová, M. Vrchlabský and J. Havel, *ibid.*, 1980, **10**, 13.
14. E. Mikanová, M. Bartušek and J. Havel, *ibid.*, 1980, **10**, 3.
15. M. Suchánek and L. Šůcha, *Sb. Věd. Pr., Vys. Šk. Chemickotechnol., Praha*, 1978, **H13**, 41.
16. M. Meloun and J. Pancl, *Collect. Czech. Chem. Commun.*, 1976, **41**, 2365.
17. M. Meloun and J. Chýlková, *ibid.*, 1979, **44**, 2815.
18. M. Meloun, *Iraqi J. Sci.*, 1981, **22**, 26.
19. M. Meloun, J. Chýlková and J. Pancl, *Collection Czech. Chem. Commun.* 1978, **43**, 1027.
20. M. Meloun and J. Pancl, *ibid.*, 1979, **44**, 2032.
21. J. Voštová and L. Sommer, *ibid.*, 1976, **41**, 1137.
22. L. Jančář, J. Havel, V. Kubáň and L. Sommer, *ibid.*, 1982, **47**, 2654.
23. P. Voznica, J. Havel and L. Sommer, *ibid.*, 1980, **45**, 54.
24. J. Komárek, J. Havel and L. Sommer, *ibid.*, 1979, **44**, 3241.
25. J. Votava and M. Bartušek, *ibid.*, 1975, **40**, 2050.
26. E. Mikanová, M. Mikešvá and M. Bartušek, *ibid.*, 1980, **46**, 701.
27. A. Mikan, J. Havel and M. Bartušek, *Scripta Fac. Sci. Nat., Univ. Purkynianae Brunensis*, 1980, **10**, 23.
28. M. Meloun and J. Havel, *Computation of Solution Equilibria. I. Spectrophotometry*, Folia Fac. Sci. Nat., Univ. Purkynianae Brunensis, Brno, 1983.
29. J. Havel and M. Meloun, in *Computational Methods for the Determination of Stability Constants*, D. J. Leggett (ed.), Plenum Press, New York, 1983.
30. M. Meloun and J. Čermák, *Talanta*, 1979, **26**, 569.
31. M. Meloun and M. Javůrek, *ibid.*, in press.
32. A. Sabatini, A. Vacca and P. Gans, *ibid.*, 1974, **21**, 53.
33. W. C. Hamilton, *Statistics in Physical Science*, Ronald Press, New York, 1964.
34. J. R. Bell, *Commun. ACM*, 1968, **11**, 498, Algorithm 334.
35. C. F. Moore, *Comp. Phys. Commun.*, 1971, **2**, 53.