

Computer Estimation of Dissociation Constants. Part V.* Regression Analysis of Extended Debye-Hueckel Law

Milan Meloun^{1,**}, Milan Javůrek¹, and Jiří Militký²

¹ Department of Analytical Chemistry, University of Chemical Technology, CS-532 10 Pardubice, Czechoslovakia

² Department of Textile Materials, Technical University, CS-461 17 Liberec, Czechoslovakia

Abstract. Nonlinear regression program DHMINOPT has been used for an analysis of a set of values expressing the dependence of mixed dissociation constant on ionic strength according to the extended Debye-Hueckel law. Efficiency of program has been examined on simulated data loaded with generated random errors. Goodness-of-fit brings various regression diagnostics enabling to prove a reliability of a regression process and parameter estimates. For five selected sulphonephthalein indicators, i.e. Bromocresol Green, Bromophenol Red, Bromocresol Purple, Bromothymol Blue and Phenol Red, the thermodynamic dissociation constant has been determined at 25°C together with the ion-size parameter and the salting-out coefficient.

Key words: dissociation constants, protonation equilibria, protonation constants, Debye-Hueckel law, ion-size parameter.

In determination of dissociation constants, one of the most important problems concerns the activity coefficients. Since individual activity coefficients are not accessible, some conventions have been adopted in defining them for electrolyte solutions. According to Brönsted [1] and Guggenheim [2] it can be assumed that only one linear term expressing the salting-out effect is used in an extended Debye-Hueckel model. In the previous works in this field the first attempts in estimation of thermodynamic dissociation constants, an ion-size parameter and the salting-out coefficient were published [3, 4]. Regression programs for model building in solution equilibria may be classified according to a blocks structure [4–6]. Several subroutines can make a part of one block, for example, the minimization block, the error analysis block, etc. The division of a program into such blocks is of great help in understanding an anatomy and function of a sophisticated regression program.

* Previous Part IV.: Mikrochim. Acta **1992**, 108, 227.

** To whom correspondence should be addressed

This paper pays attention to a reliability of parameters estimated and elucidates various regression diagnostics for proving an adequacy of the model proposed with experimental data with the use of new algorithm DHMINOPT.

Theoretical

Modus Operandi

To estimate thermodynamic dissociation constant, an ion-size parameter and the salting-out coefficient of the extended Debye-Hueckel equation expressing a dependence of mixed dissociation constant on an ionic strength, three programs DHMINUIT, DHFIT, DHLET are compared with new DHMINOPT. Using structural classification an elucidation of modus operandi and a comparison of programs is much easier:

(1) *Residuals sum of squares of $U(\beta)$* : This block contains the regression model being the same in four programs DHMINOPT, DHMINUIT, DHFIT and DHLET. The theoretical model $f(x; \beta)$ consists of a dependence of the mixed dissociation constant $K_a = a_H \cdot [L^{z-1}]/[HL^z]$ on an ionic strength when both ions HL^z and L^{z-1} have roughly the same ion-size parameter \AA in the dissociation equilibrium $HL^z = L^{z-1} + H^+$ with $K_a^T = a_H \cdot a_L/a_{HL}$ and that the overall salting-out coefficient is given by $C = C_{HL} - C_L$. This dependence is expressed by the Debye-Hückel equation

$$pK_a = pK_a^T - A\sqrt{I}(1 - 2z)/(1 + B\text{\AA}\sqrt{I}) + CI, \quad (1)$$

where $A = 0.5112 \text{ mole}^{-1/2} \cdot \text{l}^{1/2} \cdot \text{K}^{3/2}$ and $B = 0.3291 \text{ mole}^{-1/2} \cdot \text{m}^{-1} \cdot \text{l}^{1/2} \cdot \text{K}^{1/2} \cdot 10^{10}$ for aqueous solutions and 25°C . The mixed dissociation constant pK_a represents a dependent variable y while the ionic strength I is the independent variable x . Three unknown parameters pK_a^T (in general notation being β_1), \AA (being β_2) and C (being β_3) are to be estimated by minimization of a residual sum of squares function $U(\beta)$

$$\begin{aligned} U(\beta) &= \sum_{i=1}^n w_i (pK_{a,\text{exp},i} - f(I_i; pK, \text{\AA}, C))^2 \\ &= \sum_{i=1}^n w_i (pK_{a,\text{exp},i} - pK_{a,\text{calc},i})^2 = \text{minimum}, \end{aligned} \quad (2)$$

where w_i are the statistical weights.

(2) *Minimization*: This block searches for the best estimates of the parameters vector $\beta = \{pK_a^T, \text{\AA}, C\}$ by minimizing a difference between the experimental and calculated data $e_i = pK_{a,\text{exp},i} - pK_{a,\text{calc},i}$ so that $U(\beta)$ is minimal. The nonlinear estimation problem is simply a problem of optimization in the parameter space in which the pK_a and I values are known and given while the β values are the variables. The theorem of calculus tells us that the function $U(\beta)$ must have its smallest value at a point where (i) all derivatives $\delta U(\beta)/\delta \beta_j = 0$, $j = 1, 2, 3$, and is denoted as a stationary point; (ii) some derivatives $\delta U(\beta)/\delta \beta_j$ do not exist, and is denoted as a cusp; (iii) the value of β_j is on the boundary of the allowed region, and is denoted as an edge point.

The MINUIT procedure [7] in DHMINUIT applies three different minimization algorithms each may be used alone or in combination with two others: first, a Monte-Carlo searching non-derivative method is used at the beginning of a minimization when no reasonable starting point is known. Second, the Nelder and Mead simplex method is based on a n -dimensional convex polyhedron (simplex) specified by $(n + 1)$ vertices, i.e. a triangle in two dimensions, a tetrahedron in three, etc. Third, a derivative minimization subroutine based on a variable metric method by Fletcher of the original Davidon-Fletcher-Powell algorithm is extremely fast near a minimum or in any “nearly-quadratic” region but slower if the function $U(\beta)$ is badly behaved. DHMINUIT employs some “global” logic being built into the program: if derivative method fails, it automatically causes the Nelder and Mead method to be called to make another attempt. In addition, the minimization can be guided or separated into steps which cause a variable parameter to be fixed at a constant value or restored to variable status in between minimization steps.

DHFIT employs the algorithm FIT [8] based on the Morrison derivative method.

DHLET employs the algorithm LETAG [4] based on the modified Sillen’s LETAGROP-VRID method [9]. The principle of the method of “pit-mapping” (in Swedish leta-grop) is the approximation of a criterion function $U(\beta)$ in vicinity of $\beta^{(i)}$ of the i -th iteration by the m -dimensional elliptic hyperparaboloid. The coefficients of this hyperparaboloid are calculated from $(m + 1)(m + 2)/2$ points $\{\beta_j, U(\beta_j)\}$. Substituting these points into the equation for m -dimensional hyperparaboloid we get $(m + 1)(m + 2)/2$ linear equations for an estimation of their coefficients. Knowing these coefficients from the analytical derivation a minimum of approximate paraboloid may be calculated and therefore also the vector $\beta^{(i+1)}$.

DHMINOPT employs the modified double-dog strategy as an alternative to Marquardt method proposed by Militky [11, 14]. The principle of this derivative method is to reach an acceptable minimization step between direction of linear approximation of $f(x; \beta)$ and antigradient direction. This step is calculated as a linear combination of these two directional vectors [14].

(3) *Statistical analysis*: This block calculates confidence intervals of parameters and correlation coefficients. The square-roots of the diagonal elements of the covariance matrix are the standard deviations of the parameters.

The normalized off-diagonal elements are the paired correlation coefficients r_{ij} between the i th and j th parameters expressing the interdependence of two parameters when other parameters are not been assumed. The multiple correlation coefficient R_i is a measure of the dependence of the given i th parameter on linear combination of all the others. If correlation coefficient r_{ij} is equal to zero, the two parameters are uncorrelated, while when $r_{ij} = 1$ or -1 the two parameters are completely correlated (collinear). Highly correlated parameters indicate that the elliptic hyperparaboloid shape exhibits a shallow pit.

DHMINUIT is able to plot the $U(\beta)$ function contours in the space of any two chosen variables at a time. $U(\beta)$ contours give the most detailed description of the shape of the residual sum of squares function (cf. Fig. 1).

The suitable determination of the confidence interval of parameter β_k^* is based on the maximal length Δ_k of the projection Δ_{kj} into the parameter axis β_k . In DHLET [5–6] the estimate of standard deviation of the k -th parameter, β_k^* , is calculated by

$$\Delta_k = \max_j (\Delta_{kj}) \quad (3)$$

and the confidence interval of the parameter β_k is estimated by

$$b_k - \Delta_k \leq \beta_k^* \leq b_k + \Delta_k. \quad (4)$$

Simpler is instead of projections to search directly coordinates of extreme points on the confidence ellipsoid in directions of individual parameter axes [13].

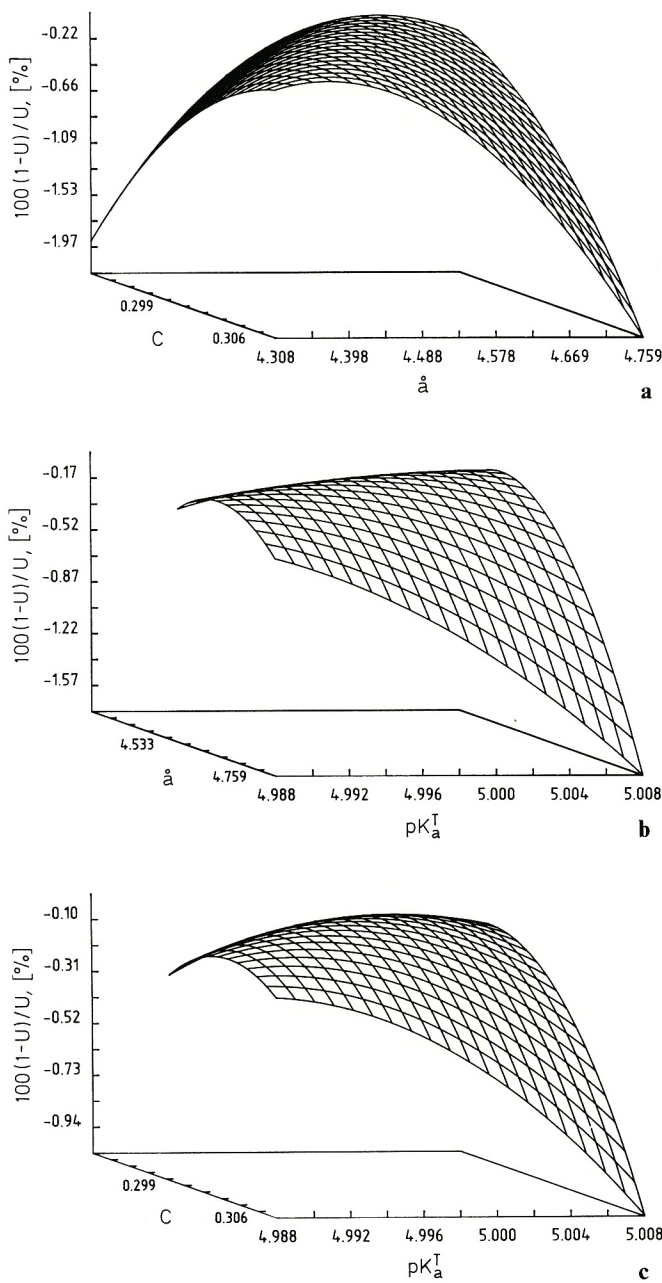


Fig. 1. The 3D graph of a $(1 - U(\beta))$ response surface for $pK_a - I$ data from Table 1 indicates **a** that \bar{a} and C are well-conditioned in the model because the surface exhibits an obvious maximum, **b** two ill-conditioned parameters \bar{a} and pK_a^T , and **c** two ill-conditioned parameters pK_a^T and C . For both cases, **b** and **c**, there is no well defined maximum $(1 - U(\beta))$

All confidence intervals are symmetrical.

If a linearization of the regression model (1) can be used the $100(1 - \alpha)\%$ th confidence interval of a prediction $f(x^*; b)$ in the point x^* may be calculated too [14]. The confidence intervals of prediction calculated in DHMINOPT for the whole range of independent variable I form the *confidence bands*.

(4) *Goodness-of-fit*: This block contains the examination of fitness achieved by the statistical analysis of residuals. The residuals are defined as the differences

$$\hat{e}_i = pK_{a,exp,i} - pK_{a,calc,i}, \quad i = 1, \dots, n, \tag{5}$$

where $pK_{a,exp,i}$ is the i -th observation and $pK_{a,calc,i}$ is the i -th prediction. If the proposed model represents the data adequately, the residuals should form random pattern. Systematic departures from randomness indicate that model is not satisfactory and detect: (i) an outlier or an extreme observation; (ii) a trend in the residuals; (iii) a sign change; (iv) an abrupt shift of level in the experiment.

In many regression programs used the statistical analysis of residuals represents the main diagnostic tool and a resolution criterion in a search of the "best" model when more than one are possible or proposed. The following statistical characteristics of residuals can be used for a fitness evaluation: (1) The *arithmetic mean of residuals* known as the *residual bias*, $E(\hat{e})$, should be equal to zero; (2) The *mean of absolute values of residuals*, $E|e|$, and the square-root of the residual variance $s^2(\hat{e})$ known as the estimate of the *residual standard deviation*, $s(\hat{e})$, should be both of the same magnitude as the (instrumental) error of dependent variable $pK_{a,exp}$, $s_{inst}(pK_a)$. Obviously it is also valid that $s(\hat{e}) \approx s_{inst}(pK_a)$; (3) The *residual skewness*, $g_1(\hat{e})$, should be for Gaussian normal distribution of residuals equal to zero; (4) The *residual curtosis*, $g_2(\hat{e})$, should be for Gaussian normal distribution equal to 3; (5) The *determination coefficient* D^2 is computed from the relation

$$D^2 = 1 - \frac{U(b)}{\sum_{i=1}^n (pK_{a,exp,i} - \overline{pK_{a,exp}})^2}, \quad (6)$$

where $\overline{pK_{a,exp}} = (1/n) \sum_{i=1}^n pK_{a,exp,i}$. The determination coefficient is for linear models equal to square of multiple correlation coefficient; (6) When determination coefficient is multiplied by 100%, we receive so called *regression rabat*, $D^2 \cdot 100[\%]$; (7) In chemometrics the *Hamilton R-factor of relative fitness* is often used being expressed by

$$R = \sqrt{\frac{U(b)}{\sum_i pK_{a,exp,i}^2}}. \quad (7)$$

For $\overline{pK_{a,exp}} = 0$ it is valid that $R^2 = 1 - D^2$ and so that the following relation between R and D is valid

$$R = \sqrt{(1 - D^2) - \frac{(1 - D^2)n \overline{pK_{a,exp}}^2}{\sum_i pK_{a,exp,i}^2}}. \quad (8)$$

The Hamilton R-factor of relative fitness exhibits a difference between two models, $pK_a = f(I; \beta)$ and $pK_a = 0$. This rule is not correct for models with an intercept term and the values of the Hamilton R-factor are *incorrectly low*. It should be noted that D^2 also R-factor are continuous function of the number of parameters. While D^2 is an increasing function of the number of parameters, the R-factor is decreasing function of this number. Therefore, both statistics are not convenient as the resolution diagnostic for a search of models of different numbers of parameters. (8) To distinguish between various models the *Akaike information criterion* AIC is more suitable to apply which is defined by relation

$$AIC = -2L(b) + 2m, \quad (9)$$

where m is a number of estimated parameters (here is 3). The "best" model is considered to be a model for which this criterion reaches a minimal value. Using the least-squares and models which do not belong into the same class the AIC criterion may be expressed

$$AIC = n \ln \left[\frac{U(b)}{n} \right] + 2m, \quad (10)$$

where m is 3. To above application of statistical analysis of classical residuals \hat{e} it should be critically noted that the diagnostic use of classical residuals is not rigorous but of a rather approximate character. The classical residuals do not exhibit zero mean, they are biased and they are a linear combination of errors ε . Moreover, they are dependent on true values of parameters β^* which are unknown.

While for linear regression models all characteristics for an identification of influential points are function of residuals \hat{e}_i and diagonal elements H_{ii} of projection matrix $H = X(X^T X)^{-1} X^T$ for nonlinear regression models the situation is rather more complicated as the parameter estimates and residuals cannot be expressed so simply as the linear combination of experimental data. When the Taylor type linearization of original nonlinear model is used, all methods of identification of influential points in linear models can be used. The procedure starts from the one-step approximation of the parameter estimate computed without i -th point

$$b_{(i)}^1 = b - (J^T J)^{-1} J_i \frac{\hat{e}_i}{1 - P_{ii}}, \quad (11)$$

where P_{ii} are elements of a projection matrix, $P = J(J^T J)^{-1} J^T$, cf. ref. [14].

The influential points may be easily identified on base of the one-step approximation of the jackknife residuals \hat{e}_{ji} calculated by

$$\hat{e}_{ji} = \frac{\hat{e}_i}{\hat{s}_{(i)} \sqrt{1 - P_{ii}}}, \quad (12)$$

where $\hat{s}_{(i)}^2$ is residual variance computed by using estimates $b_{(i)}$

$$\hat{s}_{(i)}^2 = \frac{U(b) - \frac{\hat{e}_i^2}{1 - P_{ii}}}{n - 4}. \quad (13)$$

Jackknife residuals higher than 3 indicate highly influential points.

Nonlinear measure of an influence of the i -th point on the parameter estimates is represented by the *likelihood distance*

$$LD_i = 2[\ln L(b) - \ln L(b_{(i)})]. \quad (14)$$

In case of the least-squares the likelihood distance is expressed by

$$LD_i = n \ln \left[\frac{U(b_{(i)})}{U(b)} \right]. \quad (15)$$

When $LD_i > \chi_{1-\alpha}^2(2)$ is valid the i -th point is *strongly influential*. The significance level α is usually chosen to be equal to 0.05 then $\chi_{0.95}^2(2) = 5.992$.

(5) *Data simulation*: This block serves for debugging a program or for an examination of reliability of parameters estimation. For chosen parameters and their standard deviations, the “theoretical points” along the exact curve $pK_a = f(I; pK_a^T, \hat{a}, C)$ are calculated. Each theoretical point is then transformed into an “experimental” one by the addition of a random error (having, for example, a Gaussian normal distribution) obtained with the aid of a random-number generator.

Loading the theoretical points with high random error may, however, decrease the accuracy and precision of the parameters estimated. When several parameters are to be refined or ill-conditioned parameters are in model, data with a low precision may result in erroneous values of the parameter estimates.

Quality of Regression

The quality of nonlinear regression model is examined using following criteria:

- (1) *The quality of parameter estimates:* The quality of parameter estimates is considered from their confidence intervals or from their variances $D(b_j)$. Often the empirical rule is used: the parameter is considered to be significant when its estimate is greater than 3x its standard deviation. High values of parameters variance are often caused by termination of minimization process before reaching a minimum.
- (2) *The quality of curve fitting:* An agreement of proposed model with experimental data is examined by the goodness-of-fit characteristics based on the statistical analysis of residuals.
- (3) *The quality of experimental data:* For examination of a quality of experimental data the identification of influential points by regression diagnostics is used. The suitable diagnostics are the likelihood distances LD_i and jackknife residuals \hat{e}_j .

Software

DHMINOPT was applied from CHEMSTAT package [14] (Trilobyte, Pardubice) on IBM PC while other computations (DHFIT, DHLET, DHMINUIT) were performed on the EC1033 computer at the Computing Centre of the University of Chemical Technology, CS-532 10 Pardubice, Czechoslovakia.

The package CHEMSTAT is available from authors upon request.

Results and Discussion

There were 20 points of dependence $pK_a = f(I)$ calculated for pre-selected values of parameters $pK_a^T = 5.000$, $\hat{a} = 0.45$ and $C = 0.300$ and loaded with random errors generated for an instrumental error of dependent variable $s_{inst}(pK_a) = 0.005$. For an initial guess of parameters $(pK_a^T)^{(0)} = 1$, $(\hat{a})^{(0)} = 0$, $C^{(0)} = 1$ the sum of squares $U(b^{(0)})$ is 325.7. The program DHMINOPT reaches a minimum $U(b) = 4.94 \cdot 10^{-4}$ with parameters estimates $pK_a^T = 4.997$, $\hat{a} = 4.524$, $C = 0.299$ while DHMINUIT $U(b) = 4.957 \cdot 10^{-4}$ with 4.997, 4.525, 0.299, DHFIT with 4.997, 4.525, 0.299 and DHLET with 4.996, 4.559, 0.298. The point and interval estimates found by DHMINOPT are in Table 1.

As ill-conditioned parameters (here \hat{a}) in Fig. 1 have little influence on the residual sum of squares function $U(\beta)$, they have large value of parameters standard deviation and their determination is less certain. On the other hand, well-conditioned parameters (here pK_a^T , C) lead to a pronounced maximum in $(1 - U(\beta))$ so that they have a great influence on the hyperparaboloid and parameters standard deviations are rather small.

The multiple correlation coefficient $R(\hat{a}) = 0.9515$ reaching largest value proves that \hat{a} is strongly correlated with pK_a^T and C . The matrix of paired correlation coefficients shows highest negative correlation between parameters \hat{a} and C and then between \hat{a} and pK_a^T . This correlation may be elucidated by a flat shape of the maximum $(1 - U(\beta))$ in Fig. 1. Smaller correlation between pK_a^T and C proves an

Table 1. Regression analysis of simulated data of extended Debye-Hückel equation by DHMINOPT. Data are calculated for $\text{pK}_a^T = 5.0$, $\text{\AA} = 4.5$, $C = 0.3$ and loaded with random error generated for $s_{\text{inst}}(\text{pK}_a) = 0.005$.

(a) Point and interval estimates of parameters with their statistical characteristics. Accuracy is expressed by the relative systematic deviation of each parameter in percent.

Parameter	Estimate	Standard deviation	Half-length of confidence interval		Relative syst. dev. $e_{\text{rel}}(b_j)$, [%]
			Δ_j	$\Delta_{R,j}$	
β_j	b_j	$s(b_j)$			
pK_a^T	4.9973	0.0034	0.0090	0.0106	0.054
\AA	4.5237	0.0790	0.2446	0.2446	−0.527
C	0.2990	0.0024	0.0069	0.0075	0.333

(b) Matrix of partial correlation coefficients of parameters, r_{ij} .

	pK_a^T	\AA	c
pK_a^T	1	−0.844	0.612
\AA	−0.844	1	−0.917
C	0.612	−0.917	1

(c) Goodness-of-fit test by an analysis of residuals and identification of influential points (i)

i	Indep. variab. I	Response measured pK_a	Generated error ε	Classical residual $\hat{\varepsilon}$	Jackknife residual $\hat{\varepsilon}_j$	Likelihood distance LD
1	0.01	4.8646	−0.004744	−2.1330E−03	−6.8887E−01	6.8351E−03
2	0.04	4.7752	−0.000030	2.4165E−03	8.5434E−01	7.5280E−03
3	0.09	4.7019	−0.006361	−4.0980E−03	−9.7670E−01	1.3443E−02
4	0.16	4.6661	0.003561	5.6531E−03	1.2521E+00	2.1715E−02
5	0.25	4.6407	0.006528	8.4750E−03	1.5134E+00	6.9497E−02
6	0.36	4.6145	−0.005988	−4.1547E−03	−9.7966E−01	8.8595E−03
7	0.49	4.6084	−0.011194	−9.4401E−03	−1.5190E+00	1.0674E−01
8	0.64	4.6318	0.001692	3.4004E−03	1.0004E+00	7.4421E−03
9	0.81	4.6484	−0.002600	−9.0466E−04	−3.9492E−01	4.4736E−03
10	1.00	4.6726	−0.008887	−7.1712E−03	−1.3154E+00	4.2531E−02
11	1.21	4.7179	−0.003061	−1.2931E−03	−4.9730E−01	4.6641E−03
12	1.44	4.7769	0.007956	9.8064E−03	1.6213E+00	1.3634E−01
13	1.69	4.8213	−0.003755	−1.7924E−03	−6.0626E−01	4.8549E−03
14	1.96	4.8896	0.000613	2.7163E−03	9.1282E−01	5.6278E−03
15	2.25	4.9522	−0.008288	−6.0162E−03	−1.1970E+00	2.0001E−02
16	2.56	5.0424	0.003049	5.5155E−03	1.2396E+00	1.6851E−02
17	2.89	5.1242	−0.001204	1.4830E−03	7.3327E−01	4.8170E−03
18	3.24	5.2178	−0.000704	2.2302E−03	8.3698E−01	6.2688E−03
19	3.61	5.3129	−0.005627	−2.4221E−03	−7.3291E−01	7.4632E−03
20	4.00	5.4196	−0.005772	−2.2707E−03	−7.1440E−01	7.1671E−03

Goodness-of-fit test

	Residuals
Bias, E($\hat{\epsilon}$)	1.2037E-08
Mean of absolute values of residuals, E $\hat{\epsilon}$	0.0042
Mean abs. values of relative residuals, 100E $\hat{\epsilon}_{rel}$ [%]	0.08711
Variance, s ² ($\hat{\epsilon}$)	2.6021E-05
Standard deviation, s($\hat{\epsilon}$)	0.0051
Skewness, g ₁ ($\hat{\epsilon}$)	0.1527
Kurtosis, g ₂ ($\hat{\epsilon}$)	2.3874
Residuals sum of squares, RSS	4.9439E-04
Regression rabat, 100D ² , [%]	99.957
Akaike information criterion, AIC	-2.0291E+02
Hamilton R-factor, [%]	0.101

independence of these two parameters and corresponds to a well-developed maximum $(1 - U(\beta))$.

Statistical analysis of residuals is a criterion of reliability of parameters estimation. The residual standard deviation $s(\hat{\epsilon})$ is in good agreement with the pre-selected value of instrumental error, $s_{inst}(pK_a) = 0.005$. Sufficiently close fitness is proved by high value of regression rabat $D^2 = 99.95\%$ and low value of Hamilton R-factor = 0.101% expressing a good fit. The normal distribution for residuals is examined with skewness $g_1(\hat{\epsilon})$ and curtosis $g_2(\hat{\epsilon})$. Normal distribution of residuals is proved by low value of skewness $g_1(\hat{\epsilon})$ (for normal distribution $g_1(\hat{\epsilon})$ should be equal to 0), and by curtosis $g_2(\hat{\epsilon})$ (for normal distribution $g_2(\hat{\epsilon})$ should be equal to 3), and for the normality test it is $\chi^2_{exp} < \chi^2_{1-\alpha/2}(20 + 1)$.

Table 2 gives estimation of the thermodynamic dissociation constant pK_a^T , an ion-size parameter \mathring{a} and the salting-out coefficient C for five sulphonephtaleins, i.e. Bromocresol Green (BCG) cf. Fig. 2, Bromophenol Red (BPR), Bromocresol Purple (BCP), Bromothymol Blue (BTB) and Phenol Red (PR). Found estimates are in good agreement with values from literature [3]. The confidence interval of each parameter is calculated by Eq. (3) while literature value uses a threefold standard deviation of each parameter. The four studied sulphonephtaleins have their ion-size parameter within a narrow range 0.760–0.839 nm. The only exception is the ion of the simplest, non-substituted sulphonephtalein, Phenol Red, for which the ion-size 0.361 nm is in agreement with the value 0.40 nm published by Sendroy and Rodkey [15].

Conclusion

Many problems in protonation equilibria can be reduced to the problem of finding a suitable mathematical model and its unknown parameters by minimizing the difference between experimental and calculated data. Variety of regression diagnostics introduced here serves as an efficient tool in search of true model. Besides a regression model of extended Debye-Hueckel equation also other models of activity coefficients, for example, containing an interaction terms between pairs of ions of opposite charge in Debye-Hueckel equation may be applied in DHMINOPT.

Table 2. Determination of thermodynamic dissociation constant, ion-size parameter and salting-out coefficient for five sulphonephthaleins at 25°C. Data taken from Ref. [3] ($A = 0.5115$, $B = 0.3291$, $(1 - z)^2 - z^2 = 3$).

BCG		BPR		BCP		BTB		PR	
Input data									
I	pK _a	I	pK _a	I	pK _a	I	pK _a	I	pK _a
0.010	4.901	0.010	6.017	0.010	6.085	0.0025	7.143	0.039	7.801
0.022	4.871	0.022	5.970	0.022	6.039	0.011	7.075	0.061	7.758
0.040	4.834	0.040	5.935	0.040	6.009	0.024	7.029	0.128	7.677
0.060	4.808	0.060	5.908	0.060	5.986	0.045	6.979	0.261	7.575
0.116	4.765	0.116	5.872	0.116	5.941	0.067	6.954	0.353	7.538
0.232	4.709	0.200	5.841	0.232	5.901	0.127	6.924	0.635	7.465
0.392	4.691	0.392	5.797	0.392	5.871	0.251	6.878	0.789	7.455
0.594	4.677	0.594	5.775	0.594	5.861	0.423	6.844	1.420	7.450
0.920	4.664	1.004	5.760	0.923	5.856	0.635	6.838	2.250	7.465
1.330	4.662	1.445	5.765	1.330	5.863	1.004	6.822	3.610	7.520
2.050	4.686	2.260	5.788	2.000	5.884	1.445	6.826		
3.720	4.785	4.000	5.865	3.720	5.947	2.260	6.843		
						4.000	6.915		
Parameters estimates									
pK _a ^T	5.034 ± 0.014	6.138 ± 0.011		6.202 ± 0.008		7.199 ± 0.012		8.041 ± 0.032	
Liter.	5.033 ± 0.013	6.138 ± 0.010		6.198 ± 0.017		7.199 ± 0.011		8.035 ± 0.018	
Å	7.60 ± 0.73	7.85 ± 0.60		8.39 ± 0.45		7.76 ± 0.66		3.61 ± 0.70	
Liter.	7.59 ± 0.60	7.84 ± 0.51		8.76 ± 0.93		7.44 ± 0.60		3.75 ± 0.38	
C	0.068 ± 0.012	0.055 ± 0.009		0.056 ± 0.006		0.055 ± 0.010		0.107 ± 0.033	
Liter.	0.068 ± 0.010	0.055 ± 0.008		0.055 ± 0.012		0.055 ± 0.010		0.100 ± 0.018	

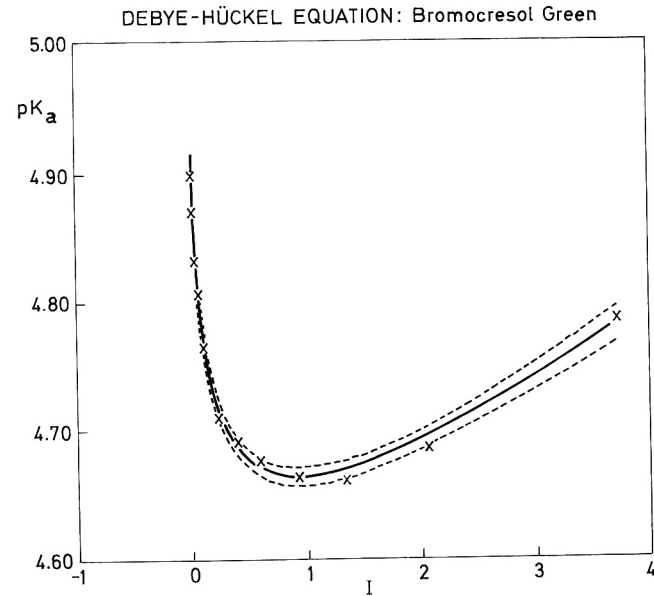


Fig. 2. Curve fit for the dependence of mixed dissociation constant on the ionic strength for Bromocresol Green expressing the extended Debye-Hueckel equation

References

- [1] J. N. Bronsted, *J. Am. Chem. Soc.* **1922**, 44, 877.
- [2] E. A. Guggenheim, *Philos. Mag.* **1935** 19, 588; **1936**, 22, 322; *Discuss. Faraday Soc.* **1957**, 23.
- [3] M. Meloun, S. Kotrlý, *Collect. Czech. Chem. Commun.* **1977**, 42, 2115.
- [4] M. Meloun, J. Čermák, *Talanta* **1984**, 31, 947.
- [5] M. Meloun, J. Havel, E. Högfeltdt, *Computation of Solution Equilibria*, Ellis Horwood, Chichester, 1988, p. 82.
- [6] M. Meloun, J. Havel, in: *Computational Methods for the Determination of Formation Constants* (D. J. Leggett, ed.), Plenum, New York, 1985, p. 284.
- [7] F. James, M. Ross, *Comput. Phys. Commun.* **1976**, 10, 343.
- [8] J. Lang, R. Muller, *Comput. Phys. Commun.* **1971**, 2, 79.
- [9] L. G. Sillen, N. Ingri, *Acta Chem. Scand.* **1962**, 16, 173.
- [10] J. E. Dennis, H. H. W. Mei, *J. Opt. Theor. Appl.* **1979**, 28, 453.
- [11] J. Militký, J. Cap, *Proceedings Conf. CEF'87, Taormina, Sicilia, May 1987*.
- [12] M. Javůrek, *Phil. Thesis*, University of Chemical Technology, Pardubice, 1988.
- [13] L. Schwartz, *Anal. Chim. Acta* **1980**, 122, 291.
- [14] M. Meloun, J. Militký, M. Forina, *Chemometrics in Instrumental Analysis, Part 1 and 2*, Ellis Horwood, Chichester, 1992.
- [15] J. Sendroy, jr., F. L. Rodkey, *Clin. Chem.* **1961**, 7, 646.

Received September 19, 1991. Revision May 25, 1992.