

Prediction of Equilibrium Constants in Aqueous Solution

I. The Extrapolation of Equilibrium Constants to Zero Ionic Strength Using PLS, Artificial Neural Networks, and Genetic “Soft” Modelling

M. FARKOVÁ*, P. LUBAL, and J. HAVEL

*Department of Analytical Chemistry, Faculty of Science, Masaryk University, CZ-611 37 Brno
e-mail: farkova@chemi.muni.cz*

Received 26 May 2003

Extrapolation of formation constants to zero ionic strength using “soft” modelling with partial least-squares, genetic algorithm, and artificial neural networks (ANN) methods was examined and results of individual approaches were compared. The methods allow a rapid and sufficiently accurate prediction of thermodynamic formation constants, ion-size parameters, and salting-out coefficients from experimental equilibrium data, among them the ANN method was found most reliable.

The knowledge of stability constants is important in all areas of chemistry, chemical technology, environment, *etc.*, ranging from the need to understand speciation when developing analytical procedures, speciation of metal-ion complexes in the environment [1] for modelling processes in nuclear waste, biological liquids, *etc.* However, the measurement of a large number of stability constants under varying chemical conditions (ionic strength) is not an attractive option. Therefore accurate and reliable methods for the prediction of stability constants are desirable [2].

There are several options to solve the task. Quite a general one is to predict the thermodynamic equilibrium constants from fundamental properties. This treatment allows the prediction of the equilibrium constant of any metal-ion complex and is extremely useful in obtaining constants for those complexes which cannot be studied experimentally [3].

The other possibility is extrapolation of the equilibrium constants from data estimated at several ionic strength values to zero. The thermodynamic equilibrium constant pK^T is then obtained together with other parameters. There exist several methods relating equilibrium constants to ionic strength. Usually, various types of extended Debye–Hückel, Pitzer or Bromley equations are used. The calculations are performed by the least-squares computer programs [4, 5], using “hard” models, *i.e.* using exactly defined equation with several parameters. A series expansion of the extended Debye–Hückel equation for prediction of stability constants was used by *Baeza* [6]. However, the functional ionic strength dependence chosen for

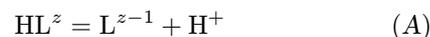
the extrapolation of the equilibrium constants is questionable. The problem of selection of functions for the extrapolation to zero ionic strength is described in detail by *Anderegg* [7].

Contrary to the “hard” models described in the literature [6, 7], the aim of this paper is to examine the possibility of prediction (extrapolation) of equilibrium constants to zero ionic strength with the use of “soft” modelling, like partial least-squares (PLS), genetic algorithm (GA), and artificial neural networks (ANN) methods. For the sake of simplicity this study will be limited only to the extended Debye–Hückel equation.

The ANN method has been successfully applied in differential pulse polarography to evaluate equilibria for fully inert [8] or labile [9] metal complexes. Recently it was shown that the ANN can be severally used for the chemical equilibria computation for any type of data [10].

THEORETICAL

The equations used for the extrapolation of equilibrium constants to zero ionic strength are often based on Debye–Hückel equation [7, 11]. In the case of a weak acid dissociation



the ionic strength (I) dependence of $-\log K_a$ ($= pK_a$) of a dissociation constant may be expressed by the extended Debye–Hückel equation

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

*The author to whom the correspondence should be addressed.

assuming that the ion-size parameters α for both ions HL^z and L^{z-1} are approximately equal, and the overall salting-out coefficient, C , is defined by $C = C_{\text{HL}} - C_{\text{L}}$. The value of constants A and B is $0.5115 \text{ mol}^{-1/2} \text{ dm}^{3/2}$ and $0.3291 \text{ m}^{10} \text{ mol}^{-1/2} \text{ dm}^{3/2}$, respectively, for aqueous solutions at 25°C , z is the charge of HL.

An equation similar to eqn (1) can be written for equilibrium constant of a complexation reaction



with the stability constant

$$\beta = \frac{[\text{ML}]}{[\text{M}][\text{L}]} \quad (2)$$

A more general equation is

$$\log \beta = \log \beta_0 - \frac{A' \Delta z^2 \sqrt{I}}{1 + B' \alpha \sqrt{I}} - CI \quad (3)$$

where $\Delta z^2 = (m - z)^2 - m^2 - (-z)^2$, A' and B' are the parameters of the solvent (constant for given experimental conditions, *e.g.* pressure and temperature), $\log \beta_0$, α , and C are adjustable parameters related to the given experimental system.

COMPUTATIONAL METHODS

There are principally two approaches, the use of “hard” and “soft” modelling. In the so-called “hard” model eqn (1) and a general least-squares approach are used, in the “soft” modelling eqn (1) is not used explicitly.

For example, thermodynamic dissociation constants at $I = 0$ may be estimated from the data of a series of $\text{p}K_{\text{a}}$ values at several different ionic strengths by a nonlinear least-squares analysis. This is the most commonly used approach belonging to the “hard” model. An example of this approach is DHLET program [4, 5], where Debye—Hückel equation is applied. On the other hand, LIANA program used by Foti *et al.* [12] applies the Pitzer and Bromley equations.

The independent variable is I and the dependent variable is measured $\text{p}K_{\text{a}}$. The parameters which are estimated by the general least-squares approach are $\text{p}K_{\text{a}}^{\text{T}}$, α , and C . The minimization function is

$$U = \sum_{i=1}^N w_i (\text{p}K_{\text{a,calc},i} - \text{p}K_{\text{a,exp},i})^2 \quad (4)$$

where w_i is the weighting factor, $\text{p}K_{\text{a,calc},i}$ are the values calculated according to eqn (1) and $\text{p}K_{\text{a,exp},i}$ are experimental values. The summation is defined for all N experimental points.

Principles of the “soft” modelling approaches are briefly given in the following paragraphs.

PLS is the name for a class of methods used for relating the blocks of variables measured on sets of objects. PLS is a multivariate regression method that provides an overview of large data sets. Modelling with PLS has a large potential as a method of data analysis in many areas of science.

PLS is used to compress the predictor data matrix $\mathbf{X} = [x_1, x_2, \dots, x_p]$, that contains the values of p predictors for n samples, into a set of D latent variables or factor scores $\mathbf{T} = [t_1, t_2, \dots, t_D]$, where $D \leq p$. The factors t_d , $d = 1, 2, \dots, D$, are determined sequentially using the nonlinear iterative partial least-squares (NIPALS) algorithm. The orthogonal factor scores are used to fit a set of n observations to m dependent variables $\mathbf{Y} = [y_1, y_2, \dots, y_m]$ [13].

Data in this work were processed on a Pentium IV PC computer. PLS method was executed using the PLS2 algorithm as implemented in the software package Unscrambler [14].

Principles of biological evolution and selection were founded by Charles Darwin and the history of this discovery is described in [15]. GA is a search algorithm based in some way on the rules of biological evolution. Several centuries afterwards a computer algorithm has been developed, the principles of which are described in papers [16–18].

The search for optimal parameters is initialized with a random population of possible solutions. Each member of the population representing a candidate solution is tested against some criteria and the members of the population are ranked according to their “fitness”. Competition is the basis of the natural selection: nonaccommodated members are rejected, and accommodated members are reproduced, creating offsprings that take the place of rejected members. Reproduction is the mix of the characteristics of the pairs of good members to produce new members that possess some characteristics of one parent, and other characteristics of the other parent. The second way of changing the population is called mutation. Mutation randomly chooses a member of the population and randomly changes some of its characteristics. Fit solutions are allowed to live and breed while unfit solutions die. An iteration is performed until either the populations or quality of the solutions converge [16].

In GA, the optimized variables ($x_1, x_2, x_3, \dots, x_n$) are represented as the genes of a chromosome. A collection of values of the variables to be optimized is called a chromosome, and the variables themselves are called genes [17]. The fitness of the chromosome is determined by computing the response function score ($\text{fitness} = f(x)$).

GA algorithm consists of four basic steps.

i) The initial population of chromosomes is created either randomly or by random perturbation of an input chromosome. The population size N_{p} , a user-controlled option, remains constant throughout the optimization.

ii) The fitness of each chromosome in the population is computed.

iii) The third step is the exploitation, or the natural selection step. The chromosomes with the largest fitness scores are placed one or more times into a mating subset in a semirandom fashion. Chromosomes not selected for the mating subset are removed from the population. The chromosome with the better fitness score is placed in the mating subset, and both chromosomes are returned to the gene pool. This process continues until the mating subset is full. This method gives the chromosomes with higher fitness scores a higher probability of being included in the mating subset than the chromosomes with lower fitness scores.

iv) The fourth step, exploration, consists of applying the recombination and mutation operators. Two chromosomes (parents) from the mating subset are randomly selected to be mated. If the parents are allowed to mate, a recombination operator is used to exchange genes between the two parents to produce two children. If they are not allowed to mate, the parents are placed into the next generation unchanged.

Following recombination, a mutation operator is used to increase the diversity in the population.

After the exploration step, the population consists of the newly created chromosomes, and the steps two through four are repeated. This process continues for a user-selected number of generations or until a certain termination criterion is reached (*e.g.* most chromosomes in the population are the same) [18].

Data in this work were processed using GA as implemented in the Pascal program written by *Kvasnička*, Department of Mathematics, Slovak University of Technology, Bratislava, Slovakia.

ANN are mathematical systems that simulate biological neural networks. The description of ANN can be found in several monographs [19–21] and in papers [8–10] and thus only a brief description will be given here.

The fundamental processing element of an ANN is the node (an analogue of a neuron in biological systems). The nodes are arranged in layers that make up the global network architecture. The networks comprise three or more node layers:

1. a layer of input units,
2. one or more hidden layers,
3. a layer of output units.

The nodes are connected and the weights are assigned to each connection. The values of the weights are modified in the course of the network operation. Each unit has an activity level that is determined by an input signal received from the other units in the network. Neural networks are trained by repeatedly presenting examples.

There are many different neural network architectures, but one of the most common is the feedforward neural network. During the learning stage of these

nets, the outputs come to approximate the target values given by the inputs in the training set. In the second stage, the results of training are to be proved on a test set.

The data in this work (first cases) were processed using Back Propagation Neural Networks (BPNN), applying the generalized delta rule, as included in the Parallel Distributed Processing (PDP) software package [20], while the data in the last case were processed using Trajan 3.0 software package [22].

RESULTS AND DISCUSSION

We will examine several examples concerning prediction of equilibria constants to zero ionic strength from experimental values obtained for a series of different ionic strengths with the use of PLS, GA, and ANN methods.

Initially we dealt with prediction of the thermodynamic dissociation constant of a weak acid. In this case we have first applied simulated data, with no experimental noise superimposed. Using extended Debye–Hückel eqn (1) twelve values of a weak acid dissociation constant, with $\text{p}K_{\text{a}}^{\text{T}} = 6.5$, $\alpha/\text{m}^{-10} = 8.0$, $C/(\text{mol}^{-1} \text{dm}^3) = 0.053$ and for $I/(\text{mol dm}^{-3}) = 0.01, 0.02, 0.04, 0.06, 0.1, 0.2, 0.4, 0.6, 1, 2, 3,$ and 5 were calculated. Then the thermodynamic equilibrium constant was estimated from these data (12 points) with the use of PLS, GA, and ANN methods.

However, in order to perform PLS or ANN prediction, one needs a calibration set. Therefore for the estimation of the values $\text{p}K_{\text{a}}^{\text{T}}$, α , and C by the PLS and ANN methods using eqn (1) for the same ionic strength values, the calibration set represented by 64 curves was computed for $\text{p}K_{\text{a}}^{\text{T}} = 6.40, 6.45, 6.50,$ and 6.55 , $\alpha/\text{m}^{-10} = 7.4, 7.8, 8.2,$ and 8.6 , $C/(\text{mol}^{-1} \text{dm}^3) = 0.045, 0.050, 0.055,$ and 0.060 (Fig. 1). These 64 curves represent a calibration set where, for each curve, twelve $\text{p}K_{\text{a}}$ values are given in input and $\text{p}K_{\text{a}}^{\text{T}}$, α , and C values representing output data were used. A PLS model with 12 latent variables was applied.

Similarly, these 64 curves were used as a training set for ANN. A feedforward network was constructed by using three-layer ANN architecture and a systematic study was made for the different neural parameters. Output values were scaled so as to be between 0 and 1 (the range of the sigmoidal transfer function outputs), and the normalized outputs values were used for the output nodes. The number of hidden nodes is an adjustable parameter and its optimal value was searched. Together with the number of nodes in the hidden layer another adjustable parameter was learning rate (*lr*ate). The ANN structure with 12 input, 3 hidden, and 3 output artificial neurons (Fig. 2) approximated the data quite well and thus this structure was used for the prediction.

The obtained thermodynamic dissociation constants, ion-size parameters, and salting-out coeffi-

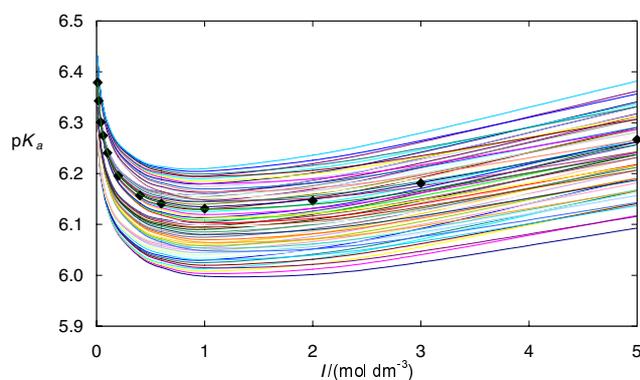


Fig. 1. Dependence of the pK_a values (\blacklozenge) of the weak acid on ionic strength ($pK_a^T = 6.5$, $\alpha/m^{-10} = 8.0$, $C/(\text{mol}^{-1} \text{dm}^3) = 0.053$, $I/(\text{mol} \text{dm}^{-3}) = 0.01, 0.02, 0.04, 0.06, 0.1, 0.2, 0.4, 0.6, 1, 2, 3, 5$) and simulated curves ($pK_a^T = 6.40, 6.45, 6.50$, and 6.55 , $\alpha/m^{-10} = 7.4, 7.8, 8.2$, and 8.6 , $C/(\text{mol}^{-1} \text{dm}^3) = 0.045, 0.050, 0.055$, and 0.060) for the estimation of thermodynamic equilibria constants.

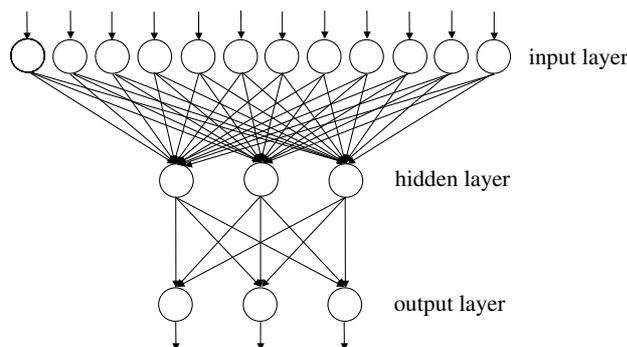


Fig. 2. An example of the ANN architecture (12:3:3) used in some cases of this work.

icients as obtained using different approaches are given in Table 1. The results are quite satisfactory, the lowest deviation from the correct values of parameters is observed for the ANN method.

Two levels of noise in the pK_a values were examined. To 12 values of dissociation constants calculated in the previous case the noise with normal distribution and with the mean value 0 and the standard deviation

0.01 (0.02) was superimposed. Then we continued as in the foregoing case.

The thermodynamic dissociation constants, ion-size parameters, and salting-out coefficients estimated by GA are given in Table 1. For PLS we used the same calibration set and model as in the previous case, and for the ANN method the same training set and ANN architecture were employed as in the case without noise. Predicted values and comparison of all three methods are given in Table 1. The results reported in Table 1 show also in this case quite satisfactory agreement with the correct pK_a^T , α , and C values.

Dissociation constants of bromthymol blue were determined [4] for the different ionic strength values. From these experimental data the thermodynamic equilibrium constants and Debye–Hückel equation parameters were determined with the use of the GA, PLS, and ANN methods.

First the use of GA was examined. For the estimation of the pK_a^T , α , and C values by PLS and ANN methods 64 curves were simulated using the extended Debye–Hückel equation for the same ionic strength values, where $pK_a^T = 7.10, 7.15, 7.20$, and 7.25 , $\alpha/m^{-10} = 7.0, 7.5, 8.0$, and 8.5 , $C/(\text{mol}^{-1} \text{dm}^3) = 0.050, 0.055, 0.060$, and 0.065 (Fig. 3).

For PLS these 64 curves were used as a calibration set, for every curve 13 pK_a values as input and pK_a^T , α , and C values as output data. A PLS model with 13 latent variables was applied. In the prediction phase we took as the input data 13 values of dissociation constants from literature and the pK_a^T , α , and C values were predicted as the output data.

For the ANN method these 64 curves were used as a training set, for every curve 13 pK_a values as input and pK_a^T , α , and C values as output data. A feed-forward network was constructed using a three-layer ANN architecture. The output values were scaled to be between 0 and 1, and the normalized output values were used for the output nodes. The ANN structure with 13 input, 4 hidden, and 3 output artificial neurons approximated the data quite well, therefore this structure was then used for the prediction. In the prediction phase, we took as the input data 13 values of dissociation constants from [4] and the output data –

Table 1. Comparison of Results of Different Prediction Approaches for the Determination of Debye–Hückel Equation Parameters of the Weak Acid

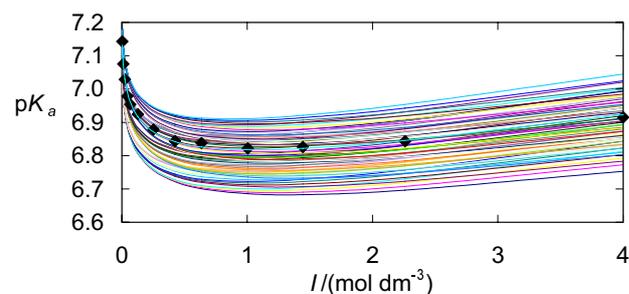
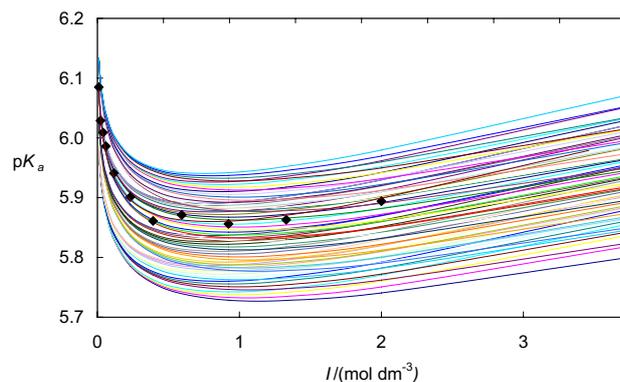
Parameter	Correct value	Method								
		PLS Noise s			GA Noise s			ANN Noise s		
		0	0.01	0.02	0	0.01	0.02	0	0.01	0.02
pK_a^T	6.500	6.500	6.498	6.496	6.450	6.498	6.496	6.502	6.499	6.496
α/m^{-10}	8.000	8.031	8.310	8.608	8.009	8.297	8.619	8.025	8.339	8.607
$C/(\text{mol}^{-1} \text{dm}^3)$	0.053	0.053	0.053	0.052	0.053	0.053	0.052	0.053	0.053	0.052

Table 2. Comparison of Results of Different Prediction Approaches for the Determination of Debye–Hückel Equation Parameters for the Bromthymol Blue

Parameter	Method			
	DHLET	PLS	GA	ANN
pK_a^T	7.199 ± 0.004	7.199	7.199	7.195
α/m^{-10}	7.763 ± 0.209	7.783	7.753	7.790
$C/(\text{mol}^{-1} \text{ dm}^3)$	0.054 ± 0.003	0.055	0.055	0.054

Table 3. Comparison of Results of Different Prediction Approaches for the Determination of Debye–Hückel Equation Parameters for the Bromcresol Purple

Parameter	Method			
	DHLET	PLS	GA	ANN
pK_a^T	6.197 ± 0.006	6.197	6.197	6.197
α/m^{-10}	8.807 ± 0.328	8.830	8.812	8.828
$C/(\text{mol}^{-1} \text{ dm}^3)$	0.055 ± 0.004	0.055	0.055	0.055

**Fig. 3.** Dependence of the pK_a values (◆) of bromthymol blue on ionic strength and simulated curves ($pK_a^T = 7.10, 7.15, 7.20, \text{ and } 7.25, \alpha/m^{-10} = 7.0, 7.5, 8.0, \text{ and } 8.5, C/(\text{mol}^{-1} \text{ dm}^3) = 0.050, 0.055, 0.060, \text{ and } 0.065$) for the estimation of thermodynamic equilibria constants.**Fig. 4.** Dependence of the pK_a values (◆) of bromcresol purple on ionic strength and simulated curves ($pK_a^T = 6.10, 6.15, 6.20, \text{ and } 6.25, \alpha/m^{-10} = 8.0, 8.5, 9.0, \text{ and } 9.5, C/(\text{mol}^{-1} \text{ m}^3) = 0.050, 0.055, 0.060, \text{ and } 0.065$) for the estimation of thermodynamic equilibria constants.

pK_a^T , α , and C values – were then predicted. Thermodynamic dissociation constants, ion-size parameters, and salting-out coefficients estimated using GA, PLS, ANN, and general least-squares program are given in Table 2. All obtained results are acceptable.

Further 12 values of dissociation constants of bromcresol purple [4] were used as input data. The thermodynamic dissociation constant, ion-size parameter, and salting-out coefficient were determined by the GA, PLS, and ANN methods.

For the estimation of the values pK_a^T , α , and C by the PLS and ANN methods we used eqn (1) and simulated 64 curves with $pK_a^T = 6.10, 6.15, 6.20, \text{ and } 6.25, \alpha/m^{-10} = 8.0, 8.5, 9.0, \text{ and } 9.5, C/(\text{mol}^{-1} \text{ dm}^3) = 0.050, 0.055, 0.060, \text{ and } 0.065$ (Fig. 4).

64 curves represent a calibration set, while for each curve twelve pK_a values are given as the input and pK_a^T , α , and C values as the output data are used. A PLS model with 12 latent variables was applied. In the prediction phase we took as the input data 12 values

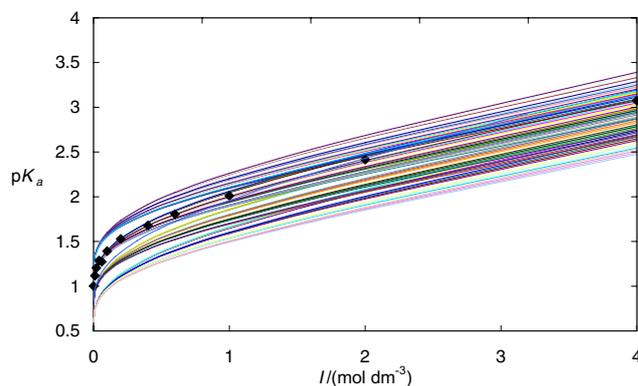
of dissociation constants from the literature [4].

For the ANN method these 64 curves were used as a training set, for every curve 12 pK_a values as input and pK_a^T , α , and C values as the output data. A feed-forward network was constructed using the three-layer ANN architecture. Output values were scaled to be between 0 and 1, and the normalized output values were used for the output nodes. The ANN structure with 12 input, 3 hidden, and 3 output artificial neurons (Fig. 2) approximated data very well and was used for prediction. In the prediction phase the values of dissociation constants from [4] were used as the input data. The predicted pK_a^T , α , and C values as obtained with GA, PLS, and ANN are given in Table 3 and the comparison with the results of the “hard” model using DHLET program [4] shows a good agreement.

In the next stage using the extended Debye–Hückel equation 12 values of dissociation constants

Table 4. Comparison of Results of Different Prediction Approaches for the Determination of Debye–Hückel Equation Parameters of Pd(II) under the Formation of [Pd(OH)⁺] Complex

Parameter	Correct value	Method		
		PLS	GA	ANN
pK_a^T	0.997	0.997	0.100	1.019
α/m^{-10}	3.530	3.267	3.267	3.500
$C/(\text{mol}^{-1} \text{ dm}^3)$	0.283	0.275	0.276	0.275

**Fig. 5.** Dependence of the pK_a values (\blacklozenge) of the [Pd(OH)⁺] complex on ionic strength ($pK_a^T = 0.997$, $\alpha/m^{-10} = 3.53$, $C/(\text{mol}^{-1} \text{ dm}^3) = 0.283$, $I/(\text{mol dm}^{-3}) = 0, 0.01, 0.02, 0.04, 0.06, 0.1, 0.2, 0.4, 0.6, 1, 2, 4$) and simulated curves ($pK_a^T = 0.6, 0.8, 1.0$, and 1.2 , $\alpha/m^{-10} = 3.0, 3.5, 4.0$, and 4.5 , $C/(\text{mol}^{-1} \text{ dm}^3) = 0.275, 0.280, 0.285$, and 0.290) for the estimation of thermodynamic equilibria constants.

of [Pd(OH)⁺] complex with $pK_a^T = 0.997$, $\alpha/m^{-10} = 3.53$, $C/(\text{mol}^{-1} \text{ dm}^3) = 0.283$ were calculated for ionic strength in the range 0–4. The values of parameters were estimated with respect to the values from [23] and the normally distributed noise with the mean value of 0 and the standard deviation of 0.02 was superimposed. Thermodynamic equilibrium constant as well as α and C parameters were then determined using the GA, PLS, and ANN methods.

For the estimation of the values pK_a^T , α , and C by the PLS and ANN methods 64 curves were simulated using eqn (1), where $pK_a^T = 0.6, 0.8, 1.0$, and 1.2 , $\alpha/m^{-10} = 3.0, 3.5, 4.0$, and 4.5 , $C/(\text{mol}^{-1} \text{ dm}^3) = 0.275, 0.280, 0.285$, and 0.290 (Fig. 5).

For PLS these 64 curves serve as a calibration set, for every curve 12 pK_a values as the input and pK_a^T , α , and C values as the output data were used. A PLS model with 12 latent variables was applied. In the prediction phase we took 12 calculated values of dissociation constants as the input data.

For the ANN method these 64 curves serve as a training set, for every curve 12 pK_a values as input and pK_a^T , α , and C values as output data were used. A feedforward network was constructed by using a three-layer ANN architecture. The output values were scaled

to be between 0 and 1 and were used for the output nodes. The ANN structure with 12 input, 3 hidden, and 3 output artificial neurons (Fig. 2) approximated data very well and was used for prediction. In the prediction phase we took 12 calculated values of dissociation constants as the input data. Estimated thermodynamic dissociation constant, ion-size parameter, and salting-out coefficient obtained by different methods are given in Table 4. The results of all methods are satisfactory, the best agreement between experimental and predicted values was obtained by the ANN method.

Dissociation constants of water in lithium chloride were determined for the LiCl concentrations in the range 0.1–3 mol dm⁻³ [6].

The calibration set was again prepared in a similar way as in the previous cases using the extended Debye–Hückel equation. 64 curves were computed for $pK_a^T = 14.06, 14.07, 14.08$, and 14.09 , $\alpha/m^{-10} = 5.360, 5.365, 5.370$, and 5.375 , $C/(\text{mol}^{-1} \text{ dm}^3) = 0.080, 0.081, 0.082$, and 0.083 .

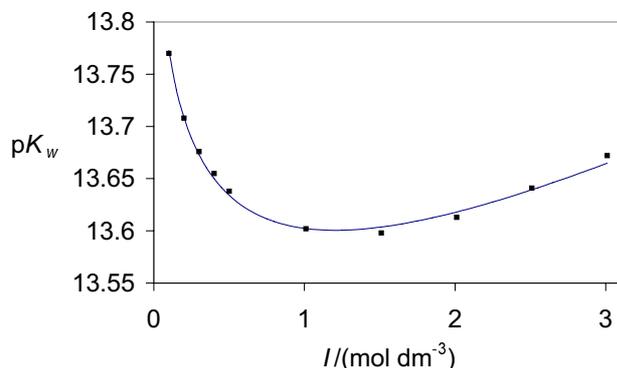
These 64 curves were used as the training set; for each curve 10 pK_w values were used as the input, pK_a^T , α , and C values were defined as the output data. From these data (10 values) thermodynamic dissociation constant of water and other parameters were determined with the use of GA and ANN method as well as the DHLET program. For the ANN a feedforward network was applied. The number of hidden nodes was searched automatically by the automatic network designer. Trajan's automatic network designer [22] determined a suitable architecture using a combination of heuristics and sophisticated optimization strategies and the obtained ANN structure (with 10 input, 4 hidden, and 3 output artificial neurons) approximated data quite well so that the results of the training were then used for the prediction. In the prediction phase, 10 pK_w values from [6] were used at the input and the values pK_a^T , α , and C were predicted (Table 5). Fig. 6 shows a curve obtained using these values. The results obtained by the DHLET program are also given for comparison in Table 5 and are in a good agreement.

CONCLUSION

Prediction of thermodynamic equilibrium constants for $I \rightarrow 0$ and Debye–Hückel parameters us-

Table 5. Comparison of Results of Different Prediction Approaches for the Determination of Debye–Hückel Equation Parameters of Water in Lithium Chloride Aqueous Solution

Parameter	Method		
	PLS	GA	ANN
pK_a^T	14.076 ± 0.015	14.079	14.076
α/m^{-10}	5.369 ± 0.032	5.268	5.368
$C/(\text{mol}^{-1} \text{ dm}^3)$	0.081 ± 0.010	0.084	0.081

**Fig. 6.** Plot of the pK_w values in lithium chloride and a comparison with the ANN model (solid line) with $pK_a^T = 14.076$, $\alpha/m^{-10} = 5.368$, and $C/(\text{mol}^{-1} \text{ dm}^3) = 0.081$; ■ experimental values.

ing the “soft” modelling with PLS, GA, and ANN yields the results which are in a good agreement with the “hard” models where the explicit equation and the general least-squares approach were applied. The “soft” modelling methods allow to predict the thermodynamic dissociation constants, ion-size parameters, and salting-out coefficients either from a series of experimental or the simulated data sets rapidly and with the sufficient accuracy.

The advantage of all “soft” modelling methods is that the explicit model is not needed and after appropriate learning the calculations are almost instantaneous.

Acknowledgements. Support from the Grant Agency of the Czech Republic, Grant No. 203/02/1103, is acknowledged.

REFERENCES

- Hancock, R. D., *Analyst* 122, 51R (1997).
- Dimmock, P. W., Warwick, P., and Robbins, R. A., *Analyst* 120, 2159 (1995).
- Brown, P. L., *Talanta* 36, 351 (1989).
- Havel, J. and Meloun, M., in *Computational Methods for the Determination of Formation Constants*, Vol. 27. (Legett, D. J., Editor.) P. 284. Plenum Press, New York and London, 1985.
- Meloun, M. and Kotrlý, S., *Collect. Czech. Chem. Commun.* 42, 2115 (1977).
- Baeza, J. J. and Ramis-Ramos, G., *Talanta* 43, 1579 (1996).
- Anderegg, G. and Kholeif, S., *Talanta* 41, 1507 (1994).
- Cukrowski, I. and Havel, J., *Electroanalysis* 12, 1481 (2000).
- Cukrowski, I., Farková, M., and Havel, J., *Electroanalysis* 13, 295 (2001).
- Havel, J., Lubal, P., and Farková, M., *Polyhedron* 21, 1375 (2002).
- Linder, P. W. and Murray, K., *Talanta* 29, 377 (1982).
- Foti, C., Sammartano, S., and Signorino, G., *Fluid Phase Equilib.* 149, 91 (1998).
- Geladi, P., *J. Chemom.* 2, 231 (1988).
- Unscrambler v 5.53*, Camo AS, Trondheim, Norway, 1993.
- Ilakovac, T., *CCACAA* 68, 29 (1995).
- Wise, B. M., Holt, B. R., Gallagher, N. B., and Lee, S., *Chemom. Intell. Lab. Syst.* 30, 81 (1995).
- Bangalore, A. S., Shaffer, R. E., and Small, G. W., *Anal. Chem.* 68, 4200 (1996).
- Shaffer, R. E. and Small, G. W., *Anal. Chem. (News & Features)* 69, 236A (1997).
- Gasteiger, J. and Zupan, J., *Angew. Chem.* 32, 503 (1993).
- McClelland, J. L. and Rumelhart, D. E., *Explorations in Parallel Distributed Processing, A Bradford Book*. MIT Press, Cambridge, England, 1988.
- Kvasnička, V., Beňušková, L., Pospíchal, J., Farkaš, I., Tiňo, P., and Král, A., *Introduction to Theory of Neural Networks*. Publisher IRIS, Bratislava, 1997.
- Trajan Neural Network Simulator, Release 3.0 D*. Copyright Trajan Software Ltd., 1996–1998.
- Kotrlý, S. and Šucha, L., *Chemical Equilibria in Analytical Chemistry*. Nakladatelství technické literatury (Publishers of Technical Literature), Prague, 1988.