ELSEVIER

Journal of Chromatography A, 793 (1998) 317–329

JOURNAL OF CHROMATOGRAPHY A

Neural networks for optimization of high-performance capillary zone electrophoresis methods
A new method using a combination of experimental design and artificial neural networks[1]

J. Havel[a,*], E.M. Peña[a,2], A. Rojas-Hernández[a,3], J.-P. Doucet[b], A. Panaye[b]

[a]*Department of Analytical Chemistry, Faculty of Science, Masaryk University, Kotlářská 2, 611 37 Brno, Czech Republic*
[b]*Institute de Topologie et de Dynamique des Systèmes, associé au CNRS URA 34, Université Paris 7, 1 Rue Guy de la Brosse, 7005 Paris, France*

Received 18 September 1996; received in revised form 20 August 1997; accepted 3 September 1997

Abstract

Artificial neural networks (ANN) offer attractive possibilities for providing non-linear modeling of response surfaces and optimization in capillary zone electrophoresis (CZE) when the underlying mechanisms are very complex or not well known or understood, in comparison with (non)-linear regression methods. The application of ANN in optimization of CZE methods has been examined and a new method, based on the combination of experimental design and ANN methods, which offers considerable effectiveness, has been developed. © 1998 Elsevier Science B.V.

Keywords: Neural networks, artificial; Experimental design; Chemometrics; Optimization

1. Introduction

Capillary electrophoresis (CE) was developed during the last decade as a powerful separation technique. CE in narrow-bore tubing was used as early as almost 30 years ago by Hjertén [1] and is now the most important method in analytical chemistry, being also applied in genetic analysis, largely driven by the Human Genome Project.

In spite of many published monographs [2–5] a complete and satisfactory theory describing the separation process in CE is missing. Separation of analytes depends on: (a) chemical reactions, (i.e., on pH via protonation/dissociation equilibria), buffer concentration, complexation/interaction reactions of solutes, the formation of inclusion complexes, host–guest interaction, interaction of analytes and/or the other solutes with capillary walls (changing zeta potential), sieving effect, (b) physical conditions (separation voltage, temperature, ionic strength), (c) technical features (ways of injection, conditioning

*Corresponding author.

[1] Presented at the 10th International Symposium on Capillary Electrophoresis and Isotachophoresis, Prague, 17–20 September, 1996.

[2] On leave from Department of Analytical Chemistry, Nutrition and Food Science, University of La Laguna, La Laguna, Tenerife, Spain.

[3] On leave from Department of Chemistry, UAM-Iztapalapa and FESC-UNAM, Mexico D.F., Mexico.

0021-9673/98/$19.00
PII S0021-9673(97)00918-7

procedures of the capillary tubes, etc.) and other factors.

The optimization of separations performed with CE is sometimes complex and difficult due to the high number of parameters affecting the separation. Further complications can arise from the mutual interaction of the parameters to be optimized. Corstjens et al. [6] have recently given an overview on statistical approaches in the optimization of separations in CE and on optimization procedures based on physicochemical models. All procedures have in common the fact that they provide guidelines to achieve an adequate selectivity with the minimum number of experiments. Pre-selection of parameters and the parameter space to be optimized, a model or algorithm to describe the migration behavior of the solutes and a criterion to evaluate the resulting chromatograms are indispensable for this purpose [7].

Computer simulations of electrophorectic separations have been described in several textbooks [8,9] and a number of papers published in recent years [10,11] where different algorithms for transient modeling of CE have been described. Programs based on such algorithms were applied to various modes of CE. Although the basic differential equations governing such separations are well defined, it appears from the literature that different numerical approaches for solving the equations can be chosen. There are several methods available, which are limited in both applicability of practical conditions and calculation speed. By using several alternate numerical treatment, the limitations were dealt with. Another (minor) point that the existing programs have in common is the lack of a practical user-interface, which makes it difficult for other people to use the software as if it were a word processor [12].

Many parameters are to be optimized when developing a CE method. Conventionally, a 'step-by-step' approach is used, which however involves a large number of independent analyses, and could be with advantage, replaced by statistically designed experimental protocols in which several factors are simultaneously varied. This multivariate approach has advantages in terms of a reduction in the number of experiments, improved statistical interpretation possibilities and reduced overall analysis time requirements. The use of chemometric tools like experimental design can be extremely beneficial in the evaluation of capillary electrophoresis methods [13,17]. Altria et al. [18] and Small et al. [19] concluded that the use of experimental designs and statistical data evaluation in conjunction with personal computer-controlled CE autosamplers and instruments are of great benefit in the optimization and robustness of evaluation of CE methods.

Central composite design (CCD) is used for systematic optimization and it offers an efficient route for rapid optimization of resolution with multiple interacting parameters in CE. The utility of response surface methodology for locating optimum conditions in CE is available in the literature [18,19].

On the other hand, artificial neural networks (ANNs) lately became a very powerful and practical method solving various problems in chemistry [20–22]. Recently, they were applied, e.g., to kinetics [23,24]. In spite of principal differences, CE and kinetics are in some way similar to each other as in both techniques the signal (absorbance, fluorescence) is, of course, time dependent. Also experimental variables such as temperature, ionic strength, pH, etc. determine the development of the process.

Paralleled distributed processing (PDP) provides a new way of thinking about perception, memory, learning and thinking about basic computational mechanisms for intelligent information processing in general. These new ways of thinking have all been captured in simulation models [25].

A network of artificial neurons, usually called ANN, is a data processing system consisting of a large number of simple, highly interconnected processing elements in an architecture inspired by the structure of the brain. The characteristics that make ANN systems different from traditional computing and artificial intelligence are: (a) learning by example, (b) distributed associative memory, (c) fault tolerance and (d) pattern recognition.

ANNs in chemistry are still in a pioneering phase. Software simulation is still far from the performances required to compete with human brain capabilities but they still offer, today, very promising avenues, particularly thanks to their ability of mapping non-linear data and extracting complex relationships when the underlying mechanisms are up to now not well known [26]. The main chemical applications of ANNs, the theory of different net-

works and the type of learning scheme for their use have been recently reviewed by Zupan and Gasteiger [21]. ANNs can be applied to various tasks of information processing: classification, modeling, association and mapping. We highlight below only the most important aspects of the network used further in this work.

Because a complete and quite general physico–chemical model in capillary zone electrophoresis (CZE) is still missing, we have examined the possibility of using ANNs for modeling in CE. In contrast with other statistical tools, neural networks models have the advantage that 'no prior knowledge' is needed about the relations among the variables. The aim of ANN is to let the computer itself extract the rules by presenting the same facts many times.

Thus, this work examines the performance of ANNs and combined CCD and ANNs to model response surfaces in CZE, with the aim of using it for the optimization of separation and thus to obtain experimental conditions. Some results of this work were presented at Chemometrics 'IV [27] and ITP'96 conferences [28].

2. ANN theory

The application of ANNs to data are claimed to constitute so called 'soft models', i.e. the models are able to represent the experimental behavior of the system when the exact description is missing or is too complex. As for the 'hard' and 'soft' models in chemistry we can refer to the monograph by Massart et al. [29]. Neural networks consist of arrays of simple activation units linked by weighted connections (see Fig. 1). The basic processing unit in ANNs is a node or a simulated neuron. ANNs can consist of multiple layers of neurons arranged so that each neuron in one layer is connected with each neuron in the next layer.

The networks used here are multilayered feedforward neural networks using as a learning scheme, the algorithm of the back-propagation (BP) of errors and the generalized 'delta rule' [30] for the adjustment of the connection weights (we shall further call these networks, BP networks).

That is, in contrast to so called 'hard models' in chemistry, where we do have formulas, equations

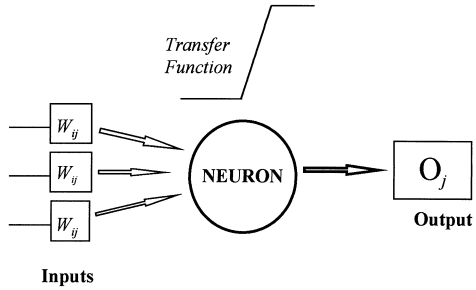


Fig. 1. Schematic architecture of the basic processing unit in an artificial neural network—the node.

and the system is exactly described by these equations and the values of physico–chemical constants. A BP network is composed of a collection of elementary units (nodes or neurons) strongly interconnected. Such nodes were developed from speculation of the actual brain neuron activity (see Fig. 1) and constitute the fundamental processing elements of an ANN [31,32]. Nodes are arranged in layers that make up the global architecture. BP networks comprise one input layer, one (possibly several) hidden layer and an output layer. The number of nodes in the input layer and the output layer are defined by the problem being solved. The input layer receives the experimental information, experimental parameters, structural descriptors, etc., the output layer delivers the response sought for property value, classification, etc. As for the hidden layer, it encodes and organizes the information received from the input layer and delivers it to the output layer.

The number of nodes in the hidden layer, which somewhat play the role of intermediate variables, may be considered as an adjustable parameter [32]. Each neuron thus has a series of weighted inputs, w_{ij}, which may be either output from other neurons or input from external sources. Each neuron calculates a sum of the weighted inputs and transforms it by a transfer function,

$$\delta_j = \frac{1}{1 + e^{\frac{1-x}{\gamma}}} \tag{1}$$

where δ is the output from the j-th neuron connected to i-th neurons in the previous layer and γ is the gain determining the slope of the sigmoid transfer function, and x is given by:

$$x = \sum_{i=1}^{n} w_{ij} o_i + \theta_j \quad (2)$$

where w_{ij} represents the weight applied to the connection from neuron i to neuron j, o_i is the output from the i-th neuron in the previous layer and θ_j is a bias term.

BP networks operate in a supervised learning mode. In a first step (training), known data are given to the networks (a series of couples of experimental parameters–response values constituting the learning set). Using the BP algorithm, the network iteratively adjusts connection weights w_{ij} and biases t_i (starting from initial random values) so as it can calculate at output, values satisfactorily matching the observed values. In the BP training algorithm, this adjustment is carried out comparing the calculated target value (t_{ij}) and the desired output (o_i) of the network by means of calculation of the total sum of squares (*tss*) of the deviation ($t_{ij} - o_{ij}$) for the n patterns of the learning set (Eq. (3)):

$$tss = \sum_{i=1}^{n} (t_{ij} - o_{ij})^2 \quad (3)$$

The weights are then adjusted for all the interconnections between the output layer and the hidden layer. Likewise, the weights are adjusted for the interconnections between the hidden layer and the input layer.

3. Data description and computation

Different data sets have been used in order to study the possibility of application of ANNs to predict the best experimental conditions for CE separations. Data were generated either by the CELET program developed by Havel et al. [33,34] or by empirical equations reported in the literature [15–17].

In the present paper, a PDP package [30] was used for the purpose and processed on a Pentium-PC compatible computer. BP networks having three layers were created with this program and optimization of the networks' parameters was then carried out varying systematically their values until the 'best' network performance was achieved in order to optimize the conditions for low errors of the prediction.

Simulated data were used to examine how adjustable parameters affect the performance of the networks. These adjustable parameters include the scaling of the data, the momentum and learning rate (used in the weights adjustments), the number of hidden nodes and the number of learning cycles. Momentum and adaptive learning rates were used in order to improve the performance. The momentum allows for preventing sudden changes in the direction in which the solution is sought for. The momentum, taking into account the preceding correction of the weights, lowers the sensitivity of the network to small details in the error surface. The adaptive learning rate attempts to keep the learning step-size as large as possible, while keeping learning stable. The use of an adaptive learning rate leads to a lower training time. But if it is set too high, the error in prediction soon starts to oscillate or grows up.

3.1. Modeling/simulating with CELET program [23,34]

CZE data for systems including complex and/or acid–base equilibria were simulated using the CELET program. For effective mobility μ_{eff} of a component B under the presence of complexing ligands C and D (where A is supposed to be mostly H^+) and assuming that a series of complexes can be formed according to the general reaction:

$$pA + qB + rC + sD \underset{\beta_{pqrs}}{\Leftrightarrow} A_pB_qC_rD_s \quad (4)$$

a general relation is valid (charges are omitted for the sake of clarity).

$$\mu_{\text{eff}} = \frac{\sum_{i=0}^{N} \mu_i [A_{p_i} B_{q_i} C_{r_i} D_{s_i}]}{\sum_{i=0}^{N} q_i \beta_i [A]^{p_i} [B]^{q_i} [C]^{r_i} [D]^{s_i}} = \frac{\sum_{i=0}^{N} \mu_i \beta_i [A]^{p_i} [B]^{q_i} [C]^{r_i} [D]^{s_i}}{B_{\text{tot}}} \quad (5)$$

where μ_i represent the mobilities of individual $[A_{p_i}B_{q_i}C_{r_i}D_{s_i}]$ species, β_i are the total stability

constants (β_0 is supposed to be equal to 1) and B_{tot} is the analytical concentration of the component B. The summation is taken over all N reactions in the system.

Assuming certain values for individual mobilities of the species in the system any kind of plots, e.g., $\mu_{eff,exp}=f(pH)$ can be calculated. Theoretical values calculated, $\mu_{eff,calc}$, can then be superimposed with a random noise:

$$\mu_{eff,exp,i} = \mu_{eff,calc,i} + SIG \cdot EPS_i = \mu_{eff,calc,i} + S_{inst,i} \quad (6)$$

where *EPS* is generated by a standard routine [35] for the generation of random numbers while the value of *SIG* can arbitrarily be chosen by the user. The product *SIG·EPS* represents an instrumental error, S_{inst}.

4. Results and discussion

Below, the results of several case studies of ANNs will be described. We have started with the most simple, rather obvious Case (1), where we know a priori, the optimal pH for the separation of these two compounds. The idea is, however, to examine the behavior of ANNs and the potentiality of this method.

4.1. Case (1): Optimization of the separation of two weak acids

First, a simulation with the CELET program was performed in order to calculate data of the effective mobilities (Eqs. (7) and (8)) for a system of two weak acids, AH and BH. Second, this data was used in the neural networks training. The inputs to the neurons were the effective mobilities and pH. As the output we have used the difference in effective mobilities, which is related to the resolution.

$$AH^{x-} \Leftrightarrow A^{(x-1)-} + H^+ \quad \mu_{eff}(AH) = \mu_{AH^{x-}}[AH^{x-}] + \mu_{A^-}[A^{(x-1)-}] \quad (7)$$

$$BH^{x-} \Leftrightarrow B^{(x-1)-} + H^+ \quad \mu_{eff}(BH) = \mu_{BH^{x-}}[BH^{x-}] + \mu_{B^-}[B^{(x-1)-}] \quad (8)$$

We will study two different cases further, the first one with the same mobilites of AH and BH species (Fig. 2a) and the other one with different values of the mobilities of all species (Fig. 2b).

4.1.1. Case (1.1): System of two weak acids with equal mobilities for the anions and zero mobilities for the non-dissociated molecules

Even if this is a very basic example (see Fig. 2a) of a most simple and well understood system, we have studied this case in order to prove the capability of ANNs to fit the data by using BP algorithm. Furthermore, the effect of random noise in input data on the training procedure for ANN learning was also studied to prove the robustness of the method (see Fig. 3a). The levels of standard deviations, S_{inst}, of the noise loaded to the exact values of the effective mobilities were: 0, 0.1, 0.5, 1, 2.5 and 5.

Prior to processing, the input data were scaled into

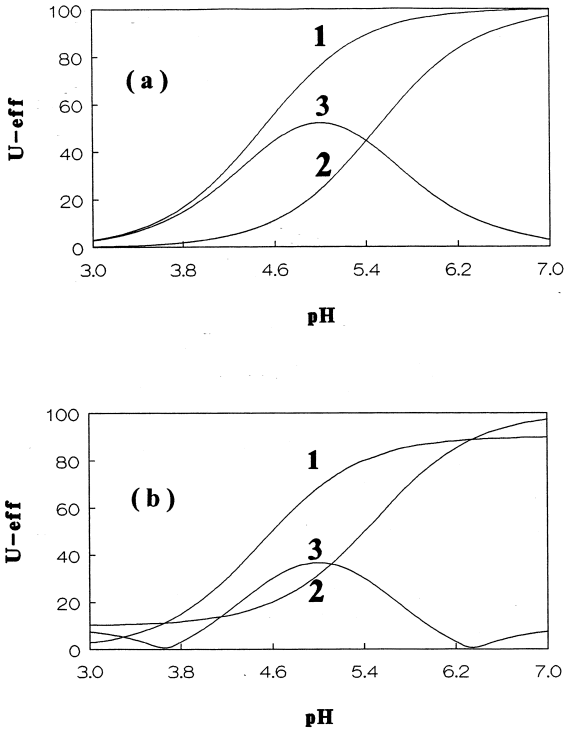


Fig. 2. Simulated effective mobilities and their difference in the case of two weak acids ($pK_{a1}=4.5$, $pK_{a2}=5.5$). (a) Equal mobilities for the anions and zero mobilities for the non-dissociated molecules. (b) Two weak acids with different mobilities for all the species.

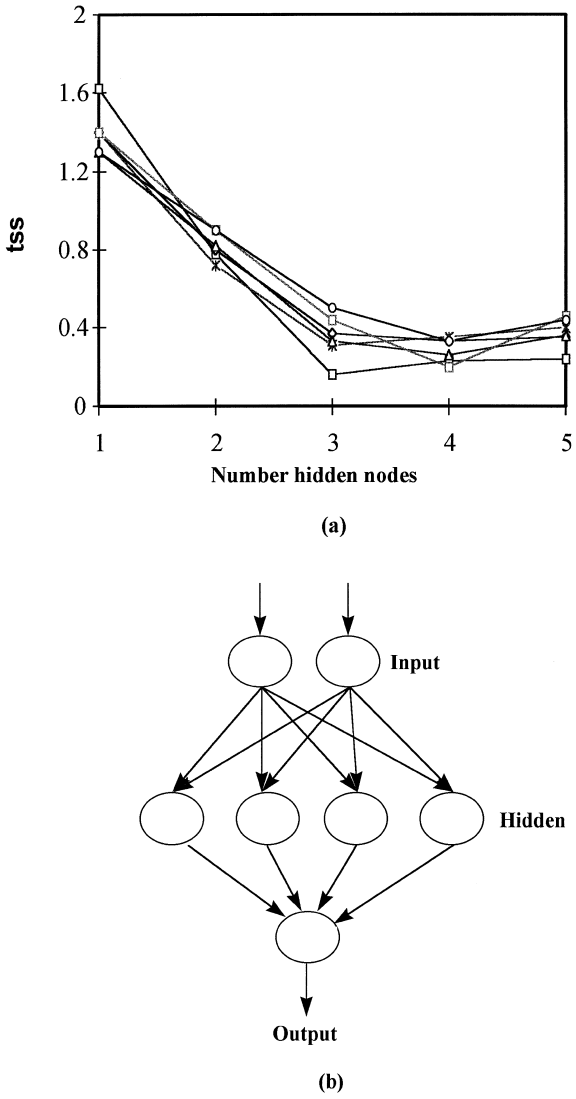


Fig. 3. (a) Root mean square error values versus the number of nodes in the hidden layer for several random noise levels of input data (level of standard deviation of the noise loaded to the exact values of the effective mobilities). Values of S_{inst}: 0, (*) 0.1, (◇) 0.5, (△) 1, (□) 2.5, (○) 5. (b) Optimal ANN structure used to solve Case (1.2).

the range 0.1 to 0.9 and rearranged to present to the ANN, three inputs (transformed pH, and two transformed values of the mobilities) and one output (transformed absolute value of the difference in mobilities). To facilitate the interpretation, the results obtained in the quantitation of the samples in the training and test sets are expressed as root mean square errors (*RMS*):

$$\%RMS = 100\sqrt{\frac{tss}{n \cdot m}} \tag{9}$$

where n is the number of variables in the pattern and m is the number of the samples.

In order to determine an optimal set of conditions for training, an artificial neural network with a low error for prediction and requiring the lowest possible number of iterations, were trained by using the simulated data superimposed as an instrumental noise. Training conditions were: momentum 0.9, l_{rate} 0.5 to 0.7, w_{range} 0.1 to 0.8. The number of cycles was fixed to 20 000. No significant improvement in the ANN results was observed beyond 20 000 cycles.

The results of the search for the optimal number of nodes in the hidden layer are shown in Fig. 3b for all simulated data sets and different levels of noise. The optimum number of nodes in the hidden layer of an ANN for which there was a good fit of the data, was found to be three in this case for low levels of random noise with standard deviations ($S_{inst} = 0.1$, 0.5, and 1) and it was equal to four for the higher level of the instrumental error ($S_{inst} = 2.5$ and 5).

4.1.2. Case (1.2): System of two weak acids, AH and BH, with different mobilities of all the species

The application of the same procedure as in the example of Case (1.1) for the transformation of data and training of the network, led to very similar results to those presented for Case (1.1). In this case (see Fig. 2b) it was found that by loading the various levels of noise to data of effective mobilities, the number of nodes in the ANN architecture are maintained the same (three nodes) (see Fig. 3a). Only for higher levels of noise studied (5%) was it found that four nodes in the hidden layer were needed.

We can conclude from Case (1.1) and Case (1.2) that in both cases, ANN architecture (see Fig. 3b) can describe quite well the relation between the input and output. This also means that ANNs can predict the best separation conditions without any necessity to use whichever ‘hard model’.

4.2. Case (2): Prediction of optimum separation of a mixture of metal ions by CZE

4.2.1. Combination of the experimental design and artificial neural networks

On the basis of encouraging results in Case (1) we have continued with examining the application of ANNs for more complicated situations. Prediction and optimization of the separation conditions in CZE is a crucial point of this method and can be in the case of many parameters (V, pH, C_L, injection mode, etc.) quite time consuming. Several methods and algorithms can be applied for this task [35–38], for example Simplex.

In the case of many parameters which are to be varied, the usual 'one-factor-at-time' method may not give an optimum set of conditions [14]. For that reason, it is necessary to use more intelligent methods but at the same time to reduce the number of experiments. The statistical methods that can be applied here are methods of experimental design already well described in the literature [18,26]. There are several possibilities; the most general is a full factorial design, Plackett–Burman design, central composite design, etc. We have applied CCD here.

CCD represents a simultaneous experimental design for k factors or variables. Having established the key factors affecting the performance of a method by using a screening design it may then be appropriate to optimize the method by obtaining a response surface, which permits the response surface to be modeled by fitting a second-order polynomial in $k+1$ dimensions. These plots can provide a graphical representation of the data over the ranges studied and can be used to predict areas of optimal performance.

Recently Quang and Khaledi [15] derived an empirical model to relate the electrophoretic mobility (μ_{ep}) in CZE of 14 different cations (e.g. alkaline, alkaline-earth and transition divalent cations) with pH and complexing agent concentration (C_{HL}), as described by Eq. (10).

$$\mu_{ep} = k_0 + k_1 C_{HL} + k_2 \text{ pH } C_{HL} + k_3 C_{HL}^2 \qquad (10)$$

where k_0, k_1, k_2, k_3 are empirical constants found by a general regression.

This experimental dependence is important in order to predict the resolution for a complex mixture, because the resolution (R_s) depends on these parameters, in agreement with Eq. (11).

$$R_s = [(N)^{1/2}/4]\{(\mu_a - \mu_b)/(\mu_{av} + \mu_{eo})\} \qquad (11)$$

where μ_a and μ_b are electrophoretic mobilities of two neighbouring metal cations ($\mu_a > \mu_b$), and μ_{av} and μ_{eo} are average and electroosmotic mobilities, respectively.

In order to study the potentiality of experimental design-ANN to predict optimal CZE separation of several metals ions, the response resolution surface (Fig. 4) has been simulated. The surface in Fig. 4 represents resolution for three different cations, M_1, M_2 and M_3, using the maximum for the minimal resolution as the separation criterion. The empirical values assumed for Eq. (10) were taken from Table 1 to calculate the values of $\mu_1(M_1)$, $\mu_2(M_2)$ and $\mu_3(M_3)$. The concentration of the complexing agent was varied from 2 to 20 mM and the pH values were taken between 3.5 and 4.5. The value of R_s was calculated by taking the minimum absolute value between $|\mu_2 - \mu_1|$, $|\mu_3 - \mu_1|$ and $|\mu_3 - \mu_2|$, the corresponding μ_{av} and assuming constant values for N and μ_{eo} in the working conditions.

4.2.2. Step 1. Search of optimal ANN architecture

First of all the data were normalized and rearranged to present two inputs (normalized C_{HL} and pH) and one output (normalized resolution), see Table 2.

In order to get good results in the training step, the values of the learning rate (l_{rate}) and weight range (w_{range}) were varied. In this empirical work, a starting set of weights was used in order to achieve the minimal *tss* values for $l_{rate} = 0.3$, $w_{range} = 0.05$ and 20 000 cycles. The training of ANNs was done for several different ANN structures and it lead finally to the optimal performance for 2:7:1 (input:hidden:output) architecture, shown in Fig. 5.

The residuals shown in Fig. 6a for the training set led us to conclude that the maximum deviation corresponds to the position of the two maxima as shown in Fig. 4. If the patterns of these two maxima are omitted in the training set, a much lower *tss* value was obtained. On the other hand, the residuals

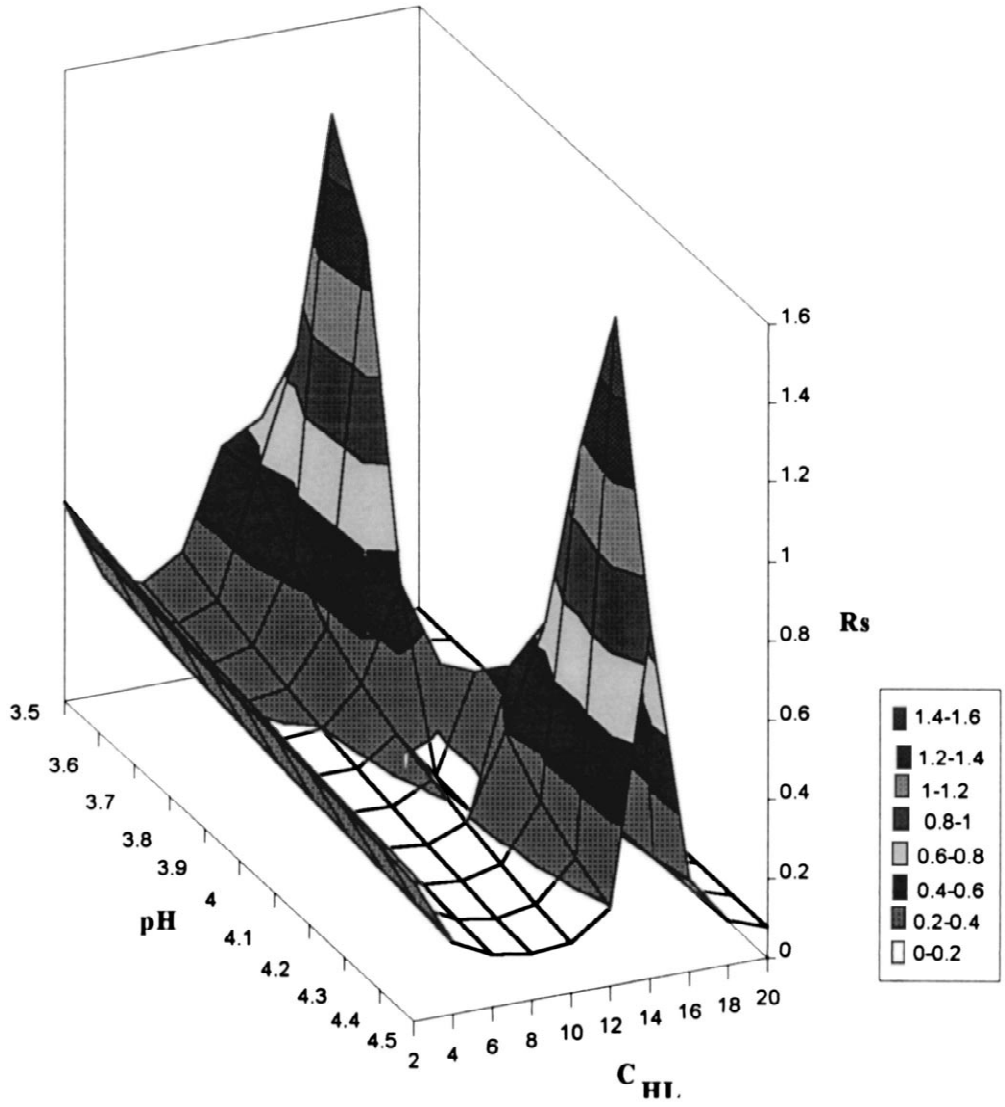


Fig. 4. Response resolution surface for capillary electrophoretic separation of three different cations, modelled using Eqs. (4) and (5). Concentration of HIBA, C_{HL} (m*M*).

Table 1
Empirical constants used to calculate electrophoretic mobilities for each cation with Eq. (10)

Empirical constants in Eq. (1)	M1	M2	M3
k_0	−60.0	−99.0	50.0
k_1	14.0	15.0	15.1
k_2	25.0	−45.0	−48.0
k_3	−10.0	13.4	14.0

shown in Fig. 6b for the test set demonstrate a very good prediction ability using the weights obtained in the learning procedure of the optimal ANN architecture. High deviations of *tss* values for five patterns were observed, see Fig. 6b. These deviations are due to the lack of information, i.e., to the absence of the other patterns in the training set close to these points.

Table 2
Normalized resolution values used to study the performance of ANNs, Case (2)

	Normalized resolution C_{HL} (mM)									
pH	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.778	0.313	0.187	0.164	0.201	0.358	0.392	0.128	0.077	0.056	0.046
0.800	0.301	0.177	0.151	0.176	0.366	0.552	0.139	0.079	0.057	0.046
0.822	0.290	0.168	0.140	0.156	0.281	0.981	0.153	0.082	0.058	0.046
0.844	0.279	0.159	0.130	0.139	0.225	0.828	0.171	0.085	0.059	0.046
0.867	0.269	0.151	0.121	0.125	0.186	0.332	0.196	0.089	0.060	0.046
0.889	0.260	0.144	0.113	0.113	0.157	0.079	0.234	0.094	0.061	0.047
0.911	0.251	0.137	0.105	0.103	0.134	0.074	0.295	0.100	0.062	0.047
0.933	0.243	0.131	0.099	0.094	0.116	0.253	0.414	0.108	0.064	0.047
0.956	0.235	0.125	0.093	0.086	0.102	0.192	0.742	0.117	0.065	0.048
0.978	0.228	0.119	0.087	0.079	0.090	0.152	1.000	0.130	0.068	0.048
1.000	0.221	0.114	0.082	0.072	0.080	0.124	0.557	0.147	0.070	0.048

4.2.3. Step 2. The prediction of optimum CZE separation using a combination of experimental design and ANN

In order to predict optimum CZE separation of several metal ions the following strategy was applied:

(1) The bold black numbers in Table 3 can be taken as a starting point for the process. With these five patterns (1 to 5 in Fig. 7) the ANN was trained and so the corresponding weights (w_{ij}) were obtained.

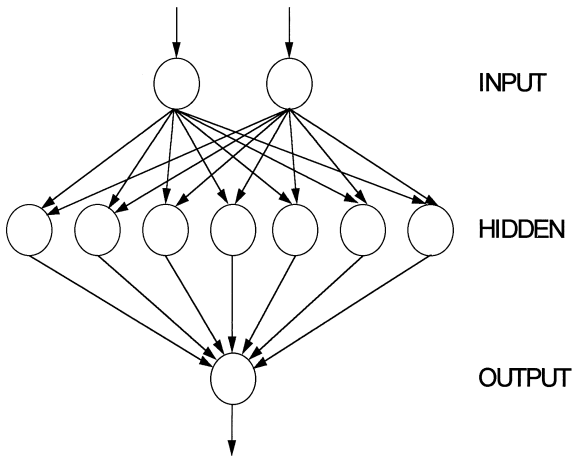
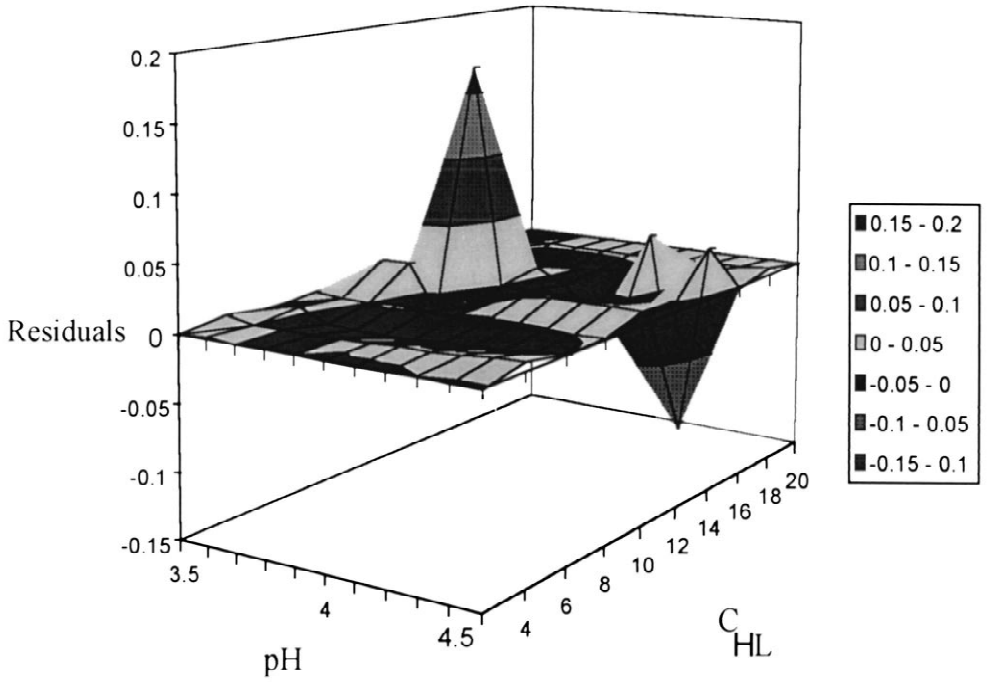


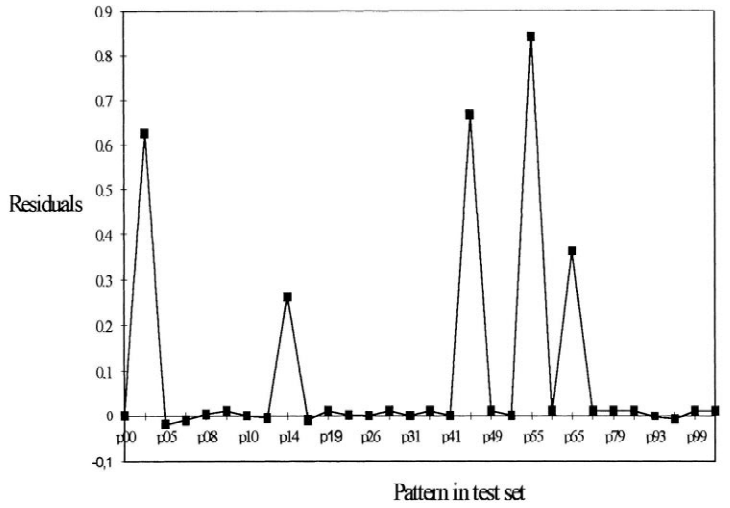
Fig. 5. The optimal ANN architecture (2:7:1) with 7 nodes in the hidden layer, using the back-propagation algorithm.

(2) We then applied the BP program and performed the prediction. The weights (w_{ij}) obtained during the neural networks training in the 1st step (1) were used to predict the values of the 1st test set, constituted by the bold italic black numbers, as shown in Table 3. Fig. 8 shows the comparison between predicted data obtained with ANNs and simulated data (see Table 3). It is worth mentioning the similarity in the trends of the curves between the 'predicted' (with ANNs) and the 'experimental' (simulated) curves. Nevertheless, there is a great absolute deviation observed for the maximum value. As it may easily be seen in Fig. 8, the predicted resolution value for pattern No. 1 is greater than the resolution value obtained (simulated) for the other points used in the 1st test set. It can be concluded that it could be useful to realize new 'experiments' around this point. The additional 'experiment' has been done (see Fig. 7, points 6 to 10) and the results confirmed that there really is a maximum resolution around these pH and C_{HL} conditions.

(3) To be certain about the results achieved above in (2), we made additional experiments. Another experimental design (bold underlined or bold italic underlined numbers in Table 3) was applied in the region centered on the greatest value obtained during the previous prediction. In this way we obtained a



(a)



(b)

Fig. 6. Results obtained using 2:7:1 ANN architecture. (a) Residuals for the training set in the simulated space of pH and C_{HL}. Concentration of HIBA, C_{HL} is in m*M*. (b) Residuals for the test set.

Table 3
Subset of the normalized resolution data to train the network

C_{HIBA} (m*M*)	pH				
	0.5	0.6	0.7	0.8	0.9
0.77778	**0.35782**	0.39159	0.12817	0.07653	0.05601
0.80000	**0.3655**	0.55182	**0.13892**	0.07908	**0.05667**
0.82222	0.28057	**0.98130**[a]	0.15268	*0.08202*	0.05762
0.84444	**0.22502**	0.82834	**0.17093**	0.08545	0.05856
0.86667	0.18585	0.33160	0.19627	0.08949	0.05961
0.88889	0.15674	0.07913	**0.23386**	0.09434	0.06080
0.91111	*0.13427*	0.07367	**0.29538**	0.10025	0.06214
0.93333	0.11639	**0.25319**	0.41441	***0.10763***	0.06367
0.95556	*0.10182*	0.19210	***0.74228***	0.11709	0.06545
0.97778	0.08973	0.15203	**1.00000**[a]	0.12968	0.06751
1.00000	**0.07953**	0.12373	0.55741	0.14721	**0.06995**

Bold numbers are marked to show the reduced factorial design used to search for the maximum response surface as shown in Fig. 4.
[a] Numbers correspond to maxima as shown in Figs. 4 and 8.

2nd training set, shown in Table 4. The quality of the fit achieved in the second training procedure was excellent. Therefore, it really was assured that adding a new experiment did not yield any further improvement, which was evidence that the optimum had already been found.

A question might arise here concerning the second maximum, see Fig. 4. We have also examined the above described procedure 1 to 3 and starting around this second maximum, similar results have been obtained by applying the procedure for the points marked as bold italics underlined in Table 3.

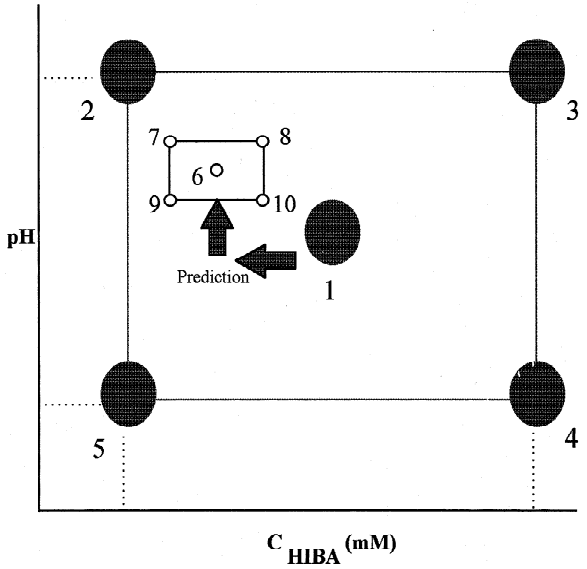


Fig. 7. Schematic diagram of experimental design–ANN optimization algorithm. Points 1, 2, 3, 4 and 5 correspond to the starting central composite design (CCD). Point 6 is a predicted point which can be the centre of a new, successive CCD consisting of points 6, 7, 8, 9 and 10.

In conclusion, it is possible to say that it has been demonstrated that using the experimental design–ANN approach we have obtained similar results to Khaledi et al. [15] or Massart et al. [16,17] but in contrast to the results of these authors, we can predict optimal separation with a much lower number of experiments. This approach was successful for resolving a complex system, such as the separation of many cations with a complicated response surface as shown in Fig. 4.

5. Conclusions

Using ANNs, the optimal separation conditions (and/or the resolution particularly) in CZE can be successfully predicted without using or knowing any explicit model of the separation process. The combination of experimental design and artificial neural networks with the back-propagation algorithm was found to be a powerful tool in predicting optimal separation conditions from a low number of experiments. This new experimental design–ANN approach proposed here, is quite general and can also be used for the optimization of any other separation processes.

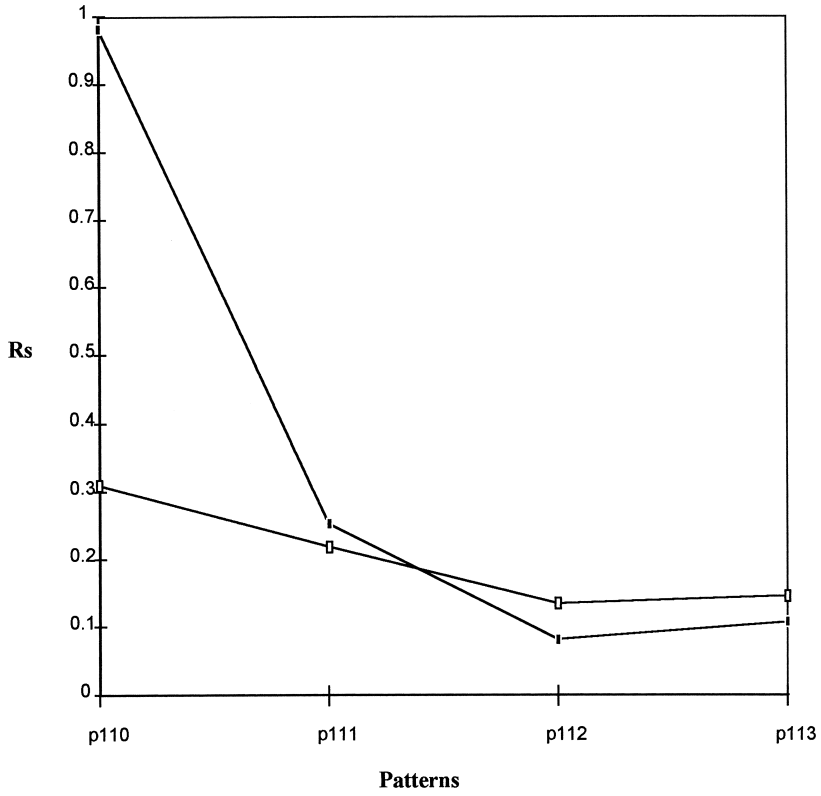


Fig. 8. Comparison between predicted and simulated data for test set in the 1st step of optimization (■ predicted, □ simulated). The 'experiment' confirms that a p110 pattern shows a high resolution value.

Acknowledgements

We wish to acknowledge the support from Grant Agency of Czech Republic, grant No. 203/96/0478. E.M.P. thanks to Dirección General de Universidades del Gobierno de Canarias by Scholarship for the post-doctoral stay in Brno. A.R.-H. acknowledges UAM-Iztapalapa and FESC-UNAM for partial financial support.

Table 4
Training set used in the 2nd step of the optimization procedure

Pattern no.	Normalized concentration C_{HIBA}	Normalized pH	Resolution
T06	0.5	0.80000	0.366
T07	0.7	0.80000	0.139
T08	0.6	0.82222	0.981
T09	0.5	0.84444	0.225
T10	0.7	0.84444	0.171

References

[1] S. Hjertén, Chromatogr. Rev. 9 (1967) 122.
[2] S.F.Y. Li, Capillary Electrophoresis, Journal of Chromatography Library, Elsevier, New York, 1992.
[3] F. Foret, L. Křivánková, P. Boček, Capillary Zone Electrophoresis, VCH, New York, 1993.
[4] R. Weinberger, Practical Capillary Electrophoresis, Academic Press, New York, 1993.
[5] R.L. St. Claire, Anal Chem. 68 (1996) 569.
[6] H. Corstjens, A.E.E. Oord, H.A.H. Billiet, J. Frank, K.C.A.M. Luyben, J. High. Resolut. Chromatogr. 18 (1995) 551.
[7] U. Bütenhord, U. Pyell, J. Chromatogr. A 772 (1997) 27.

[8] F.M. Everaerts, J.L. Beckers, Th.P.E.M. Verheggen, Isotachophoresis–Theory, Instrumentation and Applications, Journal of Chromatography Library, Vol. 6, Elsevier, Amsterdam, 1976.
[9] R.A. Mosher, D.A. Saville, W. Thormann, The Dynamics of Electrophoresis, VCH, Weinheim, 1992.
[10] S.V. Neomakav, M. Yu Zhukov, L. Capelli, P.-G. Righetti, Anal. Chem. 67 (1995) 2957.
[11] J.C. Reingenga, E. Kenndler, J. Chromatogr. A 659 (1994) 403.
[12] J.H.P.A. Martens, J.C. Reigenga, J.H.M. ten thije Boon Kamp, R.M.M. Mattheij, F.M. Everaerts, J. Chromatogr. A 772 (1997) 49.
[13] C.P. Palmer, B.G.M. Vandengiste, J. Chromatogr. A 718 (1995) 153.
[14] D. K Bempong, I.L. Honigberg, J. Pharm. Biomed. Anal. 15 (1996) 233.
[15] Ch. Quang, M.G. Khaledi, J. Chromatogr. A 659 (1994) 459.
[16] Q. Yang, J. Smeyers-Verbeke, W. Wu, M.S. Khots, D.L. Massart, J. Chromatogr. A 688 (1994) 339.
[17] M. Jidmar, T. Hamoir, W. Degezelle, D.L. Massart, S. Soyens, P. Van der Winkel, Anal. Chim. Acta 284 (1993) 217.
[18] K.D. Altria, B.J. Clark, S.D. Filbey, M.A. Kelly, D.R. Rudd, Electrophoresis 16 (1995) 2143.
[19] T.S. Small, A.D. Fell, M.W. Coleman, J.C. Berridge, Chirality 7 (1995) 226.
[20] J. Zupan, J. Gasteiger, Anal. Chim. Acta 248 (1991) 1.
[21] T.B. Blank, S.D. Brown, Anal. Chim. Acta 277 (1993) 273.
[22] M. Bos, A. Bos, E. Van der Linden, Analyst 118 (1993) 323.
[23] S. Ventura, M. Silva, D. Pérez-Bendito, Anal. Chem. 67 (1995) 4458.
[24] M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, M. Redón, Anal. Chem. 67 (1995) 4477.
[25] M. Volmer, B.G. Wolthers, H.J. Metting, T.H. de Haan, P.M.J. Coenegracht, W. Van der Slik, Clin. Chem. 40 (1994) 1692.
[26] H.J. Metting, P.M.J. Coenegracht, J. Chromatogr. A 728 (1996) 47.
[27] M. Farková, E.M.Peña-Méndez, A. Rojas-Hernández, J.P. Doucet, J. Havel, Chemometrics IV, Brno, 8–12 September, 1996, Abstracts, p. 63.
[28] E. M. Peña, J. Havel, ITP'96, Prague, 17–20 September, 1996, Abstracts, p. 50.
[29] D.L. Massart, B.G.M. Vandeginste, S.M. Deming, Y. Michotte, L. Kaufman, Chemometrics: A Textbook, Elsevier, Amsterdam, 1988.
[30] J.L. McCelland, D.E. Rumelhart, Explorations in Parallel Distributed Processing, MIT Press, Massachusetts, 1988.
[31] A. Panaye, J.P. Doucet, B.T. Fan, E. Feuilleaubois, P. Ladd, Chem. Intell. Lab. Sys. 24 (1994) 129.
[32] M.T. Spining, J.A. Darsey, B.G. Sumpter, D.W. Noid, J. Chem. Educ. 71 (1994) 406.
[33] J. Havel, P. Janoš, Chemometrics IV, Brno, 8–12 September, 1996, Abstracts, p. 67.
[34] J. Havel, P. Janoš, P. Jandik, J. Chromatogr. A 745 (1996) 127.
[35] J.R. Bell, Commun. Assoc. Comp. Mach. 11 (1968) 498.
[36] J.L. Glajch, J.J. Kirkland, J. Chromatogr. 485 (1989) 51.
[37] P.M.J. Coenegracht, A.K. Smilde, H.J. Metting, D.A. Doornbos, J. Chromatogr. 485 (1989) 195.
[38] E. Morgan, Chemometrics: Experimental Design, Wiley, Chichester, 1991.