JOURNAL OF CHEMOMETRICS
J. Chemometrics 2007; **21**: 567–577
Published online 4 September 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/cem.1082

Combining classification and regression approaches for the quantification of highly overlapping capillary electrophoresis peaks by using evolutionary sigmoidal and product unit neural networks

Cesar Hervás[1], Pedro Antonio Gutierrez[1], Manuel Silva[2]* and Juan Manuel Serrano[2]

[1]University of Córdoba, Computer Science, Spain
[2]University of Cordoba, Analytical Chemistry, Spain

Received 27 February 2007; Revised 13 July 2007; Accepted 25 July 2007

This is a study of the potential of neural networks built by using different transfer functions (sigmoidal, product and sigmoidal–product units) designed by an evolutionary algorithm to quantify highly overlapping electrophoretic peaks. To test this approach, two aminoglycoside antibiotics, amikacin and paramomycin, were quantified from samples containing either only one component or mixtures of them though capillary zone electrophoresis (CZE) with laser-induced fluorescence (LIF) detection. The three models assayed used as input data the four-parameter Weibull curve associated with the profile of the electrophoretic peak and in some cases the class label for each sample estimated by cluster analysis. The combination of classification and regression approaches allowed the establishment of straightforward network topologies enabling the analytes to be quantified with great accuracy and precision. The best models for mixture samples were provided by product unit neural networks (PUNNs), 4:4:1 (14 weights) for both analytes, after discrimination by cluster analysis, allowing the analytes to be quantified with great accuracy: 8.2% for amikacin and 5.6% for paromomycin within the standard error of prediction for the generalization test, SEP_G. For comparison, partial least square regression was also used for the resolution of these mixtures; it provided a minor accuracy: SEP_G 11.8 and 15.7% for amikacin and paramomycin, respectively. The reduced dimensions of the neural networks models selected enabled the derivation of simple quantification equations to transform the input variables into the output variable. These equations can be more easily interpreted from a chemical point of view than those provided by other ANN models.

KEYWORDS: neural networks; evolutionary algorithm; overlapped peaks; capillary electrophoresis

1. INTRODUCTION

A variety of approaches are found in the literature for resolving and quantifying overlapped chromatographic bands using chemometric methods, in the majority of the cases based on a two-dimensional array of data, such as that generated by high performance liquid chromatography (HPLC) with diode-array detection (DAD) [1–9]. In recent years, capillary zone electrophoresis (CZE) has gained in popularity as an alternative to HPLC for routine analysis, and its application is widespread in many fields of analytical chemistry. Despite its higher resolution with respect to HPLC, sometimes the complete separation of sample components is not fully achieved. Nonetheless, the quantification difficulties from poorly resolved peaks may be overcome mathematically by using chemometric techniques. Multivariate curve resolution methods [10–12], augmented iterative target transformation factor analysis [1], wavelet transforms [13] and second-order derivative electropherograms [14,15] are, among other approaches, the most recently reported for this purpose using the second-order data from CZE–DAD. In the last decade, artificial neural networks (ANNs) have shown their unique merits regarding the great variety of chemometric approaches reported for the classification and regression purposes in many fields of analytical chemistry. Particularly in separation

*Correspondence to: M. Silva, University of Cordoba, Analytical Chemistry, Cordoba, Spain.
E-mail: qa1sirom@uco.es

science, ANNs have been used mainly as tools to optimize the experimental conditions for carrying out separation and, to a lesser extent, for quantification in overlapped HPLC [16–20] peaks and, more recently, for unresolved peaks in CZE [21–25].

Several ANN models have been used to achieve regression analysis, including quantification in overlapped peaks, in analytical chemistry. Although multilayer perceptron (MLP) modeling with sigmoidal unit basis functions (SU) as transfer functions is the most widely used approach [16,20–23], radial basis function (RBF) neural networks [19,21] and, more recently, multiplicative neural networks [26], especially the so-called product unit neural networks (PUNNs) [17,27,28] modeling with product unit basis functions (PU) are other interesting choices depending on the analytical problem addressed. In addition to this type of models where all the nodes of the hidden layer have the same type of activation/transfer functions, hybrid models have also been proposed, where different activation/transfer functions are used for the nodes in the hidden layer.

In this context, it is worth emphasizing the papers by Duch and Jankowski [29] which propose different transfer functions for the nodes in the hidden layer and those by Cohen and Intrator [30,31], supported on the duality between functions based on projection, (SU, PU, etc.) and on kernel typology (RBF). The hybridization of models has been justified theoretically by Donoho [32] who demonstrated that any continuous function can be decomposed into two mutually exclusive functions, such as radial and crest (based on the projection) ones. But to the best of our knowledge there is not any theoretical work associated with the hybridization of the SU/PU in neural networks. Although theoretically this decomposition is justified, in practice it is difficult to separate the different locations of a function and to estimate them by means of a combination of RBFs, and then to estimate the residual function by means of a functional approach based on projections without getting trapped in local optima in the procedure of minimization of error by using gradient methods [33]. A similar idea has been considered in this work: neural network architecture with several transfer functions based on projections.

Thus, this paper deals with the evaluation of different evolutionary ANN models based on SU, PU and a combination of both basis function (SPU) neural networks as powerful tools for the quantification of analytes that provide highly overlapping chromatographic peaks by using first-order data. To test proposed approaches, two aminoglycoside antibiotics, amikacin and paramomycin, were quantified in mixtures using only the analytical information provided by their overlapped CZE peaks (t_i,S_{t_i}: first-order data) registered with a laser-induced fluorescence (LIF) detector. Further details on the analytical methodology are described in a previous paper [34]. Several chemical and chemometric strategies were merged in the proposed approaches:

1. The methodology was extended to more practical and real situations considering the possibility that the analyzed samples could either contain only one component or mixtures of them.
2. A classic procedure was used for classification, such as linear discriminant analysis (LDA) [35], in order to differentiate the three types of analyzed samples.
3. Models of networks with a limited number of inputs were tested. To do so, as in previous papers [16,17], the network inputs were estimated by the Levenberg–Marquardt method in the form of a four-parameter Weibull curve associated with the profile of the electrophoretic band.
4. An additional input was introduced, namely the label associated with the class of sample analyzed, in order to improve the capacity of generalization of the models; this proposal constitutes the first precedent on the joint employment of classification and regression approaches for resolving overlapped chromatographic peaks.
5. The potential of ANNs with hybrid models for activation/transfer functions was evaluated. This paper shows the first research on the use of hybrid models associated with two specific types of functions as functional projection approximators: product and sigmoidal functions, yielding the SPUNNs hybrid model.
6. Evolutionary algorithms were employed for the optimization of the parameters of the model as an alternative to the classic choice based on gradient methods, considering the complexity of the error surfaces obtained by minimizing the squared errors derived from the fitting of the regression models in the training set.

2. THEORY

The aim of the proposed approach is to evaluate the potential of different ANN models: SU, PU and SPUNNs, for the quantification of overlapped chromatographic bands, concretely unresolved CZE peaks, in order to predict the contribution of each component to the overall analytical signal.

2.1. Selection of ANN inputs

The first step of the approach consists of extracting the information from the analytical response (CZE peak) in order to select the inputs for the ANNs. Upon examining the responses (individual and mixture samples), it can be observed that profiles of the CZE peaks (t_i, S_{t_i}) were accurately fitted by least-square regression to a four-parameter Weibull curve defined by S_m (peak height), B (dispersion of S_{t_i} values from S_m), C (associated to the inflection points of the curve and defining the concavity and convexity regions) and t_m (residence time). In addition, the samples were also classified into three categories by using LDA. Classes were defined according to the composition of the samples as follows: G_1, amikacin; G_2, paromomycin; and G_3, mixtures of both antibiotics. In some studies, the label associated with the class was used as input for ANN models.

2.2. ANN models

ANNs modeling with SU, PU and hybrid SPU as transfer functions were used and their features compared in this work. Taking into account that the latter is the most novel approach, its foundation is now described in detail.

DOI: 10.1002/cem

Although the use of PUNNs in this context is very recent, the description of its theoretical basis is outside the scope of this paper, although some current papers from the authors can be considered for further explanations [17,27,28].

To estimate the concentrations of the antibiotics in the samples by using SPUNN models, we consider standard regression formulation under a general setting for predictive learning. In the regression formulation, the goal of learning is to estimate an unknown (target) real-valued function $g(\boldsymbol{x})$ in the relationship

$$y = g(\mathbf{x}) + \varepsilon \tag{1}$$

where ε is an independent and identically distributed zero mean random noise, (whose distribution can depend on $\boldsymbol{x}$, the distribution of $\boldsymbol{x}$ in the training set also being unknown), $\boldsymbol{x}$ is a multivariate input and y is a scalar output. The estimation is made based on a finite number (n) of samples (training data): $(\mathbf{x}_1,y_1), l = 1,2,\ldots,n$. To solve regression problems, the family of functions $f(\boldsymbol{x}, \boldsymbol{w})$ ($\boldsymbol{w}$ is a set of parameters of f) used in the approximation should be previously determined by using a learning method that selects the best model, $f(\boldsymbol{x}, \boldsymbol{w_0})$, from $f(\boldsymbol{x}, \boldsymbol{w})$; finally, the quality of the approximation can be expressed by the loss (discrepancy) measure $L(y, f(\boldsymbol{x}, \boldsymbol{w}))$, the squared error being a common loss function used for regression.

The linear models given by $f(\boldsymbol{x}, \boldsymbol{w}) = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}$ are the simplest ones used to solve regression problems. It is extremely unlikely that the true function $g(\boldsymbol{x})$ is actually linear in $\boldsymbol{x}$. Normally, $g(\boldsymbol{x})$ is nonlinear and non-additive in $\boldsymbol{x}$, and representing $g(\boldsymbol{x})$ by a linear model is usually a convenient, and sometimes necessary, approximation. Another class of useful approximation can be expressed as a linear basis expansion

$$f(\mathbf{x}) = \sum_{j=1}^{M} \beta_j B_j(\mathbf{x},\mathbf{w}_j) \tag{2}$$

where $B_j(\mathbf{x},\mathbf{w}_j)$ are a suitable set of functions or nonlinear transformations of the input vector $\mathbf{x} = (\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_p)$, $B_0(\mathbf{x},\mathbf{w}_j) = 1$ in order to consider bias in the model; β_j are the coefficients from lineal combination that are estimated from the data; $\mathbf{w}_j = (\mathbf{w}_{j1},\mathbf{w}_{j2},\ldots,\mathbf{w}_{jp})$ are the parameters associated with the basis functions; and M is the parameter of regularization of the model, which is associated with the number of basis functions that are necessary and sufficient to minimize some definite function of the error on this matter. In this work, two types of basis functions have been used, namely SU function

$$B_j(\mathbf{x},\mathbf{u}_j) = \frac{1}{1+e^{-\left(\mathbf{u}_{j0}+\sum_{i=1}^{p}\mathbf{u}_{ji}\mathbf{x}_i\right)}}, \qquad j = 1,\ldots,m_1 \tag{3}$$

and PU function

$$B_k(\mathbf{x},\mathbf{w}_k) = \prod_{i=1}^{p} \mathbf{x}_i^{\mathbf{w}_{ki}}, \qquad k = 1,\ldots,m_2 \tag{4}$$

whose linear combination provided the hybrid function (SPU) used for estimation:

$$f(\mathbf{x}) = \alpha_0 + \sum_{j=1}^{m_1} \alpha_j B_j(\mathbf{x},\mathbf{u}_j) + \sum_{k=1}^{m_2} \beta_k B_k(\mathbf{x},\mathbf{w}_k) \tag{5}$$

The method involves finding a sufficient number of basis functions (architecture) providing a type of approximate universal function for the estimate function, in such a way, and taking into account that both types of basis functions are universal approximators [17], that for every $\varepsilon > 0$ it should be possible to find a value of m_1 and m_2 as well as the estimators of the parameters $\alpha_0,\alpha_j,\beta_k,\hat{\mathbf{u}}_j$ and $\hat{\mathbf{w}}_k$ for $j = 1,\ldots,m_1$ and $k = 1,\ldots,m_2$, that hold:

$$\left\| f(\mathbf{x}) - \left(\alpha_0 + \sum_{j=1}^{m_1} \alpha_j B_j(\mathbf{x},\mathbf{u}_j) + \sum_{k=1}^{m_2} \beta_k B_k(\mathbf{x},\mathbf{w}_k) \right) \right\| < \varepsilon \tag{6}$$

This optimization problem is similar to that involved in the 'projection pursuit' regression model with the special feature being that the 'ridge functions' are exclusively of two types. Evolutionary algorithms similar to those reported by Angeline *et al*. [36], Yao and Liu [37] and García *et al*.[38] and that have been used in this work to obtain the ANN architecture and to estimate model coefficients.

This kind of function topology can be represented by an ANN architecture, as shown in Figure 1. The network has k inputs that represent the independent variables of the model, five in our problem, $m_1 + m_2 = 6$ nodes in the hidden layer and one node in the output layer. The activation of the jth sigmoidal and kth product nodes in the hidden layer is given by Equations (3) and (4), respectively; where $\mathbf{u}_{ji}$ is the weight of the connection between input node i and hidden sigmoidal node j and $\boldsymbol{w}_{ki}$ is the weight of the connection between input node i and hidden product node k. The activation of the node in the output layer is given by Equation (5); where α_j and β_k are the weights of the connection between the hidden sigmoidal node j or product node k and the output node. The transfer functions of each product node in the hidden layer and the output node for regression are the identity function. In this way, each function of the $f(\boldsymbol{x}, \boldsymbol{w})$ family is represented by the structure of the network.

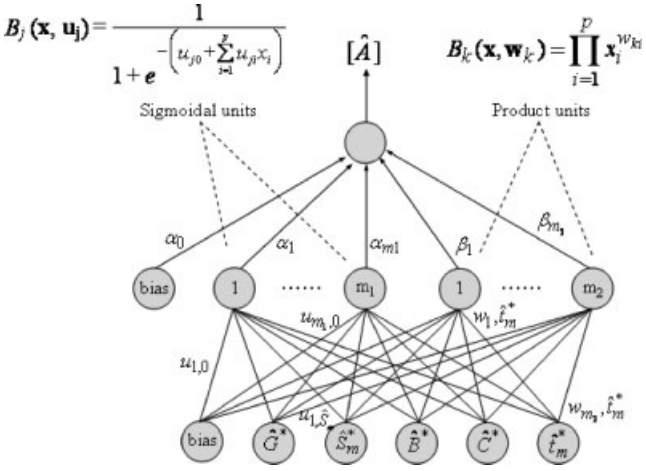

Figure 1. Functional scheme of the neural network based on sigmoidal and product units. $\alpha_0,\ldots\ldots,\alpha_{m_1}$ and $\beta_0,\ldots\ldots,\beta_{m_2}$ are the regression coefficients of the model. Other symbols are defined in the text.

DOI: 10.1002/cem

2.3. Evolutionary algorithm

The optimization of the SPUNN topology consisted of the search for the structure of the sigmoidal and product unit base functions that best fit the data of the training set, by determining the values of m_1 and m_2 associated with the optimum number of base functions for each type involved. On the other hand, the estimation of the weights of the network is closed to the evolution of the $\mathbf{u}_j$ and $\mathbf{w}_j$ vectors, which determine the coefficients in each base function, as well as with α_j and β_k, coefficients involved in the linear combination of the base functions. The population is subjected to the operations of replication and structural and parametric mutations. The general structure of the algorithm is the following:

(1) Generate the initial population.
(2) Repeat until the stopping criterion is fulfilled.
(a) Calculate the fitness of every individual in the population.
(b) Rank the individuals regarding their fitness.
(c) The best individual is copied into the new population.
(d) Ten per cent of the best individuals in the population are replicated and substitute the 10% of worst individuals.
(e) Apply parametric mutation to 10% of the best individuals.
(f) Apply structural mutation to the rest of the 90% of individuals.

For the generation of the initial population, the algorithm begins with the random generation of a larger number of networks than the number used during the evolutionary process. Initially, we generate 10 000 networks from which, the best 1000 with major fitness are extracted, being it the population size along the evolutionary process. For the generation of a network, the number of nodes in the hidden layer is taken from a uniform distribution in the interval [0,6], for $m_1 + m_2 = 6$, where m_1 and m_2 are related to the maximum number of hidden nodes in each base function in the population. The number of connections between each node of the hidden layer and the input nodes is determined from a uniform distribution in the interval [0,k], where k is the number of independent variables. There is always at least one connection between the hidden layer and the output node. Once the topology of the network is defined, each connection is assigned a weight from a uniform distribution in the interval [−5,5] for the weights between the input and hidden layers, and the same interval for the weights between the hidden layer and the output node. Two types of mutations are performed in the algorithm: parametric and structural. The parametric mutations affect the weights of the network and the structural ones affect to the network topology (hidden nodes and connections). The severity of a mutation is dictated by the function's temperature $T(f(\mathbf{x}))$ given by:

$$T(f(\mathbf{x})) = 1 - A(f(\mathbf{x}))\ 0 \leq T(f(\mathbf{x})) \leq 1 \tag{7}$$

where $A(f(\mathbf{x}))$ is the fitness value of the $f(\mathbf{x})$.

Let $D = \{(\mathbf{x}_l, y_l) : l = 1, 2, \ldots\ldots, n_T\}$ be the training data set, where the number of samples is n_T. In this context, we consider that the mean squared error MSE of an individual $f(\mathbf{x})$ of the population is given by:

$$\mathrm{MSE}\,(f(\mathbf{x})) = \frac{1}{n_T}\sum_{l=1}^{n_T}(y_l - f(\mathbf{x}_l))^2 \tag{8}$$

where the y_l are the predicted values, whereas the fitness function $A(f(\mathbf{x}))$ is defined by means of a strictly decreasing transformation of the MSE:

$$A\,(f(\mathbf{x})) = \frac{1}{1 + \dfrac{1}{n_T}\displaystyle\sum_{l=1}^{n_T}(y_l - f(\mathbf{x}_l))^2} \tag{9}$$

where $0<A(f(\mathbf{x}))<1$.

Parametric mutation consists of a simulated annealing algorithm [39] based on a strategy that proceeds by proposing jumps from the current model according to some user-designed mechanism. Structural mutation is more complex because it implies a modification of the structure of the function. Structural mutation allows the exploration of different regions in the search space and helps to keep up the diversity of the population. Five mutations are applied sequentially to each network. The first four are similar to the mutations of the GNARL model reported by Angeline *et al.* [36], namely: *Node addition* (NA), *Node deletion* (ND), *Connection addition* (CA) and *Connection deletion* (CD); to these, *Node fusion* (NF) is added. For each mutation (except NF), there is a minimum value of 1 for adding or removing nodes and connections, and a maximum value of 2 or 6 for adding or removing nodes and connections, respectively.

In this hybrid implementation of the basis functions, when a new node should be added to the networks, it is necessary to estimate the probability for adding SU or PU. In this work, we have considered with equal probability the addition of a sigmoidal or product unit hidden node, thus the probability is 0.5. These probabilities have been determined by means of trial and error, and they are supported along the whole evolutionary process. Finally, the stop criterion is reached whenever one of the following two conditions is fulfilled: (i) the algorithm achieves a given number of generations; (ii) there is no improvement for a number of generations either in the average performance of the best 20% of the population or in the fitness of the best individual. For more details about the evolutionary algorithm, see references [27] and [28].

3. EXPERIMENTAL

The whole procedure followed in this work is schematically shown in Figure 2. To obtain analytical data, 55 electropherograms provided by samples containing amikacin (15–300 μg/L), paromomycin (10–200 μg/L) and mixtures of both antibiotics with uniformly distributed concentrations of them (30–300 and 20–200 μg/L, respectively) were prepared in duplicate, as described elsewhere [34], by using 35 mM sodium borate at pH 9.0 as background electrolyte.

The Levenberg–Marquardt algorithm was used to estimate the four-parameter Weibull function ($\hat{S}_m$, $\hat{B}$, $\hat{C}$ and $\hat{t}_m$) associated with the profile of the electrophoretic peaks (convergence of the iterative process was achieved with a tolerance of 0.0001 and a maximum number of 100 iterations), whereas LDA was used to classify these electrophoregrams in three categories: G_1 and G_2 for samples containing only amikacin

DOI: 10.1002/cem

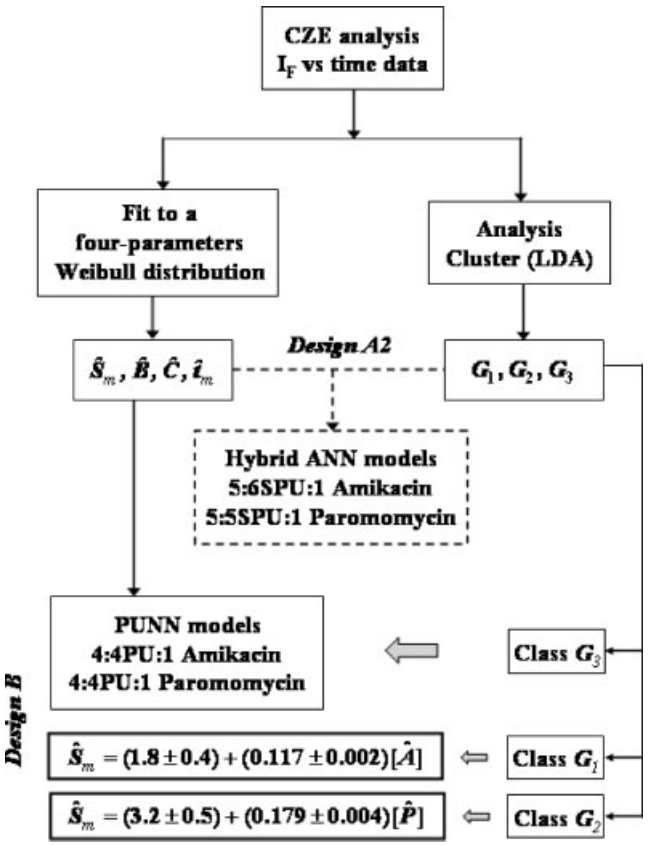

Figure 2. Flow diagram representing the whole analytical protocol.

or paromomycin, respectively and G_3 for samples containing mixtures of both antibiotics, which were generically denoted by $\hat{G}$. Prior to use these parameters as inputs for the ANN models, and in order to avoid saturation problems in the sigmoidal basis functions (preventing driving the weights to infinity) as well as to improve the learning process, each of the input (parameters of the Weibull curve and the category) and output (concentration of the antibiotics in the sample) were scaled in the range [0.1, 0.9]; this process was also useful for product basis functions because the lower bound is chosen to avoid inputs values near to zero that can produce very large values of the outputs for negative exponents, whereas the upper bound is chosen to avoid dramatic changes in the outputs of the network when there are weights with large values (especially in the exponents). Thus, the new scaled variables were expressed as follows:$\hat{S}_m^*$, $\hat{B}^*$, $\hat{C}^*$, $\hat{t}_m^*$ and $\hat{G}^*$ for the input variables and $[\hat{A}]^*$ and $[\hat{P}]^*$ for the output variables. For example $[\hat{A}]^*$ is calculated as follows:

$$[\hat{A}]^* = \frac{[A] - [A_{\min}]}{[A_{\max}] - [A_{\min}]} \times 0.8 + 0.1 \tag{10}$$

where $[A]$ is the original concentration of amikacin in the sample, $[A_{\min}]$ and $[A_{\max}]$ are the minimum and maximum values and $[\hat{A}]^*$ is the scaled concentration. After optimizing the network models, estimations should be de-scaled according to the same equation.

The experimental design was conducted using a holdout cross-validation procedure where the size of the training set was approximately $3n/4$ and $n/4$ for the generalization set, where $n = 110$ is the size of the full data set. The algorithm software for SU, PU and SPU neural networks was designed in Java using the JCLEC library [40] and was run on a portable PC Pentium IV compatible computer. The accuracy of each model was assessed in terms of the SEP for the results obtained for both data sets, that is SEP_T for the training set, and SEP_G for the generalization set. In this way, the SEP was calculated as:

$$\text{SEP} = \frac{100}{\overline{C}_i} \sqrt{\frac{\sum_{i=1}^{n} (C_i - \hat{C}_i)^2}{n}} \tag{11}$$

where C_i and $\hat{C}_i$ are the experimental and expected values for the antibiotic concentration in the mixture, $\overline{C}_i$ is the mean of the experimental values of the training set, or of the generalization set and n is the number of patterns used (n_T for the training set and n_G for the generalization set). Finally, the partial least squares (PLS) multivariate calibration algorithm was provided by the Pirouette 3.11 software from Infometrix, Inc.

4. RESULTS AND DISCUSSION

In recent years, CZE is increasingly being considered as an alternative to HPLC for the determination of a great variety of compounds because it affords such significant benefits as higher resolution, smaller sample requirements and shorter analysis time. Despite this better resolution, CZE separation is sometimes not accomplished and therefore chemometric resolution is the suitable choice. Thus, it is understandable that the quantitative resolution of overlapping electrophoretic peaks is an area of growing interest for analytical chemists. ANNs have scarcely been used for this purpose despite their higher discrimination power regarding other current chemometric tools, even when using a smaller amount of chemical information [16,17]. The goal of this work was to evaluate different ANN models supported on the use of three different transfer functions, namely SU, PU and the hybrid model: SPU, for the quantification of unresolved CZE peaks with a high degree of overlapping by using analytical data provided by a single detector (LIF detection), that is in the absence of spectral discrimination. The approach was tested on the determination of amikacin (A) and paromomycin (P), two aminoglycoside antibiotics that cannot be resolved by CZE. After labeling the antibiotics with sulfoindocyanine succinimidyl ester (Cy5), electrophoretic data were obtained by monitoring the single LIF signal with a diode-laser detector. In order to extend the scope of the proposed methodology to more practical situations, in addition to the typical approach based on samples containing mixtures of both antibiotics, samples with only one antibiotic (A or P) have also been considered. It is noteworthy that the CZE peak showed a similar profile in all cases.

4.1. Selection of data and inputs for ANN models

One important issue to be addressed in order to obtain the best generalization capability of ANN models is the composition of the data set used for training and general-

DOI: 10.1002/cem

ization. Two different data sets were tested from all the data (110 samples) obtained by preparing 55 individual and mixture synthetic samples in duplicate. Design A was constituted by the whole group of the samples (individual and mixtures) and design B only contains the samples with mixtures of antibiotics. In both cases, the size of the training and generalization sets was 3/4 and 1/4, respectively, of the whole set: 110 and 72 samples for design A and B, respectively. The so-called grid approach [41] was used for sampling mixtures including those with higher and smaller antibiotic ratios in the training set for assuring interpolation in the generalization process. This state was also considered for individual samples.

The selection of the multivariate input data for ANN models is also of great relevance because they may cause over-fitting of the trained networks. A small-sized neural network will be less prone to overtraining noise or the structure of the data in the training set, thus increasing its generalization capacity over a new data set and allowing users to gain knowledge from the trained neural network in order to achieve a better understanding of how the network solves the problem. As an alternative to principal component analysis (PCA), the most common choice used for reducing the data set, we have recently reported a useful approach based on the modeling of the chromatographic profile to a pre-determined function [16,17].

The electropherograms of pure amikacin, paromomycin and a mixture of equal concentrations of both antibiotics are shown in Figure 3A. As can be seen, the profiles of the CZE peaks were very similar, which makes clear the high overlapping grade of the bands of both antibiotics in mixtures samples. Although the precision (expressed as relative standard deviation) for the migration times in these electropherograms was within 1% (similar to that reported by us in other works using CZE with LIF detection) in order to improve their reproducibility, migration times were corrected using as reference one peak of the excess of Cy5 label.

The non-symmetric shape of these profiles suggests that they can be modeled with a pre-determined function, being the four-parameter Weibull distribution the best choice because it provides the best regression coefficients in the fitting. As stated above, the equation corresponding to this function is defined by the following parameters: $\hat{S}_m$ (peak height), $\hat{B}$ (dispersion of S_{t_i} values from S_m), $\hat{C}$ (associated to the inflection points of the curve) and $\hat{t}_m$ (residence time). Figure 3B shows the fit provided by the single four-parameter Weibull distribution on the CZE peak profiles. From these plots, it can be inferred that it was impossible to fix a common residence time interval to select the portion of the CZE profile subjected to fit for all samples. However, the fit can be easily achieved by using the following process: (1) a residence time interval of 10 data around the maximum was selected to start the fit and (2) this interval was successively increased by two data, one to every side of the interval. The number of data (the portion of the CZE profile) that provided the best fit, in terms of regression coefficient, was used to estimate the four parameters of the Weibull distribution. Upon examining the curves and from the estimated statistical parameters shown in Table I, it can be concluded that the four-parameter Weibull function is a fine tool for modeling these CZE data.

In addition, a cluster analysis was also carried out using a cross-validation holdout method (72 patterns for the training set and 38 for generalization set) adopting these four-parameters as variables for LDA. As can be seen in Table II, the confusion matrixes provide a total correct classification rate (CCR) value of 100 for both training and generalization sets; this table also includes the standardized coefficients for the canonical discriminant functions. From these results, the following labels were assigned to the different classes: $\hat{G}_1 = 1$

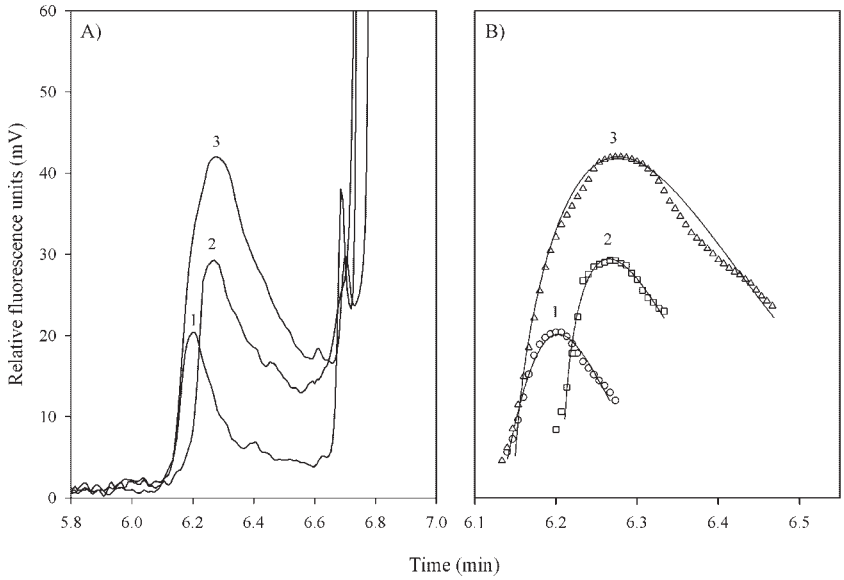

Figure 3. (A) Typical electropherograms corresponding to pure and mixture solutions of the assayed antibiotics: (1) 120 μg/L of amikacin; (2) 180 μg/L of paromomycin and (3) mixture containing 120 μg/L of amikacin and 180 μg/L of paromomycin. (B) CZE responses fitted to a four-parameter Weibull distribution (o) Experimental data and (–) Weibull curve.

DOI: 10.1002/cem

Table I. Estimated parameters obtained in the modelling process of the electrophoretic peaks shown in Figure 3 by means of a Weibull distribution

	Parameters[a]				
Sample	$\hat{S}_m$	$\hat{B}$	$\hat{C}$	$\hat{t}_m$	r
Pure amikacin	20.1 ± 0.6	0.106 ± 0.004	1.9 ± 0.1	6.203 ± 0.002	0.9889
Pure paromomycin	28.8 ± 1.6	0.15 ± 0.02	1.5 ± 0.2	6.275 ± 0.008	0.9540
Mixture of antibiotics	42.0 ± 0.8	0.245 ± 0.009	1.57 ± 0.08	6.277 ± 0.006	0.9855

[a] Mean $\pm 1.96 \times$ SD; SD = standard deviation; r = regression coefficient.

and $\hat{G}_2 = 2$ for pure amikacin and paromomycin samples, respectively and $\hat{G}_3 = 3$ for samples containing mixtures of the antibiotics. These labels were also used in some cases as inputs in addition to the Weibull parameters.

4.2. Evaluation of the ANN models

To compare the predictive ability of ANN models in terms of topology, number of connections, homogeneity (confidence interval) and accuracy (SEP_G), network models with a single node in the output layer were designed (concentration of antibiotic to be determined in the individual or mixture sample). When using experimental design A, 4:5:1 and 5:6:1 architectures were chosen to start the model selection processes, whereas a 4:4:1 topology was used for design B, in all cases over 30 runs.

As can be seen in Table III, all models provided satisfactory results in terms of accuracy (SEP_G) and homogeneity (confidence interval) for determining the concentration of each antibiotic in the samples assayed. On comparing the two designs, higher SEP values were achieved by ANN models based on design A1 that uses all samples for training and generalization sets without discrimination among them through cluster analysis. In this case, only the four-parameters of the Weibull curve were used as inputs and the computational cost was clearly lower. When the cluster analysis was carried out, two approaches have been considered (see Figure 2): using the class label as input in addition to the Weibull parameters (design A2) or employing this information to create a new data set (design B) containing only the mixture samples. Thus in the design B, only the Weibull parameters were used as inputs, and the determination of the concentrations of the antibiotic in the individual samples was supported on the construction of the classical calibration plot, such as $\hat{S}_m$ versus [antibiotic] expressed in μg/L. From experimental results, the following calibration plots can be drawn:

$$\hat{S}_m = (1.8 \pm 0.4) + (0.117 \pm 0.002)\,[A], \qquad r = 0.9965 \quad (12)$$

$$\hat{S}_m = (3.2 \pm 0.5) + (0.179 \pm 0.004)[P], \qquad r = 0.9951 \quad (13)$$

As can be seen in Table III, ANN models based on design B provided lower SEP_G values, perhaps due to the biggest homogeneity of the data sets because only mixture samples were used for both training and generalization sets. However, the use of design A2 cannot be discarded a priori because it can be an acceptable choice depending on the analytical problem addressed (higher SPG_G values vs. lower computational cost).

Regarding the type of ANN models tested, those based on PU and SPU as transfer functions provided similar results for the design A2, whereas models of PUNN provided lower SPE_G values for design B, especially for the determination of *P*. This behavior is closely related to the composition of the data set. In fact, the design A2 composed by individual (linear) samples together with mixtures (nonlinear) is better modeled by ANN using hybrid transfer functions, whereas in the case of the design B (in the absence of individual samples), PU transfer functions yielded a better modeling of the chemical information.

In order to compare the quality achieved in each ANN model for the resolution of mixtures of these antibiotics, Table IV shows the results (in terms of concentration) obtained applying the models to the synthetic mixtures included in the generalization set of design B (see Figure 2). In addition, this table also shows the results achieved when the data were analyzed by using a classical standard reference method such as PLS regression. For the construction of the different PLS models, the cross-validation method was used and the number of significant factors in each case were chosen as the lower number whose root mean standard error (RMSE) of prediction by cross-validation was not significantly different from the lowest RMSE value:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(C_i - \hat{C}_i\right)^2}{n}} \quad (14)$$

Based on the absolute error of the prediction $\left|C_i - \hat{C}_i\right|$ obtained using the data reported by the ANN and PLS

Table II. Summary of the results obtained in the LDA of the whole group of the samples (individual and mixtures) by using the estimated four-parameters of the Weibull function as variables

	Training				Generalization			
TR/PR	$G_1=1$	$G_2=2$	$G_3=3$	FC %	$G_1=1$	$G_2=2$	$G_3=3$	FC %
A. Rate of the number of cases that were classified correctly								
$G_1=1$	12	0	0	100	6	0	0	100
$G_2=2$	0	12	0	100	0	6	0	100
$G_3=3$	0	0	48	100	0	0	24	100
CCR				100				100

Variable	$\hat{S}_m$	$\hat{B}$	$\hat{C}$	$\hat{t}_m$
B. Standardized coefficients for the canonical discriminant functions				
Coefficients of function 1	−0.374	0.600	2.414	2.437
Coefficients of function 2	0.810	−0.387	0.246	0.503

PR = Predicted response; TR = Target response; FC = False correct.

DOI: 10.1002/cem

Table III. Accuracy (mean $\pm$ 1.96 $\times$ SD) and statistical results of the algorithm used with different experimental designs and transfer functions (over 30 runs)

		Connections	SEP_T			SEP_G		
Analyte	Starting topology	Mean $\pm$ 1.96 $\times$ SD	Mean $\pm$ 1.96 $\times$ SD	Best	Worst	Mean $\pm$ 1.96 $\times$ SD	Best	Worst
Design A1								
A	4:5SU:1	26.5 $\pm$ 3.4	16.7 $\pm$ 9.9	10.1	31.7	19.4 $\pm$ 11.0	10.2	36.8
	4:5PU:1	18.3 $\pm$ 3.0	14.1 $\pm$ 3.6	11.0	18.3	22.0 $\pm$ 14.1	13.2	38.8
P	4:5SU:1	26.1 $\pm$ 3.7	21.7 $\pm$ 3.7	17.6	25.1	18.0 $\pm$ 5.0	14.2	24.1
	4:5PU:1	18.3 $\pm$ 2.9	20.4 $\pm$ 4.1	16.4	24.9	18.6 $\pm$ 6.6	13.5	27.5
Design A2								
A	5(4 + 1):6SU:1	33.6 $\pm$ 4.2	12.7 $\pm$ 5.3	8.2	17.8	16.0 $\pm$ 6.8	10.0	23.6
	5(4 + 1):6PU:1	23.8 $\pm$ 7.3	11.8 $\pm$ 3.9	8.9	16.1	20.4 $\pm$ 10.6	9.7	32.9
	5(4 + 1):6SPU:1	33.3 $\pm$ 4.7	12.9 $\pm$ 4.3	10.1	18.6	19.7 $\pm$ 11.8	9.1	38.9
P	5(4 + 1):6SU:1	34.8 $\pm$ 5.4	16.7 $\pm$ 5.0	12.1	21.4	16.1 $\pm$ 3.6	13.2	20.1
	5(4 + 1):6PU:1	23.2 $\pm$ 5.9	15.6 $\pm$ 4.8	10.7	20.4	18.2 $\pm$ 7.1	12.3	27.1
	5(4 + 1):6SPU:1	32.5 $\pm$ 6.1	13.6 $\pm$ 4.2	9.4	18.1	16.6 $\pm$ 7.2	10.9	27.7
Design B								
A	4:4SU:1	20.1 $\pm$ 3.8	7.2 $\pm$ 1.6	6.1	10.0	10.5 $\pm$ 3.2	8.4	15.3
	4:4PU:1	13.8 $\pm$ 3.1	7.0 $\pm$ 1.3	5.8	8.5	10.5 $\pm$ 2.2	8.2	13.1
	4:4SPU:1	19.2 $\pm$ 4.3	7.2 $\pm$ 1.9	5.4	8.9	10.2 $\pm$ 2.4	8.4	13.0
P	4:4SU:1	19.2 $\pm$ 4.9	8.3 $\pm$ 2.5	5.8	11.4	13.0 $\pm$ 6.7	6.5	22.2
	4:4PU:1	13.6 $\pm$ 4.4	8.8 $\pm$ 3.1	5.8	11.4	12.5 $\pm$ 8.8	5.6	22.6
	4:4SPU:1	19.2 $\pm$ 3.8	8.4 $\pm$ 3.2	6.0	12.9	12.3 $\pm$ 7.5	6.4	20.1

models in Table IV, the analysis of variance (ANOVA) technique was carried out to ascertain the statistical significance of observed differences among the four corresponding means, assuming that the absolute error values obtained have a normal distribution. A Kolmogorov–Smirnov test for normality was reached with *p*-values of 0.383, 0.389, 0.866 and 0.143 for amikacin and 0.538, 0.426, 0.235 and 0.885 for paramomycin, for SUNN, PUNN, SPUNN and PLS models, respectively. The ANOVA involved a linear regression model in which $\left|C_i - \hat{C}_i\right|$ was

Table IV. Comparison of the quality achieved for the quantification of amikacin and paromomycin mixtures using the ANN models and the PLS methodology

	Amikacin					Paromomycin				
		Found (μg/L)					Found (μg/L)			
$[\hat{A}]/[\hat{P}]$	Added (μg/L)	SUNN	PUNN	SPUNN	PLS	Added (μg/L)	SUNN	PUNN	SPUNN	PLS
8:1	240.0	231.9	244.8	241.2	238.2	30.0	16.2	26.1	18.9	10.2
8:1	240.0	240.9	253.8	250.5	248.1	30.0	17.7	27.6	21.6	9.3
1:2	30.0	23.1	27.9	25.8	7.8	60.0	75.9	66.6	77.4	82.2
1:2	30.0	31.5	33.6	35.7	30.0	60.0	66.0	61.2	71.7	73.8
4:2	120.0	102.6	103.5	104.1	118.5	60.0	71.1	72.3	75.0	87.6
4:2	120.0	108.3	110.4	108.0	122.4	60.0	71.4	71.7	72.9	94.2
6:2	180.0	173.4	176.7	171.9	190.5	60.0	61.5	64.8	60.9	88.8
6:2	180.0	170.4	174.9	167.1	185.4	60.0	64.5	66.9	62.4	98.4
8:2	240.0	225.6	226.5	222.6	241.8	60.0	55.8	61.8	57.3	91.5
8:2	240.0	234.3	235.2	231.9	249.9	60.0	57.6	63.3	59.7	85.5
1:4	30.0	40.2	47.7	39.3	32.1	120.0	111.9	102.0	116.7	112.2
1:4	30.0	38.7	46.2	42.0	39.0	120.0	118.2	109.2	127.8	119.4
6:4	180.0	195.9	197.4	201.3	213.9	120.0	103.5	105.9	102.6	111.3
6:4	180.0	196.2	196.8	200.7	212.4	120.0	112.2	115.8	110.7	113.4
4:6	120.0	115.5	119.1	115.5	122.4	180.0	178.5	177.6	178.2	185.7
4:6	120.0	114.6	117.3	121.5	129.0	180.0	182.7	180.6	185.1	192.9
6:6	180.0	166.8	163.5	179.7	186.3	180.0	176.7	179.4	177.9	182.7
6:6	180.0	174.0	169.8	192.3	198.9	180.0	167.1	167.7	168.9	174.6
4:8	120.0	115.5	115.8	118.5	124.8	240.0	244.8	237.0	243.3	255.3
4:8	120.0	112.8	112.2	117.6	125.4	240.0	251.7	241.5	250.5	263.1
10:8	300.0	264.9	273.0	268.8	252.6	240.0	254.4	239.7	251.1	270.6
10:8	300.0	270.3	275.4	277.8	260.1	240.0	246.3	237.3	243.0	260.4
4:10	120.0	130.2	116.4	120.9	137.1	300.0	296.4	310.2	300.0	273.0
4:10	120.0	124.2	111.0	117.0	134.1	300.0	297.9	307.5	301.5	277.2

DOI: 10.1002/cem

Table V. Quantification equations and accuracy provided by the optimized 4:4:1 PUNN network topologies as applied to the determination of amikacin and paramomycin

	Amikacin	Paramomycin
Quantification equations	$[\hat{A}]^* = 0.52 + 2.29\,\hat{h}_1 - 1.28\,\hat{h}_2 + 0.36\,\hat{h}_3 + 2.80\,\hat{h}_4$	$[\hat{P}]^* = 0.25 - 1.97\,\hat{h}_1 + 3.40\,\hat{h}_2 - 2.07\,\hat{h}_3 - 0.06\,\hat{h}_4$
Transfer functions	$\hat{h}_1 = (\hat{S}^*_m)^{0.39}(\hat{B}^*)^{0.50}$	$\hat{h}_1 = (\hat{S}^*_m)^{0.42}(\hat{B}^*)^{0.46}(\hat{C}^*)^{0.46}(\hat{t}^*_m)^{0.06}$
	$\hat{h}_2 = (\hat{S}^*_m)^{0.92}(\hat{B}^*)^{0.52}(\hat{C}^*)^{0.86}$	$\hat{h}_2 = (\hat{S}^*_m)^{1.14}(\hat{C}^*)^{0.46}$
	$\hat{h}_3 = (\hat{S}^*_m)^{2.11}(\hat{t}^*_m)^{0.45}$	$\hat{h}_3 = (\hat{C}^*)^{1.63}(\hat{t}^*_m)^{4.91}$
	$\hat{h}_4 = (\hat{B}^*)^{3.87}(\hat{C}^*)^{1.49}$	$\hat{h}_4 = (\hat{B}^*)^{3.13}$
Effective links	14	14
SEP_T	5.8%	5.8%
SEP_G	8.2%	5.6%

the dependent variable and the independent variable was the type of model used for prediction. The comparisons were made in terms of a critical level for Snedecor's F; if the significance level, α, was higher than this critical level, p, the hypothesis of identical means was rejected. In the case of amikacin, this hypothesis was not accepted because the p-value is 0.749, higher than a standard $\alpha = 0.05$, whereas in the case of paramomycin, it was accepted because the p-value was 0.000. Based on these results, a test of multiple comparisons of Tamhane was carried out for paramomycin. Significant differences were found between PLS and ANN models according to the p-values provided by the test: 0.000 in all comparisons. In view of these results, we selected the PUNN model on the basis of its lower SEP_G values and simpler architecture (smaller effective links) than SUNN and SPUNN models.

The simplicity of the proposed ANN models permits us to derive straightforward quantitative equations for the determination of the concentration of each antibiotic using: (a) the parameters estimated by the Weibull regression of the peak and the class label provided by the cluster analysis; (b) the optimized network weights and (c) the transfer functions involved in the models. Table V shows the quantitative equations corresponding to the proposed PUNN models for both antibiotics.

4.3. Chemical interpretation of ANN models

Taking into account that the proposed PUNN models provided simple quantitative equations for the direct determination of the contribution of each antibiotic to the overlapping CZE peaks, quality chemical information could be derived in order to explain the relationship between the profile of the CZE peak (defined by the Weibull parameters) and the determination of the antibiotic in the mixture by using the proposed PUNN models. In this way, the value of each PU transfer function involved that was affected by its coefficient in the model was calculated over the scaled range studied for the input variables, $\hat{S}^*_m$, $\hat{B}^*$, $\hat{C}^*$ and $\hat{t}^*_m$, and then plotted one against another (see Figure 4).

4.3.1. *Model for amikacin*

From the quantification equation shown in Table V and from the plots in Figure 4A, it follows that the value of $[\hat{A}]^*$ depends mainly on the contribution of the PU functions $\hat{h}_1$ and $\hat{h}_2$ with opposite effects and also on the other PU

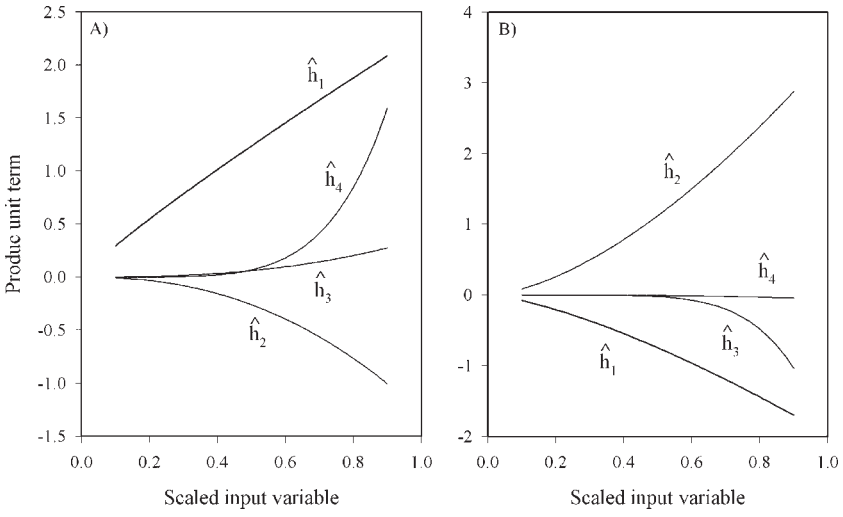

Figure 4. Relative contribution of the product unit terms used for the quantitative determination of the antibiotics provided by PUNN models. (A) amikacin and (B) paromomycin.

DOI: 10.1002/cem

functions but only at higher values of the scaled input variable: for example the response of the function $\hat{h}_4$ at higher values of $\hat{B}^*$ and $\hat{C}^*$ is assigned to mixtures with higher values of the [*A*] to [*P*] ratio. In short, this function made the model provide more accurate results when modeling this type of samples. It is also noteworthy that the contribution of $\hat{t}_m^*$ was practically negligible because it was only involved in the function $\hat{h}_3$ which, on the other hand, is the one that least contributes to the value of $[\hat{A}]^*$. From these results, the direct dependencies of $\hat{S}_m^*$ and $\hat{B}^*$ parameters on $[\hat{A}]^*$ is clear, that is the value of the peak height and the degree of dispersion of the analytical signal values from it, and both taking into account their respective exponents in $\hat{h}_1$ to a similar extent.

4.3.2. *Model for paromomycin*

By using the PUNN model for the determination of *P* in the mixture, and from the quantification equation shown in Table V and plots in Figure 4B, it follows that the value of $[\hat{P}]^*$ depends basically on the $\hat{S}_m^*$ and $\hat{C}^*$ parameters, considering the dependencies showed by the $\hat{h}_1$ and $\hat{h}_2$ functions in Figure 4B. According to the relative values of the exponents in the $\hat{h}_1$ (basis function), a higher dependence can be ascribed to $\hat{C}^*$ parameter, which is associated to the inflection points of the curve and defines the concavity and convexity regions of the CZE peaks. Regarding the other PU functions, only $\hat{h}_3$ exerts a slight contribution at higher values of the scaled input variable, which is related to mixtures with high [*P*] to [*A*] ratios; as in the case of *A*, this function is of great relevance for achieving a better modeling of these mixtures by the proposed PUNN model.

5. CONCLUSIONS

As shown in this work, quantitative analysis in CZE is possible even in the case of overlapped peaks by means of ANNs designed by an evolutionary algorithm. Three ANN strategies were evaluated for modeling the CZE data provided from a set of samples containing single and mixtures of analyte concentrations. The four parameters of the Weibull curve fitted to the profile of the CZE peaks were suitable as inputs for the three types of ANN, although the predicted results can be improved with a prior LDA analysis and using in some case the class label associated to each sample as additional input. The calculated results indicated that ANN models with SU, PU and SPU units as transfer functions were a promising tool to resolve overlapped CZE peaks with acceptable errors. The designed PUNN models provided better accurate results, smaller network architectures and more robust models and they were quite simple and easier to interpret from a chemical point of view. In summary, ANN is a powerful tool for resolving CZE overlapped peaks while also allowing an important reduction in analysis time because it avoids time consumption by finding optimal conditions for the suitable CZE resolution.

Acknowledgements

The authors gratefully acknowledge the subsidy provided by the Spanish Inter-Ministerial Commission of Science and Technology of the Ministry of Education and Science under the CTQ2007-63962 (Department of Analytical Chemistry, University of Cordoba) and TIN2005-08386-C05-02 (Department of Computer Science, University of Cordoba) Projects. FEDER also provided additional funding. The research of Pedro Antonio Gutierrez has been supported by FPU-Predoctoral Program (Spanish Ministry of Education and Science).

REFERENCES

1. van-Zomeren PV, Metting HJ, Coenegracht PMJ, de-Jong GJ. Simultaneous resolution of overlapping peaks in high-performance liquid chromatography and micellar electrokinetic chromatography with diode array detection using augmented iterative target transformation factor analysis. *J. Chromatogr. A* 2005; **1096**: 165–176.
2. Vivo-Truyols G, Torres-Lapasio JR, van-Nederkassel AM, vander-Heyden Y, Massart DL. Automatic program for peak detection and deconvolution of multi-overlapped chromatographic signals. Part II: Peak model and deconvolution algorithms. *J. Chromatogr. A* 2005; **1096**: 146–155.
3. Rodriguez-Cuesta MJ, Boque R, Rius FX, Vidal JLM, Frenich AG. Development and validation of a method for determining pesticides in groundwater from complex overlapping HPLC signals and multivariate curve resolution. *Chemometr. Intell. Lab. Syst.* 2005; **77**: 251–260.
4. Wiberg K, Jacobsson SP. Parallel factor analysis of HPLC-DAD [diode-array detection] data for binary mixtures of lidocaine and prilocaine with different levels of chromatographic separation. *Anal. Chim. Acta* 2004; **514**: 203–209.
5. Comas E, Gimeno RA, Ferre J, Marce RM, Borrull F, Rius FX. Quantification from highly drifted and overlapped chromatographic peaks using second-order calibration methods. *J. Chromatogr. A* 2004; **1035**: 195–202.
6. van-Zomeren PV, Darwinkel H, Coenegracht PMJ, de Jong GJ. Comparison of several curve resolution methods for drug impurity profiling using high-performance liquid chromatography with diode-array detection. *Anal. Chim. Acta* 2003; **487**: 155–170.
7. Gross GM, Prazen BJ, Synovec RE. Parallel column liquid chromatography with a single multi-wavelength absorbance detector for enhanced selectivity using chemometric analysis. *Anal. Chim. Acta* 2003; **490**: 197–210.
8. Fraga CG, Bruckner CA, Synovec RE. Increasing the number of analyzable peaks in comprehensive two-dimensional separations through chemometrics. *Anal. Chem.* 2001; **73**: 675–683.
9. Sanchez FC, Rutan SC, Garcia MDG, Massart DL. Resolution of multicomponent overlapped peaks by the orthogonal projection approach, evolving factor analysis and window factor analysis. *Chemometr. Intell. Lab. Syst.* 1997; **36**: 153–164.
10. Zhang F, Li H. Resolution of overlapping capillary electrophoresis peaks by using chemometric analysis: improved quantification by using internal standard. *Chemometr. Intell. Lab. Syst.* 2006; **82**: 184–192.
11. Zhang F, Li H. Resolution of overlapping capillary electrophoresis peaks by using chemometric analysis: quantification of the components in compound reserpine tablets. *Electrophoresis* 2005; **26**: 1692–1702.
12. Li H, Zhang F, Havel J. Quantification of analytes in overlapping peaks from capillary electrophoresis using multivariate curve resolution-alternating least squares methods. *Electrophoresis* 2003; **24**: 3107–3115.
13. Olazabal V, Prasad L, Stark P, Olivares JA. Application of wavelet transforms and an approximate deconvolution

DOI: 10.1002/cem

method for the resolution of noisy overlapped peaks in DNA capillary electrophoresis. *Analyst* 2004; **129**: 73–81.
14. Chen AJ, Li CH, Gao WH, Hu ZD, Chen XG. Separation and determination of active components in *Schisandra chinensis* Baill. and its medicinal preparations by non-aqueous capillary electrophoresis. *Biomed. Chromatogr.* 2005; **19**: 481–487.
15. Chen AJ, Li CH, Gao WH, Hu ZD, Chen XG. Application of non-aqueous micellar electrokinetic chromatography to the analysis of active components in radix *Salviae miltiorrhizae* and its medicinal preparations. *J. Pharm. Biomed. Anal.* 2005; **37**: 811–816.
16. Hervás C, Silva M, Serrano JM, Orejuela E. Heuristic extraction of rules in pruned artificial neural networks models used for quantifying highly overlapping chromatographic peaks. *J. Chem. Inf. Comput. Sci.* 2004; **44**: 1576–1584.
17. Hervás C, Martínez AC, Silva M, Serrano JM. Improving the quantification of highly overlapping chromatographic peaks by using product unit neural networks modeled by an evolutionary algorithm. *J. Chem. Inf. Model.* 2005; **45**: 894–903.
18. Garrido-Frenich A, Martinez-Galera M, Gil-Garcia MD, Martinez-Vidal JL, Catasus M, Marti L, Mederos MV. Resolution of HPLC-DAD highly overlapping analytical signals for quantitation of pesticide mixtures in groundwater and soil using multicomponent analysis and neural networks. *J. Liq. Chromatogr. Relat. Technol.* 2001; **24**: 651–668.
19. Li YB, Huang XY, Sha M, Meng XS. Resolution of overlapping chromatographic peaks by radial basis function neural network. *Sepu* 2001; **19**: 112–115.
20. Galeano-Diaz T, Guiberteau A, Ortiz JM, Lopez MD, Salinas F. Use of neural networks and diode-array detection to develop an isocratic HPLC method for the analysis of nitrophenol pesticides and related compounds. *Chromatographia* 2001; **53**: 40–46.
21. Zhang YX, Li H, Hou AX, Havel J. Artificial neural networks based on principal component analysis input selection for quantification in overlapped capillary electrophoresis peaks. *Chemometr. Intell. Lab. Syst.* 2006; **82**: 165–175.
22. Zhang F, Li H. Resolution of overlapping capillary electrophoresis peaks by using chemometric analysis: quantification of the components in compound reserpine tablets. *Electrophoresis* 2005; **26**: 1692–1702.
23. Zhang YX, Li H, Hou AX, Havel J. Artificial neural networks based on genetic input selection for quantification in overlapped capillary electrophoresis peaks. *Talanta* 2005; **65**: 118–128.
24. Sentellas S, Saurina J, Hernandez-Cassou S, Galceran MT, Puignou L. Quantitation in multianalyte overlapping peaks from capillary electrophoresis runs using artificial neural networks. *J. Chromatogr. Sci.* 2003; **41**: 145–150.
25. Bocaz-Beneventi G, Latorre R, Farkova M, Havel J. Artificial neural networks for quantification in unresolved capillary electrophoresis peaks. *Anal. Chim. Acta* 2002; **452**: 47–63.
26. Schmitt M. On the complexity of computing and learning with multiplicative neural networks. *Neural Comput.* 2001; **14**: 241–301.
27. Martínez-Estudillo AC, Martínez-Estudillo FJ, Hervás-Martínez C, García-Pedrajas N. Evolutionary product unit based neural networks for regression. *Neural Netw.* 2006; **15**: 477–486.
28. Martínez-Estudillo AC, Hervás-Martínez C, Martínez-Estudillo FJ, García-Pedrajas N. Hybridation of evolutionary algorithms and local search by means of a clustering method. *IEEE Trans. Syst. Man Cybern. B Cybern.* 2006; **36**: 534–546.
29. Duch W, Jankowsky N. *Transfer functions: hidden possibilities for better neural networks*, in 9th European Symposium on Artificial Neural Networks, (ESANN), Brugge (Belgium), 2001, pp. 81–94.
30. Cohen S, Intrator N. *Forward and backward selection in regression hybrid network*, in Third International Workshop on Multiplier Classifier Systems, 2002.
31. Cohen S, Intrator N. A hybrid projection-based and radial basis function architecture: initial values and global optimization. *Pattern Anal. Appl.* 2002; **5**: 113–120.
32. Donoho D. Projection based in approximation and a duality with kernel methods. *Ann. Statist.* 1989; **17**: 58–106.
33. Friedman J. Multivariate adaptive regression splines (with discussion). *Ann. Statist.* 1991; **19**: 1–141.
34. Serrano JM, Silva M. Trace analysis of aminoglycoside antibiotics in bovine milk by micellar electrokinetic chromatography with laser induced fluorescence detection. *Electrophoresis* 2006; **27**: 4703–4710.
35. Duda RO, Hart PE, Stork DG. *Pattern Classification* (2nd edn). Wiley-Interscience: New York, USA, 2001.
36. Angeline PJ, Saunders GM, Pollack JB. An evolutionary algorithm that constructs recurrent neural networks. *IEEE Trans. Neural Netw.* 1994; **5**: 54–65.
37. Yao X, Liu Y. A new evolutionary system for evolving artificial neural networks. *IEEE Trans. Neural Netw.* 1997; **8**: 694–713.
38. García N, Hervás C, Muñoz J. Covnet: cooperative coevolution of neural networks. *IEEE Trans. Neural Netw.* 2003; **14**: 575–596.
39. Kirkpatrick S Jr, Gelatt CD, Vecchi MP. Optimization by simulated annealing. *Science,* 1983; **220**: 671–680.
40. Ventura S, Romero C, Zafra A, Delgado JA, Hervás C. JCLEC: A Java framework for evolutionary computation. *Soft. Comput. D 2007. Available in http://dx.doi.org/10.1007/s00500-007-0172-0*
41. U.S. Environmental Protection Agency. *Guidance on Choosing a Sampling Design for Environmental Data Collection*. EPA: Washington, USA, 2002.

DOI: 10.1002/cem