Improving the Quantification of Highly Overlapping Chromatographic Peaks by Using Product Unit Neural Networks Modeled by an Evolutionary Algorithm

César Hervás and Alfonso Carlos Martínez

Department of Computer Science, Albert Einstein Building, Rabanales Campus, University of Cordoba,
E-14071 Cordoba, Spain

Manuel Silva* and Juan Manuel Serrano

Department of Analytical Chemistry, Marie Curie Building (Annex), Rabanales Campus,
University of Cordoba, E-14071 Cordoba, Spain

This work investigates the ability of multiplicative (on the basis of product units) and sigmoidal neural models built by an evolutionary algorithm to quantify highly overlapping chromatographic peaks. To test this approach, two *N*-methylcarbamate pesticides, carbofuran and propoxur, were quantified using a classic peroxyoxalate chemiluminescence reaction as a detection system for chromatographic analysis. The four-parameter Weibull curve associated with the profile of the chromatographic peak estimated by the Levenberg–Marquardt method was used as input data for both models. Straightforward network topologies (one output) allowed the analytes to be quantified with great accuracy and precision. Product unit neural networks provided better information ability, smaller network architectures, and more robust models (smaller standard deviation). The reduced dimensions of the selected models 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 sigmoidal neural networks, and the effect of both analytes on the characteristics of chromatographic bands, namely profile, dispersion, peak height, and residence time, can be readily established.

INTRODUCTION

Over the past decade, increased attention has been paid to the applications of "soft" models in chemistry. Soft models can be defined as approaches in which an explicit mathematical model is neither formulated nor used. The prime example of such an approach is the use of artificial neural networks (ANNs), which can be applied to various problems such as spectra modeling, analyte concentration prediction, compound classification, or multivariate calibration, among others.[1–8] Multilayer perceptron (MLP) modeling is often found to be more efficient than other methods, but it suffers from the perception of being a "black box" heuristic tool. This capability is supported by the fact that a neural network with sigmoidal transfer functions can approximate any continuous function with a prescribed degree of accuracy by determining the number of neurons in the hidden layer.[9–11] Yet, the approximation of a relationship remains a problem, especially when the information provided by the instrumentation is too unwieldy to model, as many inputs do not contain important information, such as in separation science (chromatographic analysis). Furthermore, limitations of the universal approximation capabilities of sigmoidal networks have been recently reported on the basis that the number of summation units required to approximate continuous functions can be prohibitive.[12] To overcome this difficulty, so-called product unit neural networks (PUNNs) can be a valuable alternative. These are a type of multiplicative neural

model introduced by Durbin and Rumelhart,[13] in which the inputs are multiplied after they have been raised to some power specified by an adjustable weight. Neurons with multiplicative responses are extremely powerful computational elements in neural networks and for more biologically motivated models,[14] which is supported on the basis of the super-additive principle: the merger of several sensorial factors provides more information than the factors do separately.[15] These models express multivariate polynomial equations, and it is possible that available domain knowledge will be straightforwardly embedded within the neural networks, and we can easily interpret the results.[16] In short, so far, MLP modeling has suffered from difficulties in training models, and therefore, new strategies are required to find networks with reduced architecture, which demonstrate a better generalization ability, in order to enhance the knowledge extraction from the trained models, which is important in order to gain the users' confidence and acceptance.

During the past decade, ANNs have been increasingly applied in analytical chemistry, and in separation science as well. Optimization in separation sciences is still an important demand from analysts who are looking for a certain resolution or selectivity with a limited number of experiments in minimum time.[17] In this context, ANN models have been applied for peak tracking[18] and response surface modeling[19] in high-performance liquid chromatography (HPLC) optimization, assessment of chromatographic peak purity,[20] or for deconvolution of overlapping peaks.[21] Although the use of ANNs for the resolution of overlapping chromatographic

* Corresponding author phone: +34-957-212099; e-mail: qa1sirom@uco.es.

QUANTIFICATION OF OVERLAPPING PEAKS

J. Chem. Inf. Model., Vol. 45, No. 4, 2005 **895**

analytical signals has gained popularity over the past years, few publications have dealt with this subject and some difficulties have yet to be resolved.[22−25]

In this paper, we propose a new ANN model, based on PUNNs, as a powerful tool to achieve the desired selectivity in chromatographic analysis, such as the quantification of analytes that provide highly overlapping chromatographic peaks obtained from a single detector instrument and also the extraction of quality information about the chemical problem being addressed. Although we recently dealt with this problem using classical sigmoidal networks pruned by a regularization method and with architectures designed by a real genetic code algorithm,[25] in order to evaluate the real ability of PUNN models versus classical MLP ones, a comparison was made between both models, in this case by using the same evolutionary algorithm for the optimization of the network topologies. Three strategies were merged in the proposed model: (1) The number of input neurons was significantly reduced. Because, in chromatographic analysis, the chromatogram is defined by a large number of signal/ time response pairs and considering that they do not provide important information at any time, 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 chromatographic band. (2) We used neural networks based on product units (PUNNs) because sigmoidal or radial transfer functions do not guarantee the best generalization or fast learning. These families of parameterized transfer functions provide flexible decision borders with the advantages of increased information capacity and the ability to form higher-order combinations of inputs.[26] Consequently, network architecture can be reduced and the error in approximation decreased, without detriment to its approximation capability, because product units can approximate any function with a given accuracy as well as sigmoidal neural networks. (3) PUNN model selection was used to compare different neural network models with regard to some objectives (i.e., best fitting to the data, lower model complexity, best generalization capability, etc.). A population-based evolutionary algorithm was used for architecture design and estimation of weights,[27−29] which began the search with an initial population. At each iteration, the population was updated using a population update algorithm based on replication and two types of mutation (structural and parametric) operations. The structural mutation implies a modification of the structure of the function performed by the network, which allows the exploration of different regions of the search space, by changing the nodes in the hidden layer or the number of connections associated with the input or hidden nodes with a predetermined probability. The parametric mutation modifies the coefficients of the model by using a self-adaptive annealing algorithm. In this process, a random Gaussian value was used to change the weights of connections associated with nodes of input and hidden layers. It should be pointed out that the crossover operator was not used, bearing in mind its inability to handle permutation problems related to the fact that the same PUNN model can be shown by several representations. To test this approach, two *N*-methylcarbamate pesticides, carbofuran (CF) and propoxur (P), were quantified using a classic peroxyoxalate chemiluminescence (CL) reaction as a detection system for chromatographic analysis. Further details on the analytical

methodology are described elsewhere.[30] To our knowledge, no study on the use of PUNN models trained and designed by an evolutionary algorithm has, to date, been reported.

THEORY

The first step of the proposed approach for the quantification of overlapping chromatographic peaks consists of extracting useful information from them in order to select the inputs to the ANNs. When the chromatographic peaks are examined, it can be observed that the signals set (t_i, S_{t_i}) can be accurately fitted by least-squares regression to a four-parameter Weibull curve, defined by S_m (peak height), B (dispersion of the analytical signal values from S_m), C (related to the function shape, which is associated with the inflection points of the curve and defines the concavity and convexity regions), and t_m (residence time, associated with the time corresponding to the peak maximum), when the time domain ranges from $t_i > t_m - B[(C - 1)/C]^{1/C}$. If t_i is standardized by subtracting t_m, dividing by the dispersion parameter B, and displacing the standardized variable $[(C - 1)/C]^{1/C}$ units, a new temporal variable t'_i can be defined, that is, $t'_i = (t_i - t_m)/B + [(C - 1)/C]^{1/C}$, which is characterized by the location parameter $[(C - 1)/C]^{1/C}$ and a dispersion parameter equal to 1. In this case, S_{t_i} is the response variable, which is proportional to the contribution (concentration) of each component to the chromatographic peak, and t'_i is the independent variable. If it is assumed that the change of S_{t_i} with time is proportional to the inverse of the transformation time (t'_i) and the parameter (C) associated with the convexity of the function, the following differential equation can be obtained:

$$\frac{\partial S_{t_i}}{\partial (t'_i)} = \left(\frac{C - 1}{t'_i} - C\right) S_{t_i} \tag{1}$$

Considering that S_{t_i} at time $t'_i = 0$ is given by $S_m C \exp[-(C - 1)/C]$, where $S_m = S_{t_m}$, the integration of this equation provides

$$S_{t_i} = S_m C \exp[-(C-1)/C]\,(t'_i)^{C-1}\,e^{-t'_i C} \tag{2}$$

for $t'_i \geq 0$, that is, $t_i > t_m - B[(C - 1)/C]^{1/C}$, and $C > 1$, which corresponds to a four-parameter Weibull function.

As can be derived from eq 2, there is a strong interaction of parameter C with t_m and B through t'_i and with S_m through S_{t_i}. This interaction is maintained even following the appropriate transformations of the signals for the linearization of the model. In fact, if eq 2 is expressed in its logarithm form

$$\ln S_{t_i} = \ln S_m + \ln C - \frac{C - 1}{C} + (C - 1)\ln(t'_i) - t'_i C \tag{3}$$

and the t'_i value is substituted, the following equation can be obtained:

$$\ln S_{t_i} = \ln S_m + \ln C - \frac{C - 1}{C} + (C - 1)\ln\left[\frac{t_i - t_m}{B} + \left(\frac{C-1}{C}\right)^{1/C}\right] - \left[\frac{t_i - t_m}{B} + \left(\frac{C-1}{C}\right)^{1/C}\right]C \tag{4}$$

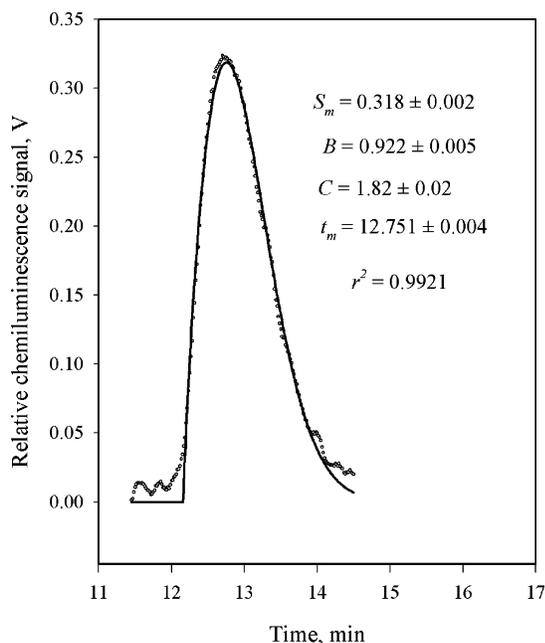in which the interactions among the parameters C, t_m, and B

Figure 1. Chromatographic peak fitted to a four-parameter Weibull function. (○) Experimental data and (−) Weibull curve. Parameter value ± standard deviation. [CF] and [P] are 30 and 150 ng/mL, respectively.

can be observed. Figure 1 shows the fit provided by the four-parameter Weibull function on a typical overlapping chromatographic peak obtained in the HPLC determination of mixtures of CF and P with CL detection. When both curves are examined and from the estimated statistical parameters, it can be concluded that the Weibull function is a fine tool for modeling this kind of chromatographic peak.

Multiplicative Neural Network Model. Let $\mathscr{R}^k$ be an n-dimensional Euclidean space and K a compact subset of it defined by $K = \{(x_1, x_2, ..., x_k) \in \mathscr{R}^k: x_i \in \mathscr{R}^+, i = 1, 2, ..., k\}$. We represent by $F(K)$ the family of functions f: $K \subset \mathscr{R}^k \rightarrow \mathscr{R}$ given by

$$f(x_1, x_2, ..., x_k) = \sum_{j=1}^{p} \beta_j (\prod_{i=1}^{k} x_i^{w_{ji}}) \qquad (5)$$

where β_j and $w_{ji} \in \mathscr{R}$, and p and $k \in N$. This typology of functions can be viewed as a polynomial with real exponents, and by appropriately choosing the exponents of the function f, it is easy to observe that the polynomial regression models are subsets of $F(K)$. For example, by appropriate selection of the exponents, $w_{ji} \in \{0, 1, 2\}$, a second-order polynomial regression model or quadratic response surface can be obtained:

$$f(x) = \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1,i<j}^{k} \beta_{ij} x_i x_j \qquad (6)$$

Let us now observe the data set $D_E \{x_i, y_i\}$ for $l = 1, 2, ..., n$, for which the regression model can be expressed by means of a lineal potential base function topology or PUNNs as eq 5. In these models, the product units can be defined as follows:
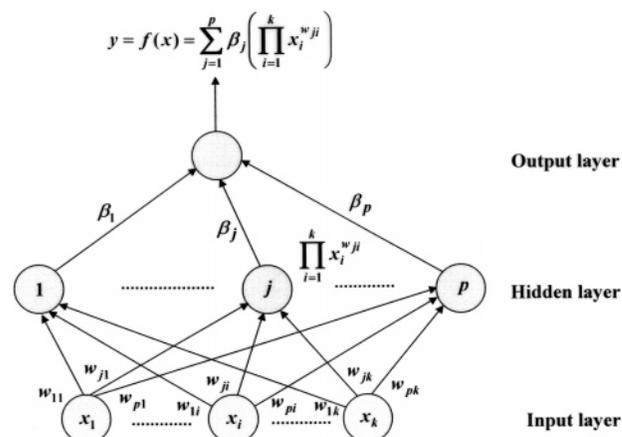
$$\prod_{i=1}^{k} x_i^{w_{ji}} \qquad (7)$$

Figure 2. Functional scheme of the multiplicative neural network based on product units.

where $w_{ji} = (w_{j1}, w_{j2}, ..., w_{jk})$ is a parameter set for the potential base functions.

This kind of function topology can be represented by a neural network architecture, as shown in Figure 2, with the following features: one input layer for the input variables, one hidden layer with a suitable number of nodes, and one output layer. Furthermore, the nodes of one layer cannot be connected themselves and there are no direct connections between the input and output layers. As stated above, in the chemical problem addressed in this study, the quantification of analytes that provide highly overlapping chromatographic peaks, the k network inputs, which represent the independent variables ($x_1, x_2, ..., x_k$), are represented by the four-parameter Weibull curve associated with the profile of the chromatographic peak and estimated by the Levenberg−Marquardt method; the p nodes in the hidden layer represent the term numbers of the model and, therefore, the number of product units considered, and the one node in the output layer corresponds to the concentration of the analyte to be analyzed in the sample, the N-methylcarbamate pesticide, CF or P in our case.

The transfer function of the jth node of the hidden layer is given by

$$\prod_{i=1}^{k} x_i^{w_{ji}} \qquad (8)$$

where $w_{ji} \in [0, L]$ is the weight for the connection between the ith node of the input layer and the jth nodes of the hidden layer. The linear transfer function of the node of the output layer is given by

$$\sum_{j=1}^{p} \beta_j (\prod_{i=1}^{k} x_i^{w_{ji}}) \qquad (9)$$

where $\beta_j \in [-M, M]$ is the weight for the connection between the jth node of the hidden layer and the node of the output layer. In brief, the topology for the functions defined in eq 5 can be readily represented by a PUNN.

Evolutionary Algorithm. This section deals with the framework of the evolutionary process to achieve the selection of the PUNN model. The population (PUNN models) is subjected to the operations of replication and
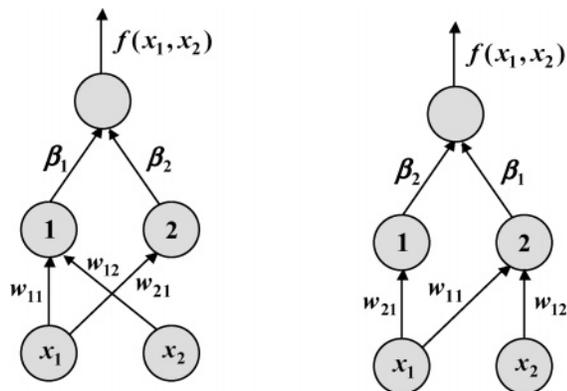
QUANTIFICATION OF OVERLAPPING PEAKS

J. Chem. Inf. Model., Vol. 45, No. 4, 2005 **897**

Figure 3. Effect of the permutation process on the network topology.

mutation (parametric and structural). Crossover is not used because of its disadvantages in deceptive environments;[31] in fact, many-to-one mapping should be considered for the representation (genotype) of the phenotype because two networks with their hidden nodes contained differently in their chromosomes will still be equivalent (see Figure 3). In general, any permutation of the hidden nodes will produce functionally equivalent networks with different chromosome representations. This problem makes the crossover operator very inefficient and ineffective in producing good offspring. With these features, the algorithm falls in the class of evolutionary programming.[32] Evolutionary programming is preferable to genetic algorithms when separating the search, and evaluation space does not afford an advantage. The general structure of the evolutionary algorithm, which is applied to an initial population of N_m individuals (PUNN models), can be verified in the following steps:

1. Evaluate the fitness score for each individual of the population on the basis of the objective function.

2. Copy the best individual to the next generation.

3. The best 10% of the population substitutes the worst 10% of individuals.

4. Apply parametric mutation operators to the best 10% of the population.

5. Apply structural mutation parametrics to the rest of the population.

These steps should be repeated until the population converges or the time is up.

The algorithm initializes the population with randomly generated networks. The number of hidden nodes for each network is chosen from a uniform distribution in the interval $[1, p]$, where p corresponds to the maximum number of hidden nodes considered. The number of initial connections from each hidden node to an input node is chosen similarly from a uniform distribution in the interval $[1, k]$, where k is the number of independent variables of the model (the four-parameter Weibull curve associated with the profile of the chromatographic peak in our case). Once a topology has been chosen, all connection weights are randomly assigned uniformly over the range $[0, L]$ for the connections from an input node to hidden nodes and the range $[-M, M]$ for the connections from hidden nodes to the output node.

In each generation of the search, the networks are first evaluated by the fitness function

$$A(f) = \frac{1}{1 + E(f)} \qquad 0 < A(f) \le 1 \qquad (10)$$

in which the relative error, $E(f)$, is given by

$$E(f) = \frac{1}{n_T} \sum_{l=1}^{n_T} (y_l - \hat{y}_l)^2 \qquad (11)$$

y_l and $\hat{y}_l$ being the experimental and expected values for the analyte concentration in the mixture, respectively, and n_T being the number of patterns used in the training set. The new population is generated replicating the best 10% of the former population, which substitutes the worst 10% of individuals; therefore, the number of individuals in the population is constant during the evolutionary process.

Mutations are separated into two classes: parametric mutation alters the values of exponents and coefficients of the functions of the population, and structural mutation alters the space of the parameters. Parametric mutation operators were applied to the best 10% of the population and structural mutation to the rest of the population. The severity of a mutation to an individual f is dictated by the function's temperature $T(f)$ given by

$$T(f) = 1 - A(f) \qquad 0 \le T(f) \le 1 \qquad (12)$$

where $A(f)$ is the fitness function. Thus, the temperature is determined by how close the function is to being a solution to the problem. Parametric mutations are accomplished each parameter w_{ji} and β_j of a function f with Gaussian noise, where the variance of normal distribution depends on the function's temperature. So, networks with a high temperature are mutated severely and those with a low temperature only slightly. This allows a coarse-grained search, initially, and progressively finer-grained searches as the network approaches the solution to the problem. To be exact, the exponents w_{ji} of the function, which represent the weights of the connections between an input and hidden nodes, are modified as follows:

$$w_{ji}(t + 1) = w_{ji}(t) + \xi_1(t) \qquad 1 \le i \le k; 1 \le j \le p \qquad (13)$$

where $\xi_1 \in N[0, \alpha_1(t) T(f)]$ represents a normally distributed one-dimensional random variable with mean 0 and variance $\alpha_1(t) T(f)$. The coefficient β_j of the function f, representing the weights of the connections between a hidden node and the output node, is modified as follows:

$$\beta_j(t + 1) = \beta_j(t) + \xi_2(t) \qquad 1 \le j \le p \qquad (14)$$

where $\xi_2 \in N[0, \alpha_2(t) T(f)]$ represents, in a similar way, a normally distributed one-dimensional random variable with mean 0 and variance $\alpha_2(t) T(f)$.

It should be pointed out that the modification of the exponents is different so that coefficients, that is $\alpha_1(t) \ll \alpha_2(t)$, are adaptively changed in every generation by some predefined rule. In essence, the functions $\alpha_1(t)$ and $\alpha_2(t)$

define the mutation strength in each case, and specifically for $i = 1, 2$, they are defined by

$$\alpha_i(t) = (1 + \beta) \, \alpha_i(t)$$
$$\text{if } A(s) > A(s - 1), \forall \, s \in \{t, t - 1, ..., t - r\}$$
$$= (1 - \beta) \, \alpha_i(t)$$
$$\text{if } A(s) = A(s - 1), \forall \, s \in \{t, t - 1, ..., t - r\}$$
$$= \alpha_i(t) \quad \text{in the other cases} \tag{15}$$

where $A(s)$ represents the fitness of the best individual in the sth generation and the parameters β and r are fixed, user-defined parameters. Taking into account that a generation is defined as successful if the best individual of the population is better than the best individual of the previous generation, if many successful generations are observed, this indicates that the best solutions lie in a better region in the search space. In this case, we increase the mutation strength in the hope of finding ever better solutions closer to the optimum solution. If the fitness of the best individual is constant in different generations, we decrease the mutation strength. In the other cases, the mutation strength is constant.

Once the mutations are made, the fitness of the individual is recalculated and the usual simulated annealing criterion is applied. Given that ΔA is the difference in the fitness function before and after the random step, if $\Delta A \geq 0$, the step is accepted, and if $\Delta A < 0$, then the step is accepted with a probability

$$P(\Delta A) = \exp\left(-\frac{\Delta A}{T}\right) \tag{16}$$

where T is the current temperature.

Structural mutation is more complex because it implies a modification of the structure of the network. There are five different structural mutations:

1. *Addition of a node.* The node is added with no connections to enforce the behavioral link with its parents.

2. *Deletion of a node.* A node is selected randomly and deleted together with its connections.

3. *Addition of a connection.* A connection is added, with weight 0, to a randomly selected node. There are two types of connection: from an input node to a hidden node and from a hidden node to the output node.

4. *Deletion of a connection.* A connection is selected and removed.

5. *Join nodes.* Two hidden nodes a and b, selected randomly, are replaced by another node c.

In this process, the common connections are preserved as are the noncommon connections, although only the former provide a probability higher than 0.5. The weights for the common connections are calculated as follows:

$$\beta_c = \beta_a + \beta_b \tag{17}$$

for connections between nodes of the hidden and output layers and

$$w_{ic} = \frac{w_{ia} + w_{ib}}{2} \tag{18}$$

for connections between input and hidden layers. An illustrative example of this process in shown in Figure 4.
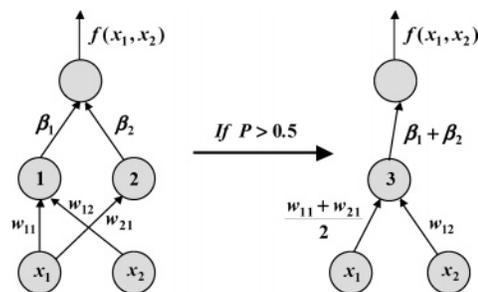

Figure 4. Estimation of the weights for the common connections by the evolutionary algorithm.

All of the above mutations are made sequentially in the same generation on the same network. For each mutation, there is a minimum value, Δ_m, and a maximum value, Δ_M, and the number of elements (nodes and connections) involved in the mutation is calculated as follows:

$$\Delta = \Delta_m + [U(0, 1) \, T(f) \, (\Delta_M - \Delta_m)] \tag{19}$$

Finally, the system is evolved until the average fitness of the network population stops growing, when functions $\alpha_1(t)$ and $\alpha_2(t)$ are near to zero. The values of the parameters for all experiments are shown in Table 1. It should be pointed out that the algorithm is quite robust to the modification of these parameters.

EXPERIMENTAL SECTION

Twenty five chromatograms provided by samples containing uniformly distributed concentrations of CF (30−150 ng/mL) and P (30−210 ng/mL) were prepared in triplicate as described elsewhere.[30] The experimental design consisted of an interpolation between different concentrations of each pesticide based on a grid, in which the validation samples lie nested within the training grid as follows: mixtures containing a pesticide at concentration levels of 30, 90, 150, or 210 ng/mL were selected for the training set, whereas those containing 20, 40, or 60 ng/mL were chosen for validation. In this way, 20 chromatograms for the training set and 5 for the validation set were used, both in triplicate. The Levenberg−Marquardt algorithm was used to estimate the four-parameter Weibull function associated with the profile of the overlapping chromatographic peaks. These parameters were defined as follows: $\hat{S}_m$ (peak height), $\hat{t}_m$ (residence time), $\hat{B}$ (dispersion of the analytical signal values from S_m), and $\hat{C}$ (related to the function profile, which is associated with the inflection points of the curve). The population-based evolutionary algorithm for ANN models, written in C language, was run on a Pentium IV compatible PC. To use the four parameters of the Weibull curve as inputs to the assayed network topologies, they were scaled over the range 0.1−0.9 for MLP models and from 0.2 to 0.8 for PUNN models. Thus, the new scaled variables were expressed as follows: $\hat{S}_m^*, \hat{t}_m^*, \hat{B}^*$, and $\hat{C}^*$. In a similar way, the concentrations of the analytes in the sample defined by [CF] and [P], which are used as outputs for the tested network models, were also scaled, in this case, over the range 0.1−0.9 for MLP models and between 1.0 and 2.0 for PUNN models. These new scaled independent variables were designed by [$\hat{CF}^*$] and [$\hat{P}^*$]. After the network models are optimized, estimations should be descaled according to the

QUANTIFICATION OF OVERLAPPING PEAKS

J. Chem. Inf. Model., Vol. 45, No. 4, 2005 **899**

Table 1. Parametric Values Used by the Evolutionary Algorithm

population parameters		structural mutation parameters: interval, $[\Delta_m, \Delta_M]$		parametric mutation parameters of eq 16	
parameter	value	parameter	value	parameter	value
size, N_m	1000	add nodes	[1, 2]	$\alpha_1(0)$	1
maximum number of hidden nodes, p	8	delete nodes	[1, 3]	$\alpha_2(0)$	5
number of independent variables, k	4	add connections	$[1, 2p]$	β	0.5
exponent interval, $[-M, M]$	$[-5, 5]$	delete connections	$[1, 2p]$	r	10
coefficient interval, $[0, L]$	$[0, 5]$				

Table 2. Accuracy and Statistical Results of the MLP and PUNN Models (over 30 Runs)

analyte	starting topology	connections		SEP_T				SEP_G			
		mean	SD	mean	SD	best	worst	mean	SD	best	worst
CF	4:3:1 MLP	15.2	1.79	4.52	0.71	3.47	5.88	3.13	1.17	1.60	5.67
	4:3:1 PUNN	12.6	1.10	3.70	0.34	3.01	4.53	3.12	0.36	2.38	4.07
P	4:2:1 MLP	9.97	1.38	4.20	0.86	2.98	6.34	3.72	1.16	2.28	6.42
	4:2:1 PUNN	8.10	1.24	3.92	1.05	2.68	6.54	3.82	1.09	2.24	5.81

same equation. The performance of the algorithms was tested using various network topologies that were run 30 times. The accuracy of each model was assessed in terms of the standard error of prediction (SEP) for the results obtained for both data sets, that is, SEP_T for the training set and SEP_G for the generalization set:

$$ SEP = \frac{100}{\bar{A}_i} \sqrt{\frac{\sum_{i=1}^{n}(A_i - \hat{A}_i)^2}{n}} \qquad (20) $$

where A_i and $\hat{A}_i$ are the experimental and expected values for the analyte concentration in the sample, $\bar{A}_i$ 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). The nonparametric Kolmogorov−Smirnov (K−S) and Levene tests were performed using SPSS 12.0 statistical software[33] and used to evaluate the performance of the different models in selecting the most suitable network topology.

RESULTS AND DISCUSSION

The quantitative analysis in HPLC requires the resolution of the analytes being studied, but sometimes chromatographic separation is not accomplished. In these cases, chemometric resolution of these compounds is possible. The mathematical approach may be successful when either the spectra or the concentration profiles of the compounds are distinguishable from each other. Thus, a variety of chemometric methods have been developed over the past few years for quantitative analysis of unresolved peaks.[34−38] Most of these are based on the data matrix obtained from HPLC with the diode array detector (DAD). Although based on the information provided by both spectra and chromatographic data, these approaches do not work well for highly overlapping peaks, and therefore, ANN calibration models have been recently reported as an alternative on account of the lower standard deviation (SD) and better results achieved in the prediction.[22−24]

Recently, we have published a powerful chemometric tool based on the use of pruned standard MLPs for quantifying highly overlapping chromatographic peaks, such as those involved in the determination of two *N*-methylcarbamate pesticides (CF and P) by HPLC with CL detection.[25] This kind of detection system provides an additional degree of complexity to the analytical problem addressed because, unlike the methods based on the bilinear data matrix from HPLC/DAD, the analytical data are supplied by a single detector, and therefore, the chemometric approach should work well without the aid of spectral discrimination. Although this methodology does produce good results, they could be further improved with the use of PUNN models, on account of their greater potential, particularly in order to obtain simpler models with more chemical interpretation. In fact, PUNN models are more robust tools for the inference of possible interactions between input and output variables, chromatographic profile and the analyte (CF and P) concentrations in the sample, in our case.

Prediction Ability Features of PUNN versus MLP Models. To further compare the predictive ability of both models in terms of SEP_G, topology, the number of connections, robustness, and chemical interpretability, one output neural network model was made, considering the output layer as a single node corresponding to the concentration of the analyte in the sample to be analyzed.

Table 2 shows the statistical results obtained over 30 runs using both neural network models. The same 4:3:1 and 4:2:1 architectures were chosen to start both model selection processes for CF and P, respectively, the aim being to compare the different ANN models using the same initial architecture. As can be seen, both models provided quite good results (in terms of accuracy and precision) for determining the concentration of each pesticide from strongly overlapping chromatographic peaks. The PUNN model, however, yielded better results for all tested features: smaller SEP mean values, SD, and the number of connections, especially for the determination of CF. For this pesticide, the proposed PUNN method is more robust, taking into account the lower SD values provided for both SPE_T and SPE_G.

To verify the hypothesis of significant differences in the mean SEP_G value and the number of connections, three statistical tests were performed using the SPSS statistical

Table 3. Statistical Comparison (p values of the Levene and Student's *t* Tests) of the Generalization Ability (SEP$_G$) and Number of Connections (*n*) for ANN Models

	SEP$_G$		number of connections	
MLP vs PUNN	Levene test	*t* test	Levene test	*t* test
carbofuran	0.000	0.991	0.018	0.000
propoxur	0.868	0.743	0.299	0.000

package.[33] A normal distribution can be assumed for all of the variables to contrast because the *p* values of the K−S test are over 0.01. A Student's *t* test was performed in order to ascertain whether the differences, in the mean values, among the SEP$_G$ values and the number of connections obtained with each model were significant using, previously, a variance Levene test. The results provided by the Levene and Student's *t* tests are shown in Table 3, from which the following conclusions can be drawn: (1) The variance of the error is smaller by using PUNN models in the prediction of CF but not for the prediction of P. (2) The performances in means of the two models are not statistically significant, with a confidence level of 0.01, for the predictions of CF and P. (3) The complexity of the model, based in terms of the number of connections of the network architecture, is significantly lower for the PUNN models, both in variance and in mean.

According to this study, the following final optimal network models were those reported in Table 4, in which quantitative equation systems for the direct determination of the contribution of each pesticide to the overlapping chromatographic bands used: (a) the parameters estimated by the Weibull regression of the peak, (b) the optimized network weights, and (c) the sigmoidal and the product unit transfer functions for MLP and PUNN models, respectively, which also were included on account of the simplicity of the proposed network architectures. In summary, both topologies can be readily used for the quantification of the analytes in the sample, although the PUNN topology provides more simple and robust models (see Table 2). On the other hand, it is noteworthy that PUNN models offer better results in mean for the training set, which contains mixtures with extreme concentrations of the analytes, because in that set a stronger interaction between the input variables is observed.

Chemical Interpretation Provided by MLP and PUNN Models. Neural network modeling has been used extensively over the past decade for solving a great variety of nonlinear chemical problems, but it suffers from the perception of being a "black box" heuristic tool. The current introduction evolutionary algorithms for designing network models enable architectures with reduced dimensions to be obtained in such a way that the mathematical transformation between the input and output can be easily implemented (see equations shown in Table 4). From these simpler models, quality chemical information could be derived to explain the analytical problem at hand. Thus, the value of each transfer function involved in the model is calculated over the range studied for the input variables, $\hat{S}_m^*$, $\hat{B}^*$, $\hat{C}^*$, and $\hat{t}_m^*$, to establish their relative contribution for the determination of CF and P, which is essentially derived through the so-called base transfer function. To evaluate the chemical interpretation provided by both models, it is necessary to observe, at least from a qualitative point of view, how the variation of the concentration of both pesticides affects the shape of the chromatograms achieved using different mixtures of pesticides (see Figure 5). As can be seen, the two pesticides show quite different chromatographic behaviors; in fact, chromatograms with the same concentration of CF and a variable concentration of P exhibit a gradual increase in the peak height, S_m, and a slight decrease in the residence time, t_m (Figure 5A). Meanwhile, t_m remains virtually constant and S_m increases with an increase in the concentration of CF for chromatograms of mixtures with a constant concentration of P and a variable concentration of CF (Figure 5B). Further conclusions can be drawn regarding the variation of the profile and the dispersion of chromatographic peaks (*C* and *B* parameters, respectively) as a function of the pesticide concentration in the mixture. By comparing chromatograms in parts A and B of Figure 5, it can be inferred that parameter *B* is more related to the variation of CF in the mixture, whereas parameter *C* is more linked to the change of P.

In light of these qualitative considerations and using the quantification equations shown in Table 4, the most notable features of the proposed models are discussed below from both computational and chemical point of views.
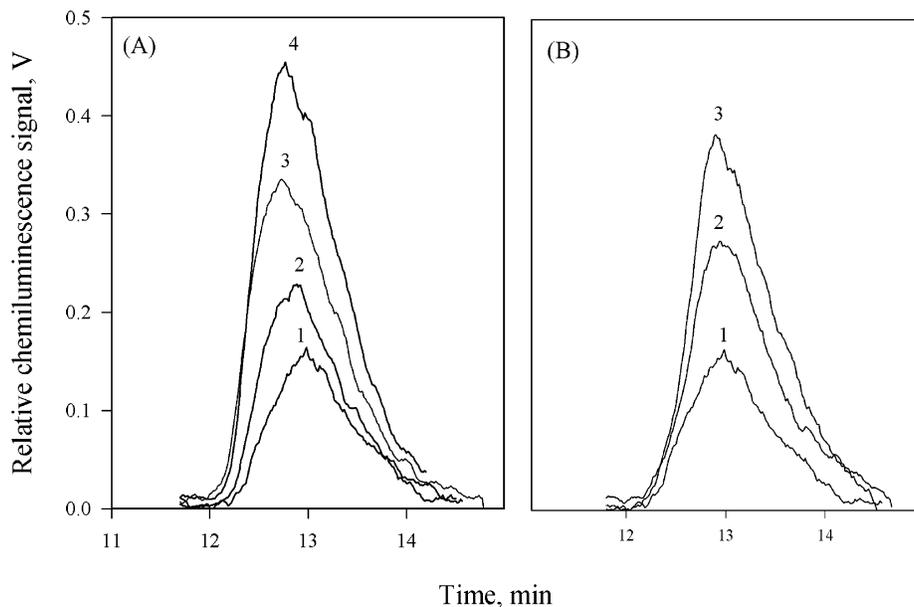
Table 4. Quantification Equations and Accuracy Provided by the Optimized MLP and PUNN Topologies as Applied to the Determination of Carbofuran Propoxur from Overlapping Chromatographic Peaks[a]

	carbofuran network models	
	4:3:1 MLP topology	4:3:1 PUNN topology
quantification equations	$[\hat{CF}]^* = -2.88 + 4.02\hat{h}_1 + 4.51\hat{h}_2 - 0.61\hat{h}_3$	$[\hat{CF}]^* = -0.30 + 0.66\hat{h}_1 + 3.33\hat{h}_2 - 0.80\hat{h}_3$
transfer functions	$\hat{h}_1 = 1/[1 + \exp(-4.91 - 0.66\hat{S}_m^* - 0.64\hat{B}^* - 0.62\hat{C}^* - 0.27\hat{t}_m^*)]$	$\hat{h}_1 = (\hat{C}^*)^{0.27}(\hat{t}_m^*)^{0.84}$
	$\hat{h}_2 = 1/[1 + \exp(-0.98\hat{S}_m^* - 0.88\hat{B}^*)]$	$\hat{h}_2 = (\hat{S}_m^*)^{0.55}(\hat{B}^*)^{0.43}(\hat{t}_m^*)^{0.07}$
	$\hat{h}_3 = 1/[1 + \exp(-0.92\hat{S}_m^* + 4.40\hat{B}^* + 1.95\hat{t}_m^*)]$	$\hat{h}_3 = (\hat{S}_m^*)^{0.67}(\hat{B}^*)^{1.03}$
	propoxur network models	
	4:2:1 MLP topology	4:2:1 PUNN topology
quantification equations	$[\hat{P}]^* = 1.29 - 4.80\hat{h}_1 - 1.20\hat{h}_2$	$[\hat{P}]^* = 0.95 + 0.39\hat{h}_1 - 0.74\hat{h}_2$
transfer functions	$\hat{h}_1 = 1/[1 + \exp(3.82 + 2.16\hat{S}_m^* - 0.36\hat{B}^* - 1.01\hat{C}^* - 0.01\hat{t}_m^*)]$	$\hat{h}_1 = (\hat{S}_m^*)^{0.99}(\hat{C}^*)^{-0.84}(\hat{t}_m^*)^{0.08}$
	$\hat{h}_2 = 1/[1 + \exp(-0.02 + 1.20\hat{S}_m^* - 3.30\hat{B}^*)]$	$\hat{h}_2 = (\hat{S}_m^*)^{0.13}(\hat{B}^*)^{0.03}(\hat{t}_m^*)^{4.66}$

[a] $[\hat{CF}]^*$ and $[\hat{P}]^* \in [0.1, 0.9]$ and $[1.0, 2.0]$ for MLP and PUNN, respectively. $\hat{S}_m^*$, $\hat{t}_m^*$, $\hat{B}^*$, and $\hat{C}^* \in [0.1, 0.9]$ and $[0.2, 0.8]$ for MLP and PUNN, respectively.

QUANTIFICATION OF OVERLAPPING PEAKS

J. Chem. Inf. Model., Vol. 45, No. 4, 2005 **901**

Figure 5. Overlapping chromatograms of mixtures containing carbofuran and propoxur to different extents. (A) Curves 1−4 correspond to samples with a fixed amount of carbofuran (30 ng/mL) and variable amounts of propoxur (30, 90, 150, and 210 ng/mL, respectively). (B) Curves 1−3 correspond to samples in which the amount of propoxur is fixed at 30 ng/mL and carbofuran varies at 30, 90, and 150 ng/mL, respectively.
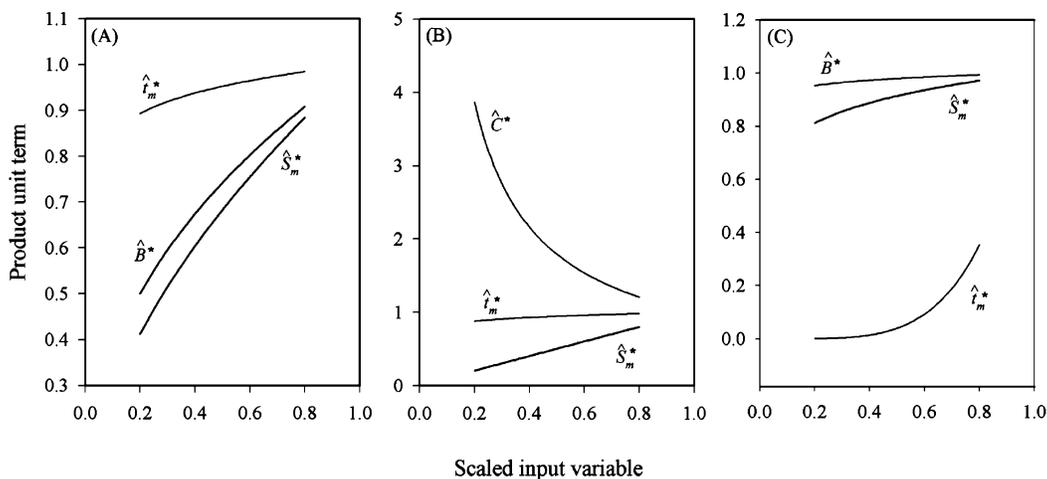

Figure 6. Relative contribution of the product unit terms used for the quantitative determination of the pesticides provided by PUNN models. (A) $\hat{h}_2$ function for CF, (B) $\hat{h}_1$ function for P, and (C) $\hat{h}_2$ function for P.

Models for Carbofuran. For the MLP model, from the equations shown in Table 4 and after evaluating the relative contribution of each sigmoidal function for the MLP model, it follows that the $\hat{h}_1$ and $\hat{h}_2$ sigmoidal functions exhibit an appreciable influence on the determination of $[\hat{CF}]^*$. The respective dependencies are as follows: $\hat{h}_1$ and $\hat{h}_2$ functions essentially depend on $\hat{S}_m^*$ and $\hat{B}^*$ parameters, in that order, and also on $\hat{C}^*$ and $\hat{t}_m^*$ (through $\hat{h}_1$), albeit to a lesser extent. In all cases, these dependencies are direct. Regarding the other sigmoidal function, $\hat{h}_3$ (related in a negative manner to $[\hat{CF}]^*$) depends on the input variables in the following order of importance: $\hat{B}^*$, $\hat{t}_m^*$, and $\hat{S}_m^*$. From the foregoing, it follows that $\hat{S}_m^*$ and $\hat{B}^*$ are the two parameters of the Weibull curve that most strongly influence the quantification of CF in the sample; the $\hat{C}^*$ and $\hat{t}_m^*$ parameters also contribute to this, although to a lesser extent, especially $\hat{t}_m^*$.

From the quantification equations shown in Table 4 for the PUNN model, it follows that (1) the value of $[\hat{CF}]^*$ depends mainly on the interaction between the parameters $\hat{S}_m^*$ and $\hat{B}^*$, the contribution of $\hat{t}_m^*$ being practically negligible, as inferred from the exponential values of the base term $\hat{h}_2$ and from Figure 6A, in which the values of each product unit term of $\hat{h}_2$ are plotted versus the scaled input variable over the range 0.2−0.8. The direct dependencies of the $\hat{S}_m^*$ and $\hat{B}^*$ parameters on $[\hat{CF}]^*$ are clear. (2) The other product unit functions, which bear a direct ($\hat{h}_1$) and inverse ($\hat{h}_3$) linear relationship with $[\hat{CF}]^*$, depend to a similar extent on the four parameters of the Weibull curve (see exponent of these parameters in $\hat{h}_1$ and $\hat{h}_3$). In short, the parameters $\hat{S}_m^*$ and $\hat{B}^*$ are the key for the determination of CF in the sample. From the relative value of the exponents for them in the transfer function $\hat{h}_2$, the following relative order of importance can be inferred: $\hat{S}_m^* > \hat{B}^* \gg \hat{t}_m^*$. Thus, from a chemical point of view, the peak height and the dispersion of the chromatographic peaks ($\hat{S}_m^*$ and $\hat{B}^*$ parameters, respectively) are closely related to the concentration of CF in the analyzed sample; the residence time, $\hat{t}_m^*$, also contributes, although to a lesser extent.

Finally, by comparing the ability of both network models to provide quality chemical information, a greater degree of interpretability of the addressed chemical problem was achieved using the PUNN model. In fact, the influence of the input parameters on the output can be more easily inferred from the relative values of their exponents in the product units. Specifically, and regarding the interpretability provided by the MLP model, the PUNN model enables us to establish the relative influence between the $\hat{S}_m^*$ and $\hat{B}^*$ parameters, and clearly, residence time, $\hat{t}_m^*$, practically does not contribute to the determination of CF in the sample (compare transfer functions for both network models proposed for CF in Table 4). These conclusions are in line with the qualitative conclusions drawn from the overlapping chromatographic peaks shown in Figure 5.

Models for Propoxur. In the case of the MLP model, both sigmoidal functions exhibited an appreciable influence on the determination of [P]*, although $\hat{h}_1$ did so to a greater extent. The respective dependencies are as follows: $\hat{h}_1$ is closely related to $\hat{S}_m^*$ and $\hat{C}^*$ and to a lesser extent to $\hat{B}^*$ and practically does not depend on $\hat{t}_m^*$; $\hat{h}_2$ depends on $\hat{B}^*$ and $\hat{S}_m^*$ in that order of importance. According to these results, we can infer that the $\hat{S}_m^*$ variable is essential for determining [P]* in the sample, whereas the contribution of the other variables, especially $\hat{C}^*$, provide a better fitting in its quantification. By using the PUNN model for the determination of P in the sample, and from the quantification equations shown in Table 4, it follows that the value of [P]* basically depends on the $\hat{S}_m^*$, $\hat{C}^*$, and $\hat{t}_m^*$ parameters according to the relative values of the exponents in the product unit functions and from the plots shown in parts B and C of Figure 6. Contrary to the MLP model, the PUNN one clearly establishes the influence of the $\hat{t}_m^*$ parameter on the determination of P, as can easily be inferred from the plots shown in Figure 6C.

CONCLUSIONS

As shown in this study, multiplicative neural networks based on product units, designed using an evolutionary algorithm, have proved to be a more computationally powerful tool than sigmoidal networks for quantifying highly overlapping chromatographic peaks provided by a single detector. So, PUNN models provide a better information ability, smaller network architectures, and more robust models (smaller standard deviation) and are quite simple and easier to interpret from both a computational and a chemical point of view, in contrast to the classical MLP models. Simple and clear relationships can be established between the input variables (the four parameters, estimated by NLR, of the Weibull curve fitted to the chromatographic peaks) and the output variable (the concentration of the analyte in the sample).

ACKNOWLEDGMENT

REFERENCES AND NOTES

(1) Zupan J.; Gasteiger, J. Neural networks: a new method for solving chemical problems or just a passing phase? *Anal. Chim. Acta* **1991**, *248*, 1−30.

(2) Smits, J. R. M.; Melssen, W. J.; Buydens, L. M. C.; Kateman, G. Using artificial neural networks for solving chemical problems. I. Multilayer feed-forward networks. *Chemom. Intell. Lab. Syst.* **1994**, *22*, 165−189.

(3) Grunenberg, J.; Herges, R. Prediction of chromatographic retention values (RM) and partition coefficients (log poct) using a combination of semiempirical self-consistent reaction field calculations and neural networks. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 905−911.

(4) Mitchell, B. E.; Jurs, P. C. Prediction of aqueous solubility of organic compounds from molecular structure. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 489−496.

(5) Lavine, B. K.; Workman, J. Chemometrics. *Anal. Chem.* **2002**, *74*, 2763−2770.

(6) Despagne, F.; Massart, D. L. Neural networks in multivariate calibration. *Analyst* **1998**, *123*, 157R−178R.

(7) Dieterle, F.; Nopper, D.; Gauglitz, G. Quantification of butanol and ethanol in aqueous phases by reflectometric interference spectroscopy—different approaches to multivariate calibration. *Fresenius' J. Anal. Chem.* **2001**, *370*, 723−730.

(8) Dieterle, F.; Busche, S.; Gauglitz G. Growing neural networks for a multivariate calibration and variable selection of time-resolved measurements. *Anal. Chim. Acta* **2003**, *490*, 71−83.

(9) Hornik, K. Multilayer feedforward neural networks are universal approximators. *Neural Networks* **1989**, *2*, 359−366.

(10) Hornik, K. Approximation capabilities of multilayer feedforward networks. *Neural Networks* **1991**, *4*, 251−257.

(11) Tchistiakov, V.; Ruckebusch, C.; Duponchel, L.; Huvenne, J. P.; Legrand, P. Neural network modelling for very small spectral data sets: reduction of the spectra and hierarchical approach. *Chemom. Intell. Lab. Syst.* **2001**, *54*, 93−106.

(12) Schmith, M., On the complexity of computing and learning with multiplicative neural networks. *Neural Comput.* **2001**, *14*, 241−301.

(13) Durbin, R.; Rumelhart, D. Product units: A computationally powerful and biologically plausible extension to back-propagation networks. *Neural Comput.* **1989**, *1*, 133−142.

(14) Salinas E.; Abbott L. F. A model of multiplicative neural responses in parietal cortex. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 11956−11961.

(15) Stein, B. E.; Meredith, M. A. *The Merging of the Senses*; MIT Press: Cambridge, MA, 1993.

(16) Saito, K.; Nakano, R. Extracting regression rules from neural networks. *Neural Networks* **2002**, *15*, 1279−1288.

(17) Siouffi, M. M.; Phan-Tan-Luu, R. Optimization methods in chromatography and capillary electrophoresis. *J. Chromatogr., A* **2000**, *892*, 75−106.

(18) Coenegracht, P. M. J.; Metting, H. J.; Van-Loo, E. M.; Snoeijer, G. J.; Doornbos, D. A. Peak tracking with a neural network for spectral recognition. *J. Chromatogr.* **1993**, *631*, 145−160.

(19) Metting, H. J.; Coenegarcht, P. M. J. Neural networks in high-performance liquid chromatography optimization: response surface modelling. *J. Chromatogr., A* **1996**, *728*, 47−53.

(20) Hu, Y. H.; Zhou, G. W.; Kang, J. H.; Du, Y. X.; Huang, F.; Ge, J. H. Assessment of chromatographic peak purity by means of artificial neural networks. *J. Chromatogr., A* **1996**, *734*, 259−270.

(21) Miao, H. J.; Yu, M. H.; Hu, S. X. Artificial neural networks aided deconvolving overlapped peaks in chromatograms. *J. Chromatogr., A* **1996**, *749*, 5−11.

(22) Garrido-Frenich, A.; Martinez-Galera, M.; Gil-Garcia, M. D.; Martinez-Vidal, J. L.; Catasus, M.; Marti, L.; Mederos, M. V. 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.

(23) Li, Y. B.; Huang, X. Y.; Sha, M.; Meng, X. S. Resolution of overlapping chromatographic peaks by radial basis function neural network. *Sepu* **2001**, *19*, 112−115.

(24) Galeano-Diaz, T.; Guiberteau, A.; Ortiz, J. M.; Lopez, M. D.; 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.

(25) Hervás, C.; Silva, M.; Serrano, J. M.; 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.

(26) Wlodzislaw, D.; Norbert. J. Survey of neural transfer functions. *Neural Comput. Surveys* **1999**, *2*, 163−212.

(27) García, N.; Hervás, C.; Muñoz, J. Multi-objective cooperative coevolution of artificial neural networks. *Neural Networks* **2002**, *15*, 1259−1278.

(28) García, N.; Hervás, C.; Muñoz, J. Covnet: A cooperative coevolutionary model for evolving artificial neural networks. *IEEE Trans. Neural Networks* **2003**, *14*, 575−596.

(29) García, N.; Ortíz, D.; Hervás, C. Cooperative coevolution of generalized multilayer perceptron. *Neurocomputing* **2004**, *56*, 257−283.

(30) Orejuela, E.; Silva, M. Monitoring some phenoxyl-type *N*-methyl-carbamate pesticide residues in fruit juices using high-performance liquid chromatography with peroxyoxalate-chemiluminescence detection. *J. Chromatogr., A* **2003**, *1007*, 197−201.

(31) Angeline, P. J.; Saunders, G. M.; Pollack, J. B. An evolutionary algorithm that construct recurrent neural networks. *IEEE Trans. Neural Networks* **1994**, *5*, 54−65.

(32) Fogel, D. B. Using evolutionary programming to greater neural networks that are capable of playing Tic-Tac-Toe. In *International Conference on Neural Networks*, San Francisco, CA; IEEE Press: Piscataway, NJ, 1993; pp 875−880.

(33) *SPSS, Advanced Models*, version 12.0; SPSS Inc.: Chicago, IL, 2003.

(34) van-Zomeren, P. V.; Darwinkel, H.; Coenegracht, P. M. J.; de Jong, G. J. 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.

(35) Sanchez, F. C.; Rutan, S. C.; Garcia, M. D. G.; Massart, D. L. Resolution of multicomponent overlapped peaks by the orthogonal projection approach, evolving factor analysis and window factor analysis. *Chemom. Intell. Lab. Syst*. **1997**, *36*, 153−164.

(36) Gross, G. M.; Prazen, B. J.; Synovec, R. E. Parallel column liquid chromatography with a single multiwavelength absorbance detector for enhanced selectivity using chemometric analysis. *Anal. Chim. Acta* **2003**, *490*, 197−210.

(37) Fraga, C. G.; Bruckner, C. A.; Synovec, R. E. Increasing the number of analyzable peaks in comprehensive two-dimensional separations through chemometrics. *Anal. Chem.* **2001**, *73*, 675−683.

(38) Liang, Y. Z.; Hamalainen, M. D.; Kvalheim, O. M.; Andersson, R. Assessment of peak origin and purity in one-dimensional chromatography by experimental design and heuristic evolving latent projections. *J. Chromatogr., A* **1994**, *662*, 113−122.

CI049697O