# Computational Neural Networks for Resolving Nonlinear Multicomponent Systems Based on Chemiluminescence Methods

César Hervás[†] and Sebastián Ventura[†]

Department of Computer Science, University of Córdoba, E-14004 Córdoba, Spain

Manuel Silva[‡] and Dolores Pérez-Bendito*[,‡]

Department of Analytical Chemistry, Faculty of Sciences, University of Córdoba, E-14004 Córdoba, Spain

Received March 20, 1998

This paper proves that computational neural networks are reliable, effective tools for resolving nonlinear multicomponent systems involving synergistic effects by using chemiluminescence-based methods developed by continuous addition of reagent technique. Computational neural networks (CNNs) were implemented using a preprocessing of data by principal component analysis; the principal components to be used as input to the CNN were selected on the basis of a heuristic method. The leave-one-out method was applied on the basis of theoretical considerations in order to reduce sample size with no detriment to the prediction capacity of the network. The proposed approach was used to resolve trimeprazine/methotrimeprazine mixtures with a classical peroxyoxalate chemiluminescent system, such as the reaction between bis(2,4,6-trichlorophenyl)oxalate and hydrogen peroxide. The optimum network design, 9:5*s*:2*l*, allowed the resolution of mixtures of the two analytes in concentration ratios from 1:10 to 10:1 with very small (less than 5%) relative errors.

## INTRODUCTION

Chemiluminescence (CL), i.e., the production of electromagnetic radiation by chemical reaction, is an interesting aspect of transient signals. The use of CL-based determinations has grown considerably in recent years as a result of their high sensitivity and expeditiousness and of its instrumental simplicity. However, the low selectivity achieved in some CL-based determinations restricts their application scope.[1−3] Although this problem can be solved by using instrumental approaches (liquid chromatographic detection systems based on CL reactions[4]), the use of more inexpensive choices such as chemometric techniques, that affording multicomponent CL-based determinations, allow in some cases to overcome this difficulty.[5,6] In general, chemical systems involved in multicomponent CL-based determinations pose several drawbacks which does not recommend the use of statistical methods: first, the corresponding differential equations describing these chemical systems are unknown; second, the presence of synergistic effects, which quite often occur in these reactions. On account of this great degree of nonlinearity, only powerful chemometric tools can offer the suitable accuracy for the resolution of these mixtures. Among them, computational neural networks were chosen in this work considering the suitability of their features to the proposed chemical problem.

CNNs are among the most exciting recent developments in computational science[7] and have grown enormously in popularity in different scientific fields (analytical chemistry included).[8−13] Multilayer feed-forward neural networks based on different versions of standard back-propagation (BP) learning algorithm have been treated by several authors[14−19] as highly powerful tools to study uniform approximation of an unknown $C^P$-function. These results encouraged us to develop a new methodology for multicomponent CL-based determinations, because it is straightforward to prove that CL signals from the mixture fulfilled the above-mentioned conditions. In addition, and taking into account that the magnitude of the error over the training set is associated with the dimension of the input space,[19] we preprocessed CL signals subject to synergistic effects by principal component analysis (PCA). A heuristic method is used to select the optimum number of PCs (CNN inputs) in order to minimize the error over the test set. Although the leave-one-out method produces higher error over the test set, it was used in this work on account that it provides the most reliable results for the fairly small data set used. For practical reasons, the use of a small data set is an important condition to develop methods to be applied. Leave-one-out is a sample reuse method for estimating the average error over the test set; it makes the most efficient use of the available data and makes the least assumptions on the statistics of the data.[20−22]

The proposed methodology was validated by the simultaneous determination of trimeprazine and methotrimeprazine (two phenothiazine derivatives with very similar kinetic behavior) in mixtures using the classical peroxyoxalate system based on the reaction between bis(2,4,6-trichlorophenyl)oxalate (TCPO) and hydrogen peroxide.[23] The reaction was implemented by using the continuous-addition-of-reagent technique, which increased the nonlinearity of the chemical system studied on account of its second-order kinetic nature.[24]

[†] E-mail: ma1hemac@uco.es.
[‡] E-mail: qa1sirom@uco.es.
* Corresponding author.

10.1021/ci980030+ CCC: $15.00 
Published on Web 09/19/1998

## THEORY

Many data modeling problems pose two difficulties, namely (a) the absence of a complete a priori model for the data generation process and (b) a limited amount of available data. This was the case with the problem addressed in this work, where the underlying distribution of the signal was unknown. In this work only 50 patterns, which conform the whole data set, were considered for practical and scientific reasons (viz., to maintain cost and run time within acceptable limits and to evaluate the influence of a fairly small data set on the network design). Restricting the amount of data makes the selection of the appropriate model complicated and therefore more challenging and crucial. In particular, it is easier to overfit a small training set and makes it very difficult to hold out a sufficiently large independent test sample.

The leave-one-out method, which has been used in this work, is a variant of the statistical cross-validation (CV) method.[25−27] In each experiment, the method trains the network with $n - 1$ patterns—where $n$ is the number of available patterns—and leaves 1 to test the error over the test set. The experiment is repeated for the $n$ patterns, and the average error over the test set is the average of the errors obtained for each pattern.

Let us to consider a set of observations $O = \{\vec{x}_i, y_i\}$ (with $i = 1, ..., n$) that are assumed to have been generated as

$$y_i = f(\vec{x}_i) + e_i$$

where $f(\vec{x})$ is an unknown function, the inputs are drawn independently with an unknown stationary probability density function $p(\vec{x})$, $e_i$ are independent random variables with zero mean and variance ($\sigma_e^2$), and $y_i$ are the observed target values.

Our task is to find an estimate, $\hat{f}_\theta(\vec{x})$, for $f(\vec{x})$ by using neural network learning, from a class of predictors or models, $\hat{f}_\theta(\vec{x})$, where, in general, $\theta \in \Theta.\Theta$ depends on a subset of available input variables $X$ used for training, on the CNN architecture, and on the set of adjustable parameters (network weights, $W$) of the selected architecture.

We used %SEP (relative standard error of prediction), the ratio between the squared root of the cross-validation average squared error, $CV^{1/2}(\theta)$, and the average of the prediction error, $\hat{e}_{\theta(j)}$ for all patterns in the prediction set. %SET (relative standard error of training) has an equivalent definition, but it is associated with the training error rather than with the prediction.

It is worth noting that, the leave-one-out method is expensive to compute for neural network models, in fact, it involves constructing $n$ different networks and training each with sets of $n - 1$ patterns, (i.e., each network is trained with all the patterns of the training set except the $j$th pattern to obtain a model $\hat{f}_{\theta(j)}(\vec{x})$ with its associated estimation weight matrix, $\hat{W}_j$). In this way, $n$ estimators of the function $\hat{f}_\theta(\vec{x})$ are obtained, from which a sufficiently acceptable global estimator can be derived, because the combination of $n$ different estimators reduces the effect of spurious patterns on the model.

In our case, we use as much data as possible for training since the higher the training set the lower the error variance over the test set. If the $n$ weight matrixes obtained in every training process ($\hat{W}_j$, $j = 1, ..., n$) are used for obtaining the error over a new test set, we should take average squared errors over the new test set ($MSE_{G(j)}$):

$$MSE_{G(j)} = \frac{1}{n}\sum_{i=1}^{n} [y_i^* - \hat{f}_{\theta(j)}(\vec{x}_i^*)]^2 \quad \forall_j = 1, ..., n$$

where $(\vec{x}_i^*, y_i^*)$ are new observations that were not used in constructing $\hat{f}_{\theta(j)}(\vec{x})$.

Although the computational cost may seen high, our experience shows that such cost is closed to the time necessary to prepare and obtain a few CL signal vs time curves in the laboratory.

## EXPERIMENTAL SECTION

Computations were performed by using the extended "delta-bar-delta" rule included in the NeuralWorks Professional II software package, which was run on a Sun workstation. Sigmoidal, *s*, and linear, *l*, functions were used for hidden and output layers, respectively. Before training was started, the values applied to the input and output nodes were normalized over the ranges from −1.0 to 1.0 and from 0.2 to 0.8, respectively; also, the connection weights were uniformly initialized between −0.1 and 0.1. The structures of the CNNs used consisted of an input layer with $x$ nodes (selected scores of PCA-preprocessed data), a single hidden layer, and an output layer containing two nodes (the concentrations of both components in the mixture). In addition, the hidden and output layers were connected to the bias (whose activation was always +1), and whose weights were also altered during the training.

An overall 50 synthetic samples containing uniformly distribution concentrations of the analytes (trimeprazine and methotrimeprazine) over the range 0.5−5.0 $\mu$g/mL were prepared as described in a previous paper.[23] The concentration of both phenothiazines in the 1:1 mixture was 3.0 $\mu$g/mL, and the other mixtures were prepared accordingly. CL data were recorded at a frequency of 100 points/s over an interval of 3 s (300 points/curve) using an instrumental setup consisting of (1) a Perkin-Elmer 650−10S spectrofluorimeter with the light source off; whose sample holder was replaced with a small magnetic stirrer holding a cylindrical glass reaction vessel; (2) a Metrohm Dossimat 665 autoburet for adding the reagent (TCPO); and (3) a NEC/Multisync 2A 33 MHz compatible computer equipped with a PC-Multilab 812 PG analog-to-digital converter.

## RESULTS AND DISCUSSION

Let $A + R \rightarrow P$ be an irreversible reaction where $A$ is the analyte, $R$ the reagent, and $P$ the formed product. If this reaction was developed using the continuous addition of reagent (CAR) technique, that is, if a solution of the reagent at a concentration $[R]_0$ is added at a constant rate $u$ to a volume $V_0$ of a solution containing the analyte $[A]_0$, the integrated rate equation for this process is given by[24]

$$\ln\frac{S_\infty - S_t}{S_\infty} = -k[R]_0 t + \frac{k[R]_0 V_0}{u}\ln\frac{V_0 + ut}{V_0} - \ln\frac{V_0 + ut}{V_0}$$

where $k$ is the second-order rate constant and $S_t$ and $S_\infty$ are the signals at time $t$ and $\infty$ (total reaction development)

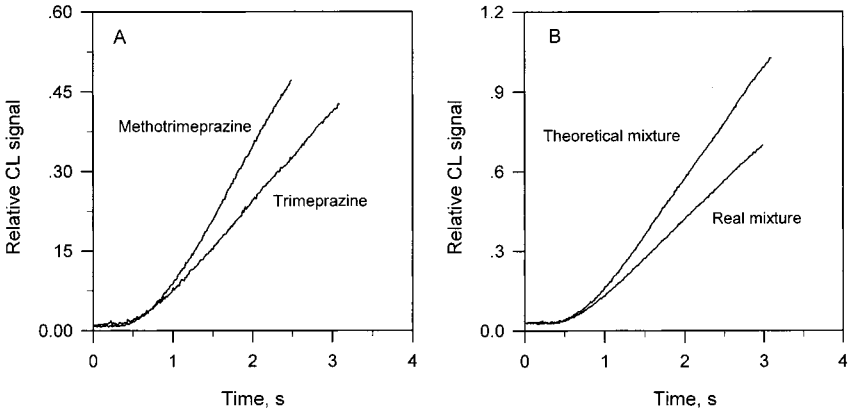


**Figure 1.** Variation of the CL intensity with time for methotrimeprazine and trimeprazine alone (A) and in mixtures (B). [Methotrimeprazine] = [Trimeprazine] = 2 $\mu$g/mL. For details, see text.

provided that the reaction was monitored via the formed product using photometric detection (absorbance measurements).

In the case of a chemiluminescence (CL) reaction, such as $A + R \rightarrow P + h\nu$, the response curve corresponds to two first-order consecutive reaction steps, and taking into account the possible rate equations that can be formulated for each reaction step, the integrated equation can be formulated as[28]

$$S_{CL} = Ck_2[R]_0 \frac{k_1}{k_1 - k_2}(e^{-k_2 t} - e^{-k_1 t})$$

where $S_{CL}$ is the CL signal at any time $t$, $k_1$, and $k_2$ are the rate constant corresponding to two opposite simultaneous first-order processes, and $C$ is a constant related to instrumental features.

When a CL reaction was developed using the CAR technique, such as the chemical system used in this work, the shape of the resulting CL signal vs time plot corresponds to a differential equation (combination of both integrated equations above) which is very difficult to obtain. As can be seen in Figure 1, the CL signal vs time plots show the characteristic initial concave portion that corresponds to a reaction developed using the CAR technique.[24] The ignorance of the differential equation that describes this process makes more difficult the resolution of mixtures of species using this chemical system. This difficulty is increased if both species exhibit a very similar behavior and interact with each other as the reaction develops (i.e., they exhibit a mutual synergistic effect). The high degree of nonlinearity showed by this chemical system does not recommend the use of statistical methods for resolving mixtures in CL-based determinations using the CAR technique, and, therefore, in this work, we have evaluated the strengths and weaknesses of the computational neural network to solve this problem.

We chose the resolution of mixtures of two structurally related phenothiazine derivatives, viz., trimeprazine and methotrimeprazine, to obtain real data on account of the highly similar kinetic behavior of these compounds (Figure 1A) and of the fact that their mixtures exhibit appreciable synergistic effects. As can be seen in Figure 1B, the synergistic effect is apparent considering the difference between the CL curves shown in it: the real mixture curve corresponds to a trimeprazine:methotrimeprazine 1:1 mixture, whereas the theoretical mixture curve corresponds to the addition of each single phenothiazine CL curve at the same concentration as in the 1:1 mixture. In addition, and in order to minimize time and costs in the analytical laboratory, we have used a fairly small sample size. This is another important parameter that shows a high influence on the performance of the CNN.

Solving this problem accurately using CNN involves several steps, namely:

(a) Linear (or nonlinear) filtration of data to reduce the dimensions of available information without detriment of the quality of the results. This filtration step was carried out by using PCA.

(b) Selecting the optimal time region of the CL curve to obtain suitable information in the least possible time.

(c) Selecting what and how many PCs to be used as input to the network with no detriment to the mixture discrimination.

(d) Optimizing the network architecture.

(e) Assessing the discriminating power of the ANN for resolving mixtures of structurally related compounds in CL-based determinations using the CAR technique (viz., the errors obtained for mixtures at different analyte concentration ratios).

(f) The leave-one-out (cross-validation) method was used throughout to calculate errors over the training and test sets.

**Preliminary Filtration of CL Data**. As a rule, the use of a CNN to solve a problem entails preprocessing available data in order to reduce the network input layer complexity and its learning time. One well-known choice for this purpose involves linear or nonlinear filtration of data to reduce the quantity of available information without detracting from quality. The filtration step (data preprocessing) was carried out by using PCA,[29] a widely employed technique for reducing the dimensions of multivariate data while preserving most of the variance. Initially, we used the most significant principal components (PC), which accounted for over 95% of the total variance, as input to the CNN. Thus, the initial CNN architecture was 10:5*s*:2*l*, i.e., one containing 10 nodes in the input layer (one input node per significant PC), 5 nodes in the hidden layer and 2 nodes in the output layer (the trimeprazine and methotrimeprazine concentrations), with sigmoidal, *s*, and linear, *l*, transfer functions, respectively. It is worth noting that this will be reoptimized as described below.

**Optimization of the Time Domain**. To solve the problem addressed in this work, it is necessary to select the

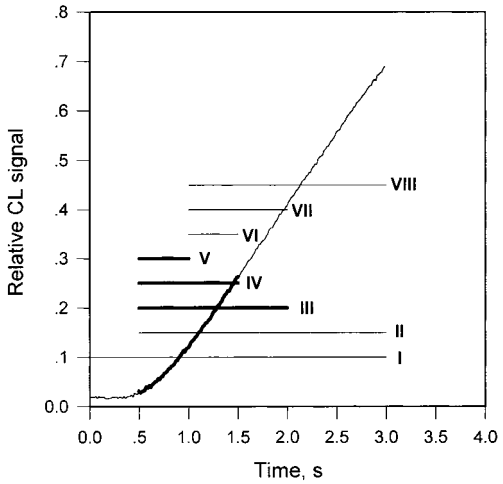


**Figure 2.** Time domains assayed for the analysis of nonlinear multicomponent mixtures in CL-based determinations. The bold lines correspond to the time domains in which the errors over the training set were fewer, and the selected portion of the CL signal vs time curve used for data preprocessing.

**Table 1.** Effect of the Time Domain on % SET in Multicomponent CL-based Determinations Using CNN

| time domain | range, s | % SET (A)[a] | % SET (B)[a] |
|---|---|---|---|
| I | 0.0−3.0 | 8.25 | 4.38 |
| II | 0.5−3.0 | 1.09 | 3.31 |
| III | 0.5−2.0 | 1.26 | 1.39 |
| IV | 0.5−1.5 | 0.84 | 1.02 |
| V | 0.5−1.0 | 1.06 | 1.27 |
| VI | 1.0−1.5 | 3.53 | 2.62 |
| VIII | 1.0−2.0 | 2.76 | 3.54 |
| VIII | 1.0−3.0 | 2.55 | 4.84 |

[a] A, methotrimeprazine; B, trimeprazine.

optimal interval of the CL signal vs time plot for data preprocessing, because the information provided depends strongly on the reaction time interval (time domain) used. Figure 2 shows the different time domain tested, and Table 1 gives the corresponding %SET obtained over the training process. As it can be seen, the selected time domain had a marked effect on the performance of the CNN. On account of the % SET values obtained (1000 epoch for the training set, each epoch corresponds to the presentation of 49 patterns to the network) several conclusions can be drawn: (1) The key for selecting the optimal time domain is the initial concave portion of the CL signal vs time plot and also a suitable interval of its linear portion, such as in III to V intervals. When the linear portion is wider, the additional information does not contribute to the discrimination process and therefore the % SET values increase, as can be seen in I and II intervals. (2) The time domains outside the initial concave portion of the CL curve (VI to VIII) yield higher % SET values in higher or lower extend. On account of these results, we chose region IV (0.5−1.5 s) for further experiments as a compromise.

The effect of the data acquisition rate over the range 20−100 points/s in the selected time domain was also examined. The effect of this variable on % SET was virtually negligible; this suggests that the key to appropriate performance of the CNN is the time domain rather than its associated data acquisition rate. To derive more information from the CL curve, 100 points/s are taken as the optimum acquisition rate.

**Number of PCs Used as Input to the Network**. As noted earlier, selecting the CNN input by preprocessing available data reduces the dimensions of the input space without detriment of the quality in the results. In many cases, using the most significant PCs for this purpose is a good choice (ours accounted for more than 95% of the total variance). However, it has recently been shown that a better discrimination between elements of the same family can occasionally be achieved by discarding the initial PCs since they are associated with the general trends of the family. So they can contribute or not to the discrimination process.[30,31] Unfortunately, there is no general rule for these situations, so heuristic optimization is required, and therefore a heuristic procedure was used in this work to estimate the optimum number of inputs (PCs) to the proposed CNN architecture.

The influence of this variable was examined by testing various neural network architectures (*X*:5*s*:2*l*), where *X* is the number of inputs corresponding to scores used after data were preprocessed by PCA. As can be seen in Figure 3A, network performance was not improved by increasing the number of PCs (>10 on account that the network architecture used up to now is 10:5*s*:2*l*). These additional PCs are probably associated with random noise rather than to a particular data trend, so they had no effect on the capacity of the neural network. However, removing the first PC improved the results, that is, using the second to 10th PCs (see Figure 3B), although, as can be seen in this figure, discarding additional PCs (the second, third, etc.) resulted in a poorer network training quality (viz., in high % SET values). A 9:5s:2l architecture was thus adopted for subsequent experiments using the second to 10th PCs as input to the CNN.

**Optimization of the Network Architecture and Training**. The network architecture was optimized by selecting the most effective number of nodes in the hidden layer (between 2 and 8, corresponding to network designs from 9:2*s*:2*l* to 9:8*s*:2*l*). The % SET values (1000 epochs for all architectures tested) for both components showed a minimum at five nodes (i.e., a 9:5*s*:2*l* architecture). It is observed that if it is slightly increased the number of nodes of the hidden layer, the generalization ability decreases. Based on these results, and taking into account that a simpler network design was desirable, the 9:5*s*:2*l* was selected for subsequent experiments. Due to the fact that a 9:5*s*:2*l* network topology poses the problem of the estimation of 62 parameters (weights) with 50 prototypes, there is certain risk of overfitting. So, it would be advisable to reduce the size of the network in a future work using pruning and genetic algorithms.[32]

Once the network architecture was established, the next step in the optimization procedure involved training it. Based on the results obtained, training could be stopped at about 1000 epochs which yields an average % SET of 1.5%, considering both components in the mixture.

**Generalization Ability of CNNs for Multicomponent CL-Based Determinations**. We assessed the generalization ability of the selected network for resolving mixtures of phenothiazine derivatives subject to synergistic effects. Figure 4 shows typical plots of estimated vs real values for both mixture components. As can be seen from the figure and the corresponding regression parameters, estimated values were quite consistent with real values. % SEP was

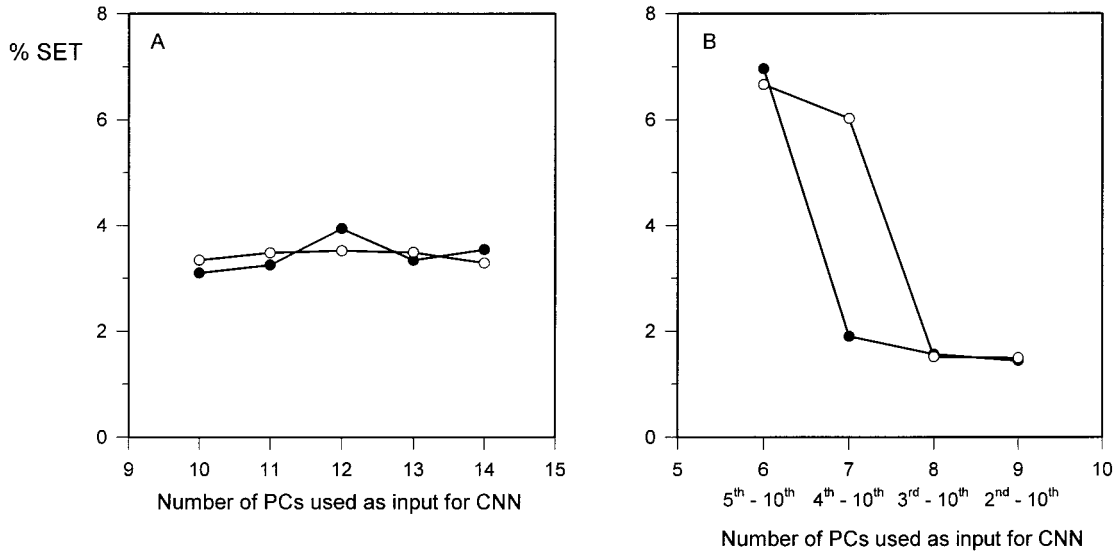


**Figure 3.** Selection of CNN input according to the number of PCs used: (A) without removing any PC; and (B) removing from the first to the fourth PCs. (o) Trimeprazine and (●) methotrimeprazine.

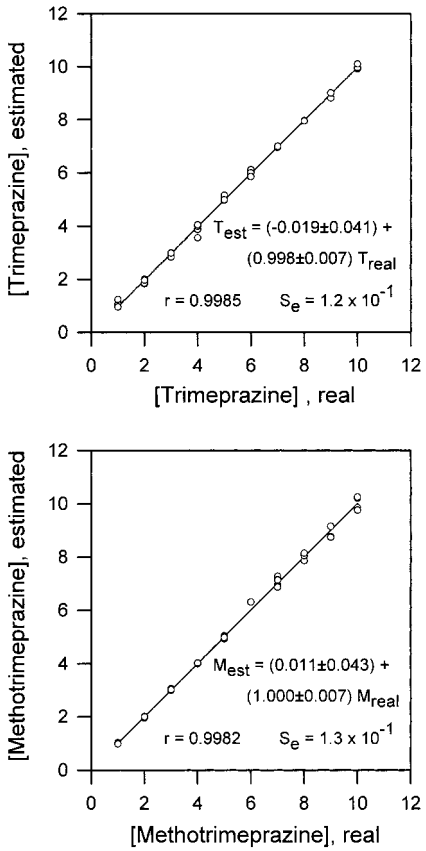


**Figure 4.** Plots of estimated vs real concentrations of both phenothiazine derivatives in the assayed mixtures. T: trimeprazine; M: methotrimeprazine; $S_e$: standard error of estimate.

1.89% and 2.37% for trimeprazine and methotrimeprazine, respectively. These results are quite good and testify to the excellent performance of neural networks in multicomponent CL-based determinations.

The magnitude of the relative errors in the concentration of each component in the mixture can be clearly envisaged from Table 2, which shows the results obtained from various synthetic binary mixtures containing variable amounts of trimeprazine and methotrimeprazine. Mixtures in ratios from 10:1 to 1:10 can be satisfactorily resolved. The relative errors made are less than 5% (except at both ends of the range, where it rises to ca. 10%) and hence quite acceptable for such a wide range of concentration ratios.

**Table 2.** Analysis of Various Trimeprazine/Methotrimeprazine Mixtures by Using the Optimal Network Design

| compound taken ($\mu$g/mL) | | trimeprazine | | methotrimeprazine | |
|---|---|---|---|---|---|
| trimeprazine | metho-trimeprazine | found ($\mu$g/mL) | rel error (%) | found ($\mu$g/mL) | rel error (%) |
| 3.00 | 3.00 | 2.99 | −0.33 | 2.99 | −0.33 |
| 2.00 | 4.00 | 1.99 | −0.50 | 3.99 | −0.25 |
| 2.00 | 5.00 | 2.01 | 0.50 | 5.01 | 0.20 |
| 1.00 | 3.00 | 0.98 | −2.00 | 2.99 | −0.33 |
| 2.00 | 8.00 | 1.92 | −4.00 | 8.10 | 1.25 |
| 1.00 | 10.00 | 1.12 | 12.0 | 10.00 | 0.00 |
| 9.00 | 1.00 | 8.76 | −2.66 | 1.10 | 10.0 |
| 10.00 | 2.00 | 10.01 | 0.10 | 2.03 | 1.50 |
| 8.00 | 2.00 | 8.00 | 0.00 | 2.00 | 0.00 |
| 6.00 | 2.00 | 5.97 | −0.50 | 1.99 | −0.50 |
| 10.00 | 4.00 | 9.99 | −0.10 | 4.01 | 0.25 |
| 10.00 | 5.00 | 10.05 | 0.50 | 5.04 | 0.80 |

## ACKNOWLEDGMENT

The authors gratefully acknowledge financial support from Spain's Dirección General Interministerial de Ciencia y Tecnología (DGICyT) to carry out this research in the framework of Project PB96-0984 (Department of Analytical Chemistry, University of Córdoba) and Project TIC97-0842-C02-02 (Department of Computer Science, University of Córdoba).

## REFERENCES AND NOTES

(1) Grayeski, M. L. Chemiluminescence analysis. *Anal. Chem.* **1987**, *59*, 1243A−1254A.
(2) Townshend, A. Solution chemiluminescence-some recent analytical developments. *Analyst* **1990**, *115*, 495−500.
(3) Robards, K.; Worsfold, P. J. Analytical applications of liquid-phase chemiluminescence. *Anal. Chim. Acta* **1992**, *266*, 147−173.
(4) Kwakman, P. J. M.; Brinkman, U. A. Th. Peroxyoxalate chemiluminescence detection in liquid chromatography. *Anal. Chim. Acta* **1992**, *266*, 175−192.
(5) Kamidate, T.; Ishikawa, A.; Segawa, T.; Watanabe, H. Simultaneous determination of cysteine and glutathione via use of time-resolved luminol chemiluminescence method. *Chem. Lett.* **1993**, *12*, 2049−2052.

(6) Navas-Díaz, A.; González-García, J. A. Nonlinear multicomponent kinetic analysis for the simultaneous stopped-flow determination of chemiluminescence enhancers. *Anal. Chem*. **1994**, *66*, 988−993.
(7) Zupan, J.; Gasteiger, J. *Neural Networks for Chemists. An Introduction*; VCH: Weinheim, 1993.
(8) Ventura, S.; Silva, M.; Pérez-Bendito, D.; Hervás, C. Artificial neural network for estimation of kinetic analytical parameters. *Anal. Chem.* **1995**, *67*, 1521−1525.
(9) Schulz, H. Neural networks in analytical chemistry-A new method or only a passing fashion? *GIT. Fachz. Lab*. **1995**, *39*, 1009−1010.
(10) Munk, M. E.; Madison, M. S.; Robb, E. W. The neural network as a tool for multispectral interpretation. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 231−238.
(11) Ventura, S.; Silva, M.; Pérez-Bendito, D.; Hervás, C. Estimation of parameters of kinetic compartmental models by use of computational neural networks. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 517−521.
(12) Zupan, J.; Novic, M.; Ruisanchez, I. Kohonen and counterpropagation artificial neural networks in analytical chemistry. *Chemom. Intell. Lab. Syst*. **1997**, *38*, 1−23.
(13) HorniK, K.; Stinchcombe, M.; White, H. Multilayer feed-forward networks are universal approximators. *Neural Networks* **1989**, *2*, 359−366.
(14) Cybenko, G. Approximation by superpositions of a sigmoidal function. *Math. Control Signals Systems* **1989**, *2*, 303−314.
(15) Borron, A. R. Universal approximation bounds for superposition of a sigmoidal function. *IEEE Transations Information Theory* **1993**, *39*, 930−945.
(16) Mhaskar, H. N. Neural networks for localized approximation of smooth and analytic functions. *Neural Networks* **1993**, *3*, 190−196.
(17) Hornik, K.; Stinchcombe, M.; White, H.; Aner, P. Degree of aproximation results for feed-forward networks approximating unknown mappings and their derivatives. *Neural Computation* **1994**, *6*, 1262−1275.
(18) Attali, J. G.; Page's, G. Approximations of functions by a multilayer perceptron: A new approach. *Neural Networks* **1997**, *10*, 1069−1081.
(19) Stone, M. Cross-validation: A review. *Math. Operations forsh. Statist. Ser. Statistics* **1978**, *9(1).*
(20) Eubank, R. L. *Spline smoothing and nonparametric regression;* Marcel Dekker Inc.: 1988.
(21) Moody, J. *Prediction risk and architecture selection for neural networks*. From *Statistics to neural network*; Springer-Verlag: 1994.
(22) Moody, J.; Utans, J. *Architecture selection strategies for neural networks: Application to corporate bond rating prediction*. In *Neural networks in the capital markets;* Refenes, A. N., Ed.; John Wiley & Sons: 1994.
(23) Cepas, J.; Silva, M.; Pérez-Bendito, D. Improved peroxyoxalate chemiluminescence based determinations by use of continuous reagent addition to remove background emission. *Anal. Chem*. **1994**, *66*, 4079−4084.
(24) Velasco, A.; Silva, M.; Pérez-Bendito, D. Processing analytical data obtained from second-order reactions using continuous reagent addition. *Anal. Chem*. **1992**, *64*, 2359−2365.
(25) Stone, M. Cross-validatory choice and assessment of statistical predictions *J. Royal Statist. Soc.* **1974**, *B36*, 111−147.
(26) Eubank, R. L. *Spline smoothing and nonparametric regression;* Marcel Dekker: New York, 1988.
(27) Wahba, G. *Spline models for observational data*; Vol. 59 of Regional Conference Series in Applied Mathematics, SIAM Press: 1990.
(28) González-Robledo, D.; Silva, M.; Pérez-Bendito, D. Performance of the stopped-flow technique in chemiluminescence spectrometry based on direct rate measurements. *Anal. Chim. Acta* **1989**, *217*, 239−247.
(29) Flury, B.; Riedwyl, H. *Multivariate statistics: A practical approach;* Chapman & Hall: London, 1988.
(30) Oja. E. Neural networks, principal components and subspaces, *Int. J. Neural Systems.* **1989**, *1*, 61−68.
(31) Oja E. Principal components, minor components, and linear neural networks. *Neural Networks* **1992**, *5*, 927−935.
(32) Shamir, N.; Marom, E. Neural net pruning based on functional behaviour of neurons. *Int. J. Neural Systems* **1993**, *4*, 143−158.

CI980030+