Estimation of Parameters of Kinetic Compartmental Models by Use of Computational Neural Networks

Sebastián Ventura,[†] Manuel Silva,[†] and Dolores Pérez-Bendito*,[†]

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

César Hervás[‡]

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

Received October 24, 1996[⊗]

A novel methodological approach to the estimation of parameters involved in multicomponent kinetic determinations from real kinetic data by use of computational or artificial neural networks (ANNs) is proposed. The ANN input data used are also estimates obtained by using the Levenberg−Marquardt method in the form of an approximate nonlinear function that is the sum of the two expressions associated with the *pseudo*-first-order kinetics of the two mixture components. The performance of the optimized network architecture, 2:4s:2l, was tested at variable rate constant ratios. The reduced dimensions of the network input space obtained using the Kolmogorov−Sprecher theorem result in improved limits of precision in estimating parameters at near-unity rate constant ratios. Experiments with real kinetic data provided a relative standard error of prediction of 2.47% and 4.23% for the two mixture components. These errors are much smaller than those obtained with existing alternative methods, particularly at the low rate constant ratio involved (1.37).

INTRODUCTION

Computational neural networks (ANNs) are among the most exciting developments in computational science in recent years. ANN-based methods are becoming increasingly popular in various branches of science including chemistry.[1] In analytical chemistry, the use of ANNs has chiefly been concerned with the analysis of a wide variety of data related to pattern recognition/classification for spectral analysis[2,3] and as a modeling method for quantitative structure−activity or structure−property relationships (QSAR and QSPR, respectively).[4] The potential of ANNs in kinetic analysis was recently assessed with the estimation of kinetic analytical parameters using real kinetic data acquired after a short reaction time[5] as well as in multicomponent kinetic determinations with the aid of spectral discrimination based on the scores of a principal component model as input data for calibration in the resolution of binary mixtures.[6]

This paper describes the use of ANNs for multicomponent kinetic determinations using a novel methodological approach for selecting input data for the ANN in order to attain the simplest architecture and shortest training time. Thus, the reaction rate constant and the observed total analytical signal change (estimated by using the Levenberg−Marquardt method[7]) of an approximate nonlinear function of the same family as each of addend functions associated with the *pseudo*-first-order kinetics of both components in the mixture were used as the input data. The experimental results obtained from analytical chemical data testify to the usefulness of the proposed methodology for establishing the optimal network topology; in fact, a small number of weights allows estimates for the concentrations of the mixture components to be obtained with minimal coefficients of variation.

Multicomponent kinetic determinations are an aspect of great interest in kinetic methods of analysis. Multivariate calibration and Kalman filter methods[8] are effective approaches to multicomponent kinetic determinations inasmuch as they yield quite accurate results are and allowed to operate with similar rate constant values by combining kinetic and spectral discrimination. Application of these approaches therefore requires some spectral differences and is unfeasible otherwise (*e.g.*, in the resolution of mixtures of many organic compounds by redox reactions, which make the core of noncatalytic kinetic applications). This problem is overcome in this paper by using ANNs for multicomponent kinetic determinations of species with rate constant ratio approaching unity without the aid of spectral discrimination.

THEORETICAL BACKGROUND AND METHODOLOGY

Let A and B be two substances to be determined in a mixture and R a reagent with which both react to yield two products P and P′ which, though different, are similar with regards to the analytical property to be measured (*e.g.*, absorbance). If the reaction rates for the two components are different and correspond to first-order kinetics in each, then the signal (absorbance) measured at time t will be given by

$$S_t = a(1-e^{-k_A t}) + b(1-e^{-k_B t}) + \epsilon \qquad (1)$$

where k_A and k_B are the respective *pseudo*-first-order rate constants; a and b are two parameters directly related to the concentrations of the two components in the mixture; and ϵ is a residual term associated with the background contribution.

Provided the shape of the analytical curves is known and can be derivated, the four model parameters (a, k_A, b, and k_B) can be estimated by nonlinear regression (*e.g.*, by using

[†] E-mail: mas2vesos@lucano.uco.es.
[‡] E-mail: ma1hemac@lucano.uco.es.
[⊗] Abstract published in *Advance ACS Abstracts,* March 15, 1997.

S0095-2338(96)00143-6 CCC: $14.00

the Levenberg−Marquardt method). The minimum of the square error function is obtained by using a descent method. One problem in the estimation of these parameters arises from the fact that when the rate constant ratio, k_A/k_B, is lower than 2 or 3, the estimated values for a and b depart from their true values, even if the true values of k_A and k_B are known. This is so because the descrimination ability for this type of curve, which depends on the time-independent experimental error, tends to disappear when $r_k = k_A/k_B$ tends to one. This can be demonstrated easily.

Let S_t^1 and S_t^2 be two kinetic curves of *pseudo*-first-order where a and b for S_t^1, and c and d for S_t^2, are such that $a + b = c + d$, *i.e.*, the signals are indistinguishable when $t \rightarrow \infty$. We selected this case since, if $a + b \neq c + d$, the two curves will be distinguishable provided the difference between the two sides of the equation is significant. The difference between the signals, ΔS, is given by

$$\Delta S = (a-c)(1-e^{-k_A t}) + (b-d)(1-e^{-k_B t}) + \epsilon_1 - \epsilon_2 \quad (2)$$

The maximum of this function can be obtained by derivation and equalling to zero:

$$\left(\frac{d\Delta S}{dt}\right)_{t=t_M} = (a-c)k_A e^{-k_A t_M} + (b-d)k_B e^{-k_B t_M} = 0 \quad (3)$$

Since $a - c = d - b$,

$$\frac{k_A}{k_B} = e^{-(k_B-k_A)t_M} \quad (4)$$

Therefore, the time t_M at which ΔS is maximal will be given by

$$t_M = \frac{1}{k_A - k_B}\ln r_k \quad (5)$$

provided $(d^2\Delta S/dt^2) > 0$, which is true if $r_k > 1$.

Substitution into eq 2 gives the maximum as a function of a, c, and r_k of the following form:

$$(\Delta S)_{max} = (a-c)((r_k)^{1/(1-r_k)} - (r_k)^{r_k/(1-r_k)}) \quad (6)$$

from which it is apparent that

$$\lim_{r_k \rightarrow 1}(\Delta S)_{max} = 0 \quad \text{(q.e.d.)}$$

In view of the fact that no accurate estimators for a and b at $r_k \rightarrow 1$ could be obtained, we developed a method for their indirect estimation that is described below.

The method for solving the problem addressed involves assuming that, for compartmental models,[10] which typically consists of linear combinations of exponential functions such as that considered here,

$$d(S_t/S) = d(S_{t_1}/S_A) + d(S_{t_2}/S_B) = -k_A S_{t_1} - k_B S_{t_2} \quad (7)$$

the solution to the differential equation can be approximately of the following form

$$S_t = S(1-e^{-kt}) \quad (8)$$

rather than that of eq 1.

If the square error made in using eq 8 instead of (1) is small enough, then parameters S and k in eq 8 can be estimated by nonlinear regression and used to estimate the true solution parameters for the differential equation system 7 from a system of nonlinear equations.

The square error made at each point in using eq 8 rather than (1) is given—constants excepted—by

$$\Delta S_t = (ae^{-k_A t} + be^{-k_B t} - Se^{-kt})^2$$

$$= a^2e^{-2k_A t} + b^2e^{-2k_B t} + S^2e^{-2kt} - 2abe^{-(k_A+k_B)t} - 2aSe^{-(k_A+k)t} - 2bSe^{-(k+k_B)t} \quad (9)$$

Therefore, the square error throughout the reaction time can be obtained by integrating eq 9 from 0 to ∞:

$$\Delta S_T = \int_0^\infty (\Delta S)^2\, dt = \frac{a^2}{2k_A} + \frac{b^2}{2k_B} + \frac{S^2}{2k} + \frac{2ab}{k_A + k_B} + \frac{2aS}{k_A + k} + \frac{2bS}{k_B + k} \quad (10)$$

In order to find the a and b values that minimize ΔS_T, one can obtain the partial derivatives with respect to such parameters and equal them to zero:

$$\frac{a}{k_A} - \frac{2b}{k_A + k_B} = \frac{2S}{k_A + k}$$

$$\frac{b}{k_B} - \frac{2a}{k_A + k_B} = \frac{2S}{k_B + k} \quad (11)$$

Substituting $k_A = r_k k_B$ into eq 11 and solving for a and b yields

$$a = \frac{b(r_k+1)}{2} - \frac{S(r_k+1)k_B}{k_B + k} \quad (12)$$

$$b = \frac{\dfrac{S(r_k+1)^2 k_B}{k_B + k} + \dfrac{2Sr_k(r_k+1)k_B}{r_k k_B + k}}{\dfrac{(r_k+1)^2}{2} - 2r_k}$$

Equation 12 provides the approximate nonlinear equations that relate a and b indirectly with S_t (via S and k) at given k_B and r_k values.

These a and b values result in a minimum for the error function since the Hessian H

$$H = \begin{vmatrix} \dfrac{1}{r_k k_B} & \dfrac{-2}{k_B(1+r_k)} \\ \dfrac{-2}{k_B(1+r_k)} & \dfrac{1}{k_B} \end{vmatrix} = \frac{(1-r_k)^2}{k_B{}^2 r_k (1+r_k)^2} > 0 \quad (13)$$

is greater than zero at any $r_k \neq 1$, and the second derivative of ΔS_T with respect to a is also greater than zero.

However, when $r_k \rightarrow 1$, from eq 12 it follows that $b \rightarrow \infty$, so $a \rightarrow \infty$ provided $\Delta S_T \rightarrow 0$; therefore, even if the Hessian H is zero at $r_k = 1$, the mean square error for eq 10 will not only exhibit a minimum but also be zero. Consequently, while the approximation is very good, the estimates obtained for a and b are not.

After S and k were estimated by using the Levenberg−Marquardt method, and their respective estimators S^* and k^* were obtained, we designed a computational neural network

Table 1. Variation of the Percentage of Standard Error of Estimation of a, b, k_A and k_B as a Function of r_k[a]

	$\kappa = 0.0001$[b]				$\kappa = 0.0005$[b]				$\kappa = 0.001$[b]			
r_k	%SE a	%SE k_A	%SE b	%SE k_B	%SE a	%SE k_A	%SE b	%SE k_B	%SE a	%SE k_A	%SE b	%SE k_B
4.0	2.26	2.15	1.41	0.76	1.97	1.62	1.98	1.83	1.54	1.35	2.01	2.12
3.0	1.80	0.87	2.14	1.56	2.40	1.70	2.56	1.81	4.25	2.12	5.45	2.74
2.0	5.68	2.10	6.74	4.59	7.26	2.53	7.32	4.78	24.58	11.11	24.81	9.34
1.5	30.85	6.30	36.59	10.11	48.57	12.87	50.07	19.45	63.48	18.45	59.42	16.45
1.3	45.92	9.43	54.25	13.90	74.95	34.74	79.90	45.70	~100.0	~100.0	~100.0	~100.0

[a] Real parameters in the kinetic curves: $a = 0.15-0.75$; $b = 0.15-0.75$; $k_A = 0.04$. Input parameters in all regressions: $a = b = 0.5$; $k_A = k_B = 0.01$. [b] The noise imposed to the kinetic data follows a normal distribution in which the mean is zero and the typical deviation depends on the signal according to $\sigma = \kappa\ 10^{\text{signal}}$.

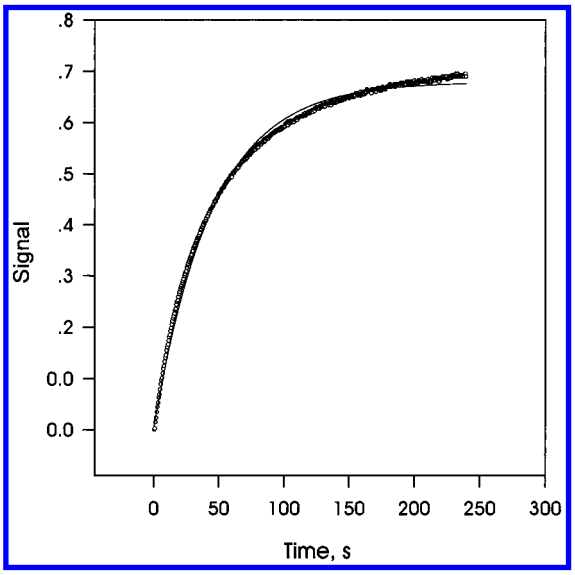

Figure 1. Experimental (●) and simulated kinetic curves (—) for a mixture resolved in experiments using real data. The simulated curve was constructed from eq 1 and the parameters obtained by using the Levenberg−Marquardt method.

capable of learning and generalizing the approximate relations of eq 12 with S^* and k^* as input and parameters a and b as output.

RESULTS

First, we used the nonlinear regression method (NLIN) to assess the quality of a and b estimates. Table 1 shows the variation of the percentage of standard error (%SE) of estimation of a, b, k_A, and k_B as a function of the rate constant ratio for simulated data, obtained by using the NLIN procedure included in the SAS/STAT software package,[8] for the estimations. The results were obtained from three training sets (25 simulated kinetic curves) at different levels of imposed random Gaussian noise of zero mean and standard deviations $10^{-4} \times 10^{\text{Abs}}$, $5 \times 10^{-4} \times 10^{\text{Abs}}$, and $10^{-3} \times 10^{\text{Abs}}$ in order to be able to extrapolate any conclusions to real data. As can be seen, the goodness of the estimate for each parameter decreased with decreasing rate constant ratio; thus, the estimates for ratios below 2 were rather unfavorable. The effect increased as the noise associated with the kinetic curves was raised.

One other reason for the poor estimation parameters lies in the fact that the descending direction of the gradient at the points on the error surface is not completely correct—as can be experimentally observed, the kinetic profile for the mixture does not coincide exactly with the sum of the two exponential functions in eq 2 (Figure 1).

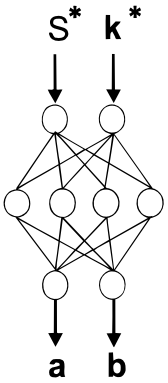

Figure 2. Network architecture used to estimate parameters a and b.

Second, due to the poor results previously obtained for the estimation of parameters a and b at $r_k \rightarrow 1$, a multilayered neural network was designed using S_t values acquired at different times as input and the parameters to be estimated (a and b) as output. The above-mentioned training sets yielded mean percentages of standard error of prediction (%SEP) greater than 20%, so this methodology was also inefficient, basically because of the high correlation between data, which resulted in covariance matrices with an incomplete rank; this called for linear or nonlinear filtering in order to compress the information without detracting from learning stability, which would be rather costly with regards to optimizing the network architecture and the time needed to train it in order to estimate a and b accurately.

Third, we used linear filtering of S_t data in order to cancel the potential effect on a and b estimates of the high correlation between both parameters. Thus, we carried out a principal component analysis to reduce the dimensions of the input space and obtain new, incorrelated variables. On the basis of a heuristic study, the first three components were chosen as optimal since they accounted for 99% of the overall variance of the initial data. Using these three variables as input and parameters a and b as output, the best network designs obtained yielded mean %SEP errors about 15%, so linear filtering methodology also proved inefficient.

Because the a and b estimates obtained with the previous methods were all unfavorable, we considered using the above-described indirect estimation methodology.

The network architecture with first-order learning algorithms that led to the best results using the same simulated kinetic data was 2:4s:2l, *i.e.*, two input data (S^* and k^*, two estimates obtained by nonlinear regression of the kinetic curve for the mixture by using a heuristically determined topological region[11]), four nodes in the hidden layer, and two output data (a and b), with sigmoidal (s) and linear (l)

Table 2. Effect of the Percentage of Standard Error of Prediction of *a and b*

	$\kappa = 0.0001$[a]		$\kappa = 0.0005$[a]		$\kappa = 0.001$[a]	
r_k	%SEP *a*	%SEP *b*	%SEP *a*	%SEP *b*	%SEP *a*	%SEP *b*
1.1	3.15	3.40	4.20	5.62	9.70	8.74
1.3	2.25	1.76	3.87	3.71	8.98	6.72
1.5	1.90	1.60	3.27	3.23	7.60	6.21
1.7	1.35	1.50	2.82	3.40	6.39	5.88
2.0	1.60	1.35	2.75	2.86	6.48	4.39
3.0	1.85	1.55	3.08	3.13	7.38	5.05
5.0	3.00	2.65	5.06	4.98	8.86	8.67
10.0	5.40	5.10	9.16	9.05	10.52	11.05

[a] The same meaning as in Table 1.

transfer functions, respectively. We used a linear transfer function in the output layer because the problem involved estimating parameters. The learning algorithm used was the extended delta-bar-delta (EDBD) rule,[12] with the following initial parameters: weighting factor ($\theta = 0.7$), scaling factors for the learning rate increment and momentum coefficient, respectively ($\kappa_\eta = 0.095$ and $\kappa_\mu = 1.0$), exponential factors for the increment of these coefficients ($\gamma_\eta = 0.1$ and $\gamma_\mu = 0.05$), and decrement factors ($\phi_\eta = 0.1$ and $\phi_\mu = 0.01$). Table 2 shows the results of this study. As can be seen, the variation of %SEP as a function of r_k for the previous network design exhibited global minima at $r_k = 1.7$ for *a* and $r_k = 2$ for *b*, which confirms the ability of ANNs to provide good estimates for *a* and *b*, especially at near-unity r_k values, where classical estimation methods are inefficient. On the other hand, the estimation errors made at $r_k > 4$ are comparable to those obtained directly by nonlinear regression, so constructing an ANN model is superfluous.

Finally, we validated the proposed indirect estimation method by using experimental kinetic data obtained in the resolution of mixtures of 2-chlorophenol (2-ClPh) and 3-chlorophenol (3-ClPh) by their oxidative coupling with *N*,*N*-diethyl-*p*-phenylenediamine,[13] the rate constant ratio for which was $r_k = k_{2-\mathrm{ClPh}}/k_{3-\mathrm{ClPh}} = 1.37$. The %SEP obtained after 45 000 iterations (1500 cycles for 30 standard mixtures of the two phenolic compounds containing concentrations over the range 2−10 μM) on the above-described network architecture were 2.47% for 2-ClPh (*a*) and 4.23% for 3-ClPh (*b*). These errors are much smaller than those achieved by other existing alternative methods, such as the Kalman filter.[13]

DISCUSSION

The proposed method provides estimates for the approximate exponential function that are used as input to an ANN with back-propagation and the EDBD learning algorithm—as shown above, the results obtained by using the curve envelope are rather poor, both with nonlinear regression and with ANNs. A very simple network design is used, so overparametrization of the model is avoided. As can be seen from the equations, parameters *a* and *b* are nonlinear functions of the estimates fed to the network.

Also worth mentioning is the fact that the use of ANNs for the quantitative representation of kinetic compartmental models relies on the Kolmogorov−Sprecher theorem[14,15] (as implemented by Nielsen[16] and Cybenko[17]) and leads to optimal precision for a small input space (of two dimensions in our case). Thus, by using preliminary estimates to reduce the dimensions of the input space, an acceptable precision limit for the representation of the multicomponent kinetic function is obtained: 1.35% of the maximum of the $a + b$ function in our case for simulated signals with a standard deviation for the error associated with the signal $\sigma = 0.0001$—by contrast, recent research revealed that the root mean square error of estimation (RMSE) for the spectral peak fitting procedure is usually not lower than 2−3% of the function maximum.[11,18] Therefore, the precision obtained in estimating *a* and *b* at near-unity r_k values is higher than that obtained by alternative methods.

CONCLUSIONS

In this work we developed a novel method for estimating multicomponent parameters based on an approximate solution of kinetic compartmental models. The method allows the nonlinear estimation of functions consisting of combined exponentials when the ratio of the parameters in the exponents is close to unity.

The proposed method provides estimates subject to a very small standard error relative to the direct use of NLR. It uses ANNs consisting of S_t signals or the optimal principal components as input.

The theoretical background for the proposed method affords application to this type of behavioral model involving a k_A/k_B ratio close to unity in other fields.

ACKNOWLEDGMENT

The authors gratefully acknowledge financial support from the Spanish *Dirección General Interministerial de Ciencia y Tecnología* (DGICyT) for the realization of this research in the framework of Projects PB91-0840 (Department of Analytical Chemistry, University of Córdoba) and TIC90-0648 (Department of Mathematics and Computer Science, University of Córdoba).

REFERENCES AND NOTES

(1) Gasteiger, J.; Zupan, J. *Neural Networks for Chemists: An Introduction*; VCH Publisher: Weinheim, 1993.

(2) Borggaard, C.; Thodberg, H. H. Optimal Minimal Neural Interpretation of Spectra. *Anal. Chem.* **1992,** *64,* 545−551.

(3) Brown, S. D.; Sum, S. T.; Despagne, F.; Lavine, B. K. Chemometrics. *Anal. Chem.* **1996**, *68*, 21R-61R.

(4) Nefati, H.; Cense, J. M.; Legendre, J. J. Prediction of the Impact Sensitivity by Neural Networks. *J. Chem. Inf. Comput. Sci.* **1996,** *36,* 804−810.

(5) Ventura S.; Silva M.; Pérez-Bendito D.; Hervás C. Artificial Neural Networks for Estimation of Kinetic Analytical Parameters. *Anal. Chem.* **1995**, *67,* 1521−1525**.**

(6) Blanco, M.; Coello, J.; Iturriaga, H.; Maspoch, S.; Redón, M. Artificial Neural Networks for Multicomponent Kinetic Determinations. *Anal. Chem.* **1995,** *67,* 4477−4483.

(7) Seber, G. A. F.; Wild, C. J. *Nonlinear Regression*; John Wiley & Sons: New York, 1989; pp 624−627.

(8) Pérez-Bendito D.; Silva M. Recent Advances in Kinetometrics. *Trends in Anal. Chem.* **1996**, *15*, 232−240.

(9) SAS/STAT. *User's Guide,* SAS; Institute Inc.: Cary, NC, 1989.

(10) Seber, G. A. F.; Wild, C. J. *Nonlinear Regressions;* John Wiley & Sons: 1989; Chapter 8.

(11) Burden, F. R. Mapping Analytic Functions Using Neural Networks. *J. Chem. Inf. Comput. Sci.* **1994,** *34,* 1229−1231.

(12) Minai, A. A.; Williams, D. *Acceleration of Backpropagation Through Learning Rate and Momentum Adaptation.* International Joint Conference on Neural Networks, 1990; Vol. Y, pp 676−679.

(13) Velasco, A.; Rui, X.; Silva, M.; Pérez-Bendito, D. Simultaneous kinetic determination of phenols by use of the Kalman filter. *Talanta* **1993**, *40*, 1505−1510.
(14) Kolmogorov, A. N. On the Representation of Continuous Functions of Many Variables by Superposition of Continuous Functions of One Variable and Addition. *Doklady Kademii Nauk SR* **1957,** *144*, 679−681.
(15) Sprecher, D. A. On the Structure of Continuous Functions of Several Variables. *Transl. Am. Math. Soc.* **1965,** *115*, 340−355.
(16) Hecht-Nielsen, R. Kolmogorov Mapping Neural Network Existence Theorem. *IEEE ICNN* **1987,** *3*, 11−13.
(17) Cybenko, G. Approximation by Superpositions of a Sigmoidal Function. *Math Control Signals Systems* **1989,** *2*, 303−314.
(18) Maggiora, G. M.; Elrod, D. W.; Trenary, R. G. Computational Neural Nets as Model-Free Mapping Devices. *J. Chem. Inf. Comput. Sci.* **1992,** *32*, 732−741.

CI960143Y