Computational Neural Networks in Conjunction with Principal Component Analysis for Resolving Highly Nonlinear Kinetics

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

Department of Analytical Chemistry, Faculty of Sciences, and Department of Mathematics and Computer Science, University of Córdoba, E-14004 Córdoba, Spain

Received June 18, 1996[⊗]

A method based on the use of an orthogonal linear filter, principal component analysis (PCA), for preprocessing data used as input for a feed-forward neural network is proposed. The method analyzes the significance of the eigenvalues of the correlation matrix associated with the first principal components of the data in order to select the subset of principal components for the sample that provides the optimum generalization value. The generalization error was estimated by using the leave-one-out method, because it provides the most reliable results for the fairly small data set used. The performance of the proposed method was assessed by applying it to the resolution of mixtures of species exhibiting a very similar kinetic behavior in the presence of a mutual kinetic (synergistic) effect. In addition, use of the continuous-addition-of-reagent (CAR) technique, a second-order approach, increased the nonlinearity of the system studied. Based on the results, the proposed designs provide accurate estimates in the kinetic resolution of binary mixtures, with errors of prediction about 5%. The results obtained in this respect are quite good taking into account that the kinetic behavior of the mixtures studied conforms to highly complex differential equations.

INTRODUCTION

Methods for kinetic multicomponent determinations have proliferated in the last few years; most use mathematical models intended to reproduce the kinetics of the reaction involved. Prominent among such methods are parametric statistical methods such as Kalman filtering.[1−4] Typically, the methods are applied to first- or *pseudo*-first-order kinetics and rely on the assumption that the reaction rates for the analytes are mutually independent. However, this assumption is incorrect for some mixtures of analytes whose reaction rates are not independent; such systems are said to be subject to a mutual kinetic or synergistic effect. Synergism is one of the major sources of error in classical kinetic methods for multicomponent determinations. While the effect of synergism can be included in the rate law for the overall kinetics of the process in various ways (*e.g.*, by including an additional cross-term in the rate equation[5]), there is no universal methodology for solving the problem as the chief result of the difficulty involved in constructing an accurately descriptive theoretical model.

In the past decade, artificial neural networks (ANNs) have aroused much interest among scientists as a powerful tool for solving a wide variety of problems. One of its most salient features is the ability to model data with no prior knowledge of the relationship between the variables that define the system. In the chemical domain, ANNs have been used to model functional relationships for a host of chemical systems in connection with industrial processes,[6,7] quantitative structure activity relationships (QSAR),[8,9] and multivariate calibration.[10−12] As regards kinetic analysis, the estimation of kinetic parameters for individual determinations[13] and the resolution of binary mixtures of species following first-order kinetics[14] were recently reported. In the latter case, the neural network was used to construct an empirical model for the complex relationship between two characteristic parameters of the curve profile obtained by nonlinear regression (NLR) and the concentrations of both analytes. Also, mixtures of species exhibiting first-order kinetics were resolved with the aid of spectral discrimination methodology,[15] data being preprocessed by PCA.

The purpose of this work was to assess the potential of ANNs for modeling complex kinetic processes (specifically, multicomponent analysis of systems involving high nonlinearity such as that resulting from the presence of synergistic effects). The proposed methodology allows data to be preprocessed in order to obtain a reduced network input space by using principal components analysis. This methodology analyzes the significance of the eigenvalues of the correlation matrix associated with the first principal components of the data in order to select the subset of principal components for the sample that provides the optimum generalization value. This methodology was validated with the simultaneous determination of cortisone and hydrocortisone in mixtures by reaction with Tetrazolium Blue, which was implemented by the continuous-addition-of-reagent (CAR) technique. This chemical system was chosen on account of the similarity between the rate constants for the two corticosteroids.

THEORY

In order to expedite convergence of the learning algorithm on a given ANN and analyze the effect of a potential reduction in the dimensions of the input space, authors usually subject input pattern, **x**, to some filtering.[16−18] The transformation used can be expressed as follows:

$$\mathbf{g} = \mathbf{WFx}$$

where $\mathbf{x}_{n\times 1}$ is a column vector of matrix $\mathbf{X}_{n\times m}$, n being the number of characteristics and m that of pattern; $\mathbf{F}_{l\times n}$ is the

[†] Department of Analytical Chemistry.
[‡] Department of Mathematics and Computer Science.
[⊗] Abstract published in *Advance ACS Abstracts,* November 1, 1996.

S0095-2338(96)00084-4 CCC: $14.00

matrix representing the filtering operation; $\mathbf{W}_{l\times l}$ is the network input weight matrix (*i.e.*, the matrix of connections between the input and hidden layer); and $\mathbf{g}_{l\times 1}$ is the vector of inputs to the hidden layer of the network. If the filter, of matrix **F**, is an orthogonal projection, then the back-propagation algorithm for the network will converge on a global minimum at least as rapidly as in the absence of filtering. Therefore, by using a reasonable filtering process, one can expect filter **F** to substantially reduce the dimension of the network input space with no loss of contractiveness in the algorithm (*i.e.*, in the number of cycles or iterations needed for convergence to be attained).[19]

Because the number of characteristics of our input pattern—in our case $n = 600$ because we used 600 points per curve—was much greater than the dimension of vector **g**, the number of nodes in the hidden layer (we assayed networks of 4 and 5 nodes), matrix $(\mathbf{WF})_{l\times n}$ has a very low rank relative to its number of columns, and one can choose a filter **F** such that the weight matrix will obey the following relation

$$\mathbf{W}^1\mathbf{x} = \mathbf{WFx}$$

for any matrix $\mathbf{W}^1{}_{n\times n}$ and input standard **x**; however, this weight matrix will result in a matrix **F** with a number of rows not greatly exceeding the rank of **W**, which is the dimension of **g**. This choice allows the number of free parameters in the learning system (*viz.*, the number of elements in **W** relative to that of elements in $\mathbf{W}^1$) to be dramatically reduced, with no change in the minimum least-squares error obtained or the learning and generalization capacity of the network. Because the previous equation holds for any standard, the aim is to have

$$\mathbf{W}^1\mathbf{X} = \mathbf{WFX}$$

Obviously, this equation cannot be exactly obeyed. However, an approximate solution to this problem can be obtained from the singular value decomposition (SVD) of **X**[20] since, if the eigenvalues of **X** are $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_l$, then it is demonstrated that the best choice of **F** for reducing the dimension of **x** and simultaneously minimizing the difference between $\mathbf{W}^1\mathbf{x}$ and **WFx** in Euclidean metrics is a truncated expansion of **F** into its PCs.[19] In this way, the error made in replacing $\mathbf{W}^1$ with **WF** will be quite small provided the number of PCs in the series, *r*, is large enough for the singular value λ_{r+1} to be negligible relative to λ_1. For this reason, the expansion of **X** into its PCs is the natural choice for a linear filter intended to reduce the rank.

On the other hand, using a few PCs for **X** many not ensure that the information preserved is useful for the estimation addressed in this work, namely kinetic multicomponent determinations, which usually involve one or two components with very large singular values—the rest having small singular values—that may be the most discriminating; therefore, in order to ensure that no useful information for the estimation is deleted, one should only discard those PCs associated with random noise or rounding errors.

If the data matrix is to be prefiltered, the ideal choice would be a filter requiring no prior knowledge of the specific data considered; however, this requirement is almost impossible to meet because matrix **W** is defined in terms of the data and only for linear networks and matrices with full-rank weights have attraction domains for certain error functions started to be developed.[21] At best, one can expect to use classes of data sets in a given context such as the multicomponent kinetic determination addressed in this work and be able to develop a heuristic filter construction methodology.

The choice of the PCs that are necessary and sufficient to discriminate, recognize, or estimate parameters in our intended application will largely depend on the variance of the random error associated with the experiment; therefore, it is preferable to conduct a preliminary heuristic study in order to obtain support for generalizable theoretical or methodological results since, in the absence of a normality hypothesis on the distribution of the population characteristics, performing hypothesis tests associated with the likelihood ratio on the eigenvalues of the correlation matrix for the population would be meaningless.[22] On the other hand, the heuristic approaches to the determination of the threshold where eigenvalues are associated with error components reported by Jollife,[23] Catell,[24] and Horn[25] lead to rather variable numbers of PCs to be considered that are usually smaller than the number used in this work. Consequently, we used the experimental results obtained to define a threshold *u* in order to automatically determine the number of PCs as follows: let $\lambda_1, \lambda_2, \ldots, \lambda_p$ be *p* real different eigenvalues arranged in a decreasing sequence for the sample matrix covariance $\mathbf{XX}^T$. The algorithm proposed for this purpose is

For $i = 1$ **to** p

If $\lambda_i > 1$, **then** the i-th PC is included since it is assumed to be associated to a characteristic with a variance exceeding that for any standardized characteristic of the network input data under the hypothesis of characteristic normality.

Else

If $(\lambda_i/\lambda_{i+1}) > (1 + \mathbf{u})$, where $0 < \mathbf{u} < 1$, **then** the i-th PC is included because it is considered to be associated with a characteristic with a small, but still much higher variance than that for the next orthogonal direction to be considered [that for the $(i+1)$-th PC].

Else, the i-th PC is not included because it is considered to be associated with rounding or random errors.

EndIf

EndIf

Depending on the particular problem addressed, the proposed algorithm must make a decision in regards to the threshold value *u* in order that the selection of the PCs to be used can be automated. As shown below, we used $u = 0.1$. In summary, the algorithm automatically calculates the number of network inputs (*i.e.*, the number of PCs to be used). For greater reliability, the number of inputs used in this work was confirmed heuristically from the experimental results obtained in the kinetic multidetermination developed.

EXPERIMENTAL SECTION

An overall 40 synthetic samples containing uniformly distributed concentrations of the analytes (cortisone and hydrocortisone) over the range 2−20 μg/mL were prepared as described in a previous paper.[26] Data (absorbance values

Table 1. Determination of the Optimum Sampling Interval[a]

	domain time, s		results	
domain	initial	final	%SET A	%SET B
I	0.2	120	11.29	11.21
II	0.2	96	9.15	9.01
III	0.2	72	12.40	16.90
IV	0.2	48	14.92	17.74
V	0.2	24	15.21	16.48
VI	12.2	108	11.25	10.87
VII	12.2	96	7.78	8.25
VIII	12.2	60	12.45	10.45
IX	12.2	36	15.25	14.25
X	24.2	120	15.87	16.48
XI	24.2	96	14.45	19.48
XII	24.2	72	16.47	10.56
XIII	24.2	48	18.25	21.17
XIV	36.2	108	12.45	10.48
XV	36.2	96	18.45	14.12
XVI	36.2	60	15.47	21.45

[a] A, cortisone; B, hydrocortisone.

at different times) were acquired at a frequency of 5 points/s over an interval of 120 s (600 points/curve), where the function exhibited and increasing trend.

RESULTS AND DISCUSSION

Developing an ANN methodology for kinetic multicomponent determinations of nonlinear (synergistic) systems involves several steps. The first is a topological study of the kinetic curve to be used to select the optimal time region that provides the maximum possible nonredundant information in the minimum reaction time. This is done by using those PCs that account for over 99% of the total variance as network inputs. After the optimal time region has been determined, the architecture of the neural network model is optimized by selecting the most suitable number of input data according to the methodology proposed here, that of neurons in the hidden layer and the transfer function leading to the maximum generalization capacity (*viz.*, the ability to correctly resolve mixtures of species not used at the training stage). As stated in the Experimental Section and considering the sample size involved, we chose to use the leave-one-out method[28] to estimate the generalization error; the method trains the network with $n - 1$ standards—where n is the number of available standards—and leaves 1 to test the generalization error. The process is repeated for the n standards, and the generalization error is calculated as the mean of the errors obtained for the different standards. Finally, sigmoidal, *s*, and linear, *l*, functions were used in the hidden and output layers, respectively, and the most suitable scaling ranges were (−1.0, 1.0) for input data and (0.2, 0.8) for output data.

The chemical system used to validate the proposed ANN model was the resolution of mixtures of two corticosteroids (cortisone and hydrocortisone) by their classical reaction with Tetrazolium Blue as implemented by using the CAR technique,[26] which increased the complexity of the system owing to its second-order nature; this resulted in high nonlinearity between the analytical signal and the concentrations of the mixture components.[27] In addition, both corticosteroids exhibit a very similar kinetic behavior and interact with each other as the reaction develops (*i.e.*, they exhibit a mutual synergistic effect). All this resulted in highly complex differential equations for the overall kinetic behavior of the system and excluded the use of parametric statistical methods for solving problem.

Table 2. Eigenvalues Associated with Each PC and Corresponding Value for the λ_i/λ_{i+1} Ratio

no. of PCs (i)	eigenvalues (λ_i)	λ_i/λ_{i+1} ratio
1	418.9098	
		401.56
2	1.04320	
		168.53
3	0.00619	
		1.33
4	0.00466	
		1.13
5	0.00412	
		1.19
6	0.00344	
		1.13
7	0.00305	
		1.08
8	0.00283	
		1.06
9	0.00267	

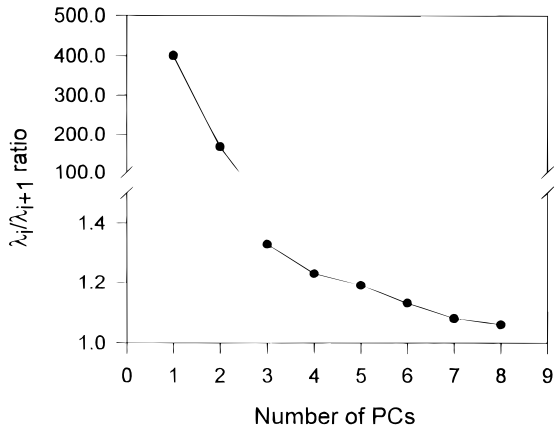

Figure 1. Plot of the λ_i/λ_{i+1} ratio against the number of PCs.

Selection of the Optimal Time Region. In the first step of the procedure, the time region of kinetic curve (absorbance *vs* time) providing the greatest amount of information for the intended kinetic multidetermination of species was selected. Table 1 shows the time domains used for this purpose—the maximum signal was obtained at 120 s, and each domain corresponds to a time portion of the kinetic curve. Data were subjected to PCA, and the 10 highest PCs were used as input to a 10:3s:2l network in each case. The results obtained are listed in Table 1. As can be seen, the best results in terms of the relative standard error of training (% SET) for the two mixture components were provided by domain VII (12.2−96 s). This interval corresponds to the linear portion of the kinetic curve and was used to obtain the measured parameter (maximum reaction rate) in the CAR technique,[27] which was employed here to obtain the experimental data.

We then determined the optimum sampling frequency; the network was found to perform similarly over the range from 5 to 0.1 point/s. This confirms the hypothesis that, in our case, the key factor was the time domain rather than the number of points or sampling frequency in such a domain.

Selection of the Optimum Network Design. After the optimum time domain and sampling frequency were determined, the significance of each PC selected from the data preprocessing procedure proposed in this paper was studied in order to select the most suitable as network input. Table 2 gives the eigenvalues obtained for the linear filter applied to the input matrix as well as their ratios; Figure 1 shows a plot of such PCs against the ratio of their associated

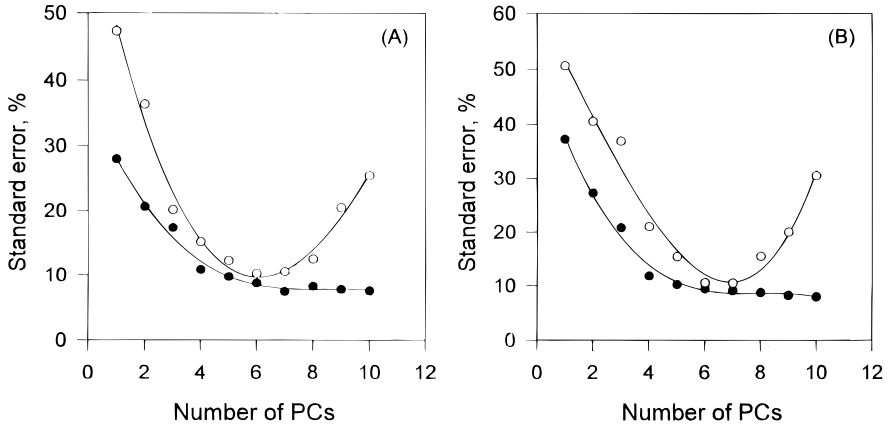

Figure 2. Determination of the optimum number of network input data. (●) Training and (○) prediction sets; (A) cortisone and (B) hydrocortisone.

Table 3. Determination of the Optimum Number of Hidden Neurons[a]

network architecture	%SET A	%SEP A	%SET B	%SEP B
6:2s:2l	10.51	19.47	12.45	18.25
6:3s:2l	8.74	10.21	9.45	10.54
6:4s:2l	3.47	5.12	4.12	5.25
6:5s:2l	2.47	5.45	2.87	6.85
6:6s:2l	7.85	10.54	7.97	12.57

[a] A, cortisone; B, hydrocortisone.

eigenvalues. As can be seen in Table 2, for a threshold $u = 0.1$, the application of the algorithm provides a number of PCs to be input to the network of 6.

In order to assess the performance of the proposed methodology, an empirical study was carried out by using an *p*:3s:2l architecture, where *p* is the number of PCs to be fed to the network. As can be seen in Figure 2, network performance improved gradually as further input data was introduced for both trainig and prediction sets. However, beyond the sixth node in the input layer, the generalization capacity of the network started to decay: the % SEP (relative standard error of prediction) gradually increases. This reflected in overdimensioning of the model (*i.e.*, the network contained more weights than needed for the desired estimation). As can be seen, the empirical results were consistent with those provided by the proposed methodology.

In order to determine the optimum number of hidden neurons, we used a 6:*q*s:2l structure with the first seven PCs as the input nodes and calculated the relative standard errors of training (% SET) and prediction (% SEP) as a function of the number of hidden neurons, *q*. As can be seen in Table 3, the best results were obtained with networks consisting of four or five neurons in the hidden layer; a 6:4s:2l network was finally chosen because, although the error of prediction was similar to that for a 6:5s:2l network, the former required 10 fewer connections and hence a smaller number of standards to train the network—this is highly desirable when a methodology of this type is to be used to solve any practical problem.

Validation of the Proposed Methodology. In order to evaluate the performance of the proposed neural network methodology for kinetic multicomponent determinations we used the cross-validation procedure known as the "leave-one-out method". Figure 3 shows the estimated value *vs* real value plot for both components and the regression parameters for the straight lines obtained. As can be seen, the network predictive ability was very good, on account of the regression parameters reflected in this figure. In addition, the % SEP for cortisone and hydrocortisone (5.12% and 5.25%, respectively) were quite good taking into account the difficulties involved in modeling the behavior of highly nonlinear kinetic systems such as that were dealt with in this work.

CONCLUSIONS

The proposed methodology is very useful for processing highly nonlinear data by using principal components (PCs) as input data for an artificial neural network (ANN). The

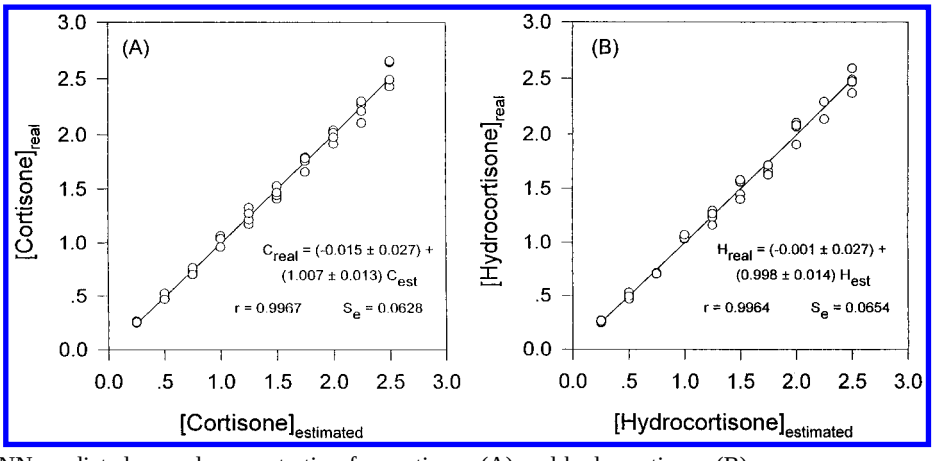

Figure 3. Plot of ANN-predicted *vs* real concentration for cortisone (A) and hydrocortisone (B).

key to its success lies in automatic selection of the eigenvalues of the correlation matrix associated with the first few PCs in such a way that the selected subset leads to an optimal generalization value. The methodology was validated on such a highly nonlinear system as the resolution of a mixture subjected to a synergistic effect by using the CAR technique, which increases the nonlinearity of the kinetic curve owing to its second-order nature.

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) Wentzell, P. D.; Karayannis, M. L.; Crouch, S. R. Simultaneous kinetic determination with the Kalman filter. *Anal. Chim. Acta* **1989**, *224*, 263−274.
(2) Xiong, R.; Velasco, A.; Silva, M.; Pérez-Bendito, D. Performance of the Kalman filter algorithm in differential reaction-rate methods. *Anal. Chim. Acta* **1991**, *251*, 313−319.
(3) Forster, E.; Silva, M.; Otto, M.; Pérez-Bendito, D. Enzymatic determination of alcohol mixtures at the nanogram level by the stopped-flow technique. *Anal. Chim. Acta* **1993**, *274*, 109−116.
(4) Quencer, B. M.; Crouch, S. R. Multicomponent kinetic determination of lanthanides with stopped-flow, diode array spectrophotometry and the extended Kalman filter. *Anal. Chem.* **1994**, *66*, 458−463.
(5) Ríos, A.; Silva, M.; Valcárcel, M. Fluorimetric determination of ammonia, hydrazine and hydroxylamine and their mixtures by differential kinetic methods. *Fresenius J. Anal. Chem.* **1985**, *320*, 762−768.
(6) Bhat, McAvoy, T. J. Use of neural nets for dynamic modeling and control of chemical process systems. *Comput. Chem. Eng.* **1990**, *14*, 573.
(7) Bos, M.; Bos, A.; van der Linden, W. E. Artificial neural networks as a tool for soft-modeling in quantitative analytical chemistry: the prediction of the water content of cheese. *Anal. Chim. Acta* **1992**, *256*, 133−144.
(8) Ball, J. W.; Jurs, P. C. Automated selection of regression models using neural networks for ^{13}C NMR spectral predictions. *Anal. Chem.* **1993**, *65*, 505−512.
(9) Wessel, M. D.; Jurs, P. C. Prediction of reduced ion mobility constants from structural information using multiple linear regression analysis and computational neural networks. *Anal. Chem.* **1994**, *66*, 2480−2487.
(10) Long, J.; Gregoriu, V. G.; Gemperline, P. J. Spectroscopic calibration and quantitation using artificial neural networks. *Anal. Chem.* **1990**, *62*, 1791−1797.
(11) Bos, M; Bos, A.; van der Linden, W. E. Artificial neural networks as a multivariate calibration tool: modeling the Fe-Cr-Ni system in x-ray fluorescence spectroscopy. *Anal. Chim. Acta* **1993**, *277*, 289−295.
(12) Blank, T. B.; Brown, S. D. Data Processing using neural networks. *Anal. Chim. Acta* **1993**, *277*, 273−287.
(13) Ventura, S.; Silva, M.; Pérez Bendito, D.; Hervás, C. Artificial neural network for estimation of kinetic analytical parameters. *Anal. Chem.* **1995**, *67*, 1521−1525.
(14) Ventura, S.; Silva, M.; Pérez Bendito, D.; Hervás, C. Multicomponent kinetic determinations using artificial neural networks. *Anal. Chem.* **1995**, *67*, 4458−4461.
(15) Blanco, M.; Coello, J.; Iturriaga, H.; Maspoch, S.; Redón. M. Artificial neural networks for multicomponent kinetic determinations. *Anal. Chem.* **1995**, *67*, 4477−4483.
(16) Bos, M; Bos, A.; van der Linden, W. E. Processing of signals from an ion-selective electrode array by a neural network. *Anal. Chim. Acta* **1990**, *233*, 31−39.
(17) Gemperline, P. J.; Long, J.; Gregoriu, V. G. Nonlinear multivariate calibration using principal components regression and artificial neural networks. *Anal. Chem.* **1991**, *63*, 2313−2323.
(18) Hervás, C.; Garrido, A.; Lucena, B.; García, N.; de Pedro, E. Near infrared spectroscopy for classification of Iberian pigs carcasses using an artificial neural network. *J. Infrared Spectrosc.* **1994**, *2*, 177−184.
(19) Ellacott, S. W. *The numerical analysis approach. Mathematical approaches to neural networks*; Elsevier: 1993.
(20) Jacobs, D. *The state of art in Numerical Analysis*; Academic Press: New York, 1977.
(21) Plumbley, M. D. Lyapunov functions for convergence of principal component algorithms. *Neural Networks* **1995**, *8*, 11−23.
(22) Anderson, T. W. *An introduction to multivariate statistical analysis*; John Wiley & Sons: 1984.
(23) Jolliffe, I. T. Discarding variables in a principal component analysis, I: Artificial data. *Applied Statistics* **1972**, *21*, 160−173.
(24) Cattell, R. B. The scree test for the number of factors. *Multivariate Behavioral Res.* **1966**, *1*, 245−276.
(25) Horn, J. L. A rationale and test for the number of factors in factor analysis. *Psychometrika* **1965**, *30*, 179.
(26) Ayllón, L; Silva, M.; Pérez Bendito, D. Improved automatic kinetic method for the determination of various corticosteroids. *J. Pharm. Sci.* **1994**, *83*, 1135−1141.
(27) Velasco, A; Silva, M.; Pérez Bendito, D. Processing analytical data obtained form second-order reactions using continuous reagent addition. *Anal. Chem.* **1992**, *64*, 2359−2365.
(28) Raudys, S. J.; Jain, A. K. Small sample size effects in statistical pattern recognition: recommendations for practitioners. *IEEE Trans. Pattern Analysis Machine Intelligence* **1991**, *13*, 252−264.

CI960084O