Available online at www.sciencedirect.com
ScienceDirect
Chemometrics and Intelligent Laboratory Systems 85 (2007) 131–139
Chemometrics and intelligent laboratory systems
www.elsevier.com/locate/chemolab

A chemometric strategy based on peak parameters to resolve overlapped electrochemical signals

J.M. Palacios-Santander *, L.M. Cubillana-Aguilera, I. Naranjo-Rodríguez, J.L. Hidalgo-Hidalgo-de-Cisneros

Departamento de Química Analítica, Universidad de Cádiz, Polígono Río San Pedro, Apartado 40, 11510, Puerto Real, Cádiz, Spain

Received 11 March 2005; received in revised form 10 March 2006; accepted 30 May 2006
Available online 17 July 2006

Abstract

A new chemometric methodology based on the use of peak parameters as direct input data into different multivariate calibration methods is proposed. Different regression techniques such as multilinear regression (MLR), partial least square regression (PLS), principal component regression (PCR) and artificial neural networks (ANN), were utilized in order to resolve hard overlapped electrochemical signals belonging to the well-known Tl^{+}/Pb^{2+} system, which was used as a benchmark. This strategy was studied as an alternative to traditional procedures that apply pre-treatment techniques (dimension reduction methods or feature selection processes) to the full voltammograms of the signals. Good predictive and effective models were obtained, being the RMS errors very similar in all cases, independent of the calibration method. However, ANN-based regression models performed slightly better. The average relative errors ranged from 5 to 10% for Tl^{+} and from 4 to 12% for Pb^{2+}. A study of the relevance of the voltammetric peak parameters was also carried out. This parameters-based strategy can involve a fast and efficient alternative to resolve multicomponent systems in those analytical techniques whose signals can be represented by peak parameters.

Keywords: Artificial neural networks; Multivariate calibration; Electrochemical peak parameters; Overlapped electrochemical signals; Voltammetry

1. Introduction

One of the main limitations to the application of electroanalytical techniques in the field of quantitative analysis is often due to lack of selectivity. In fact, it often happens that different species undergo oxidation or reduction at potential values that are very close to each other giving place to serious overlapping. Besides separation techniques, complexometric methods, or experimental manipulations like changes of pH, of the supporting electrolyte, or the use of modified electrodes, chemometrics offers efficient alternatives to solve the problem of overlapping signals.

Amongst the most used chemometric techniques for simultaneous evaluation of overlapped signals, independent of the type of signal, we can find deconvolution or semidifferential techniques coupled to curve fitting [1,2], multivariate curve resolution [3,4], and multivariate calibration [5–7]. Specifically in the field of electrochemistry, many successful applications of multivariate calibration based on different regression methods have been recently reported: multilinear regression (MLR) [8], principal component regression (PCR) [9,10], continuum regression [11], partial least squares regression (PLS) [9,10,12,13] and artificial neural networks (ANNs) [14,15].

In this paper we propose a new chemometric methodology based on the application of several regression methods such as MLR, PLS, PCR and ANNs, with the aim of resolving hard overlapped electrochemical signals, using different peak parameters as input data: position, height, half width, derivative and area of the voltammetric peaks. The well-known Tl^{+} and Pb^{2+} system, with the ions in the concentration range from 0.1 to 1.0 mg l^{-1}, which has already been resolved by the traditional PLS and PCR methods (but with higher concentrations of the ions), was employed as a benchmark [11]. The standard addition method was used to obtain the voltammetric signals for individual analytes and mixtures; no problem associated to this procedure

* Corresponding author. Tel.: +34 956 01 63 61; fax: +34 956 01 64 60.
E-mail address: josem.palacios@uca.es (J.M. Palacios-Santander).

0169-7439/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2006.05.015

was found, due to no shift of the peak potentials was observed for our standards, as it has been described in bibliography [16,17].

Until now, the multivariate calibration techniques used to resolve overlapped electrochemical signals have been applied on full voltammograms, *i.e.*, all values of intensity [18] or intervals of them [15,16] were employed as input data. To minimise the number of input variables, we recently proposed a strategy based on Fourier Transform (FT) and Wavelet Transform (WT) coupled to different regression methods (PCR, PLS and ANNs) [19] to resolve a seriously overlapped Tl^{+}/Pb^{2+} electrochemical system. Furthermore, we completed our studies applying a WT-based feature selection procedure on the same data sets [20] as a previous step to the multivariate calibration with PLS, MLR and ANNs.

The use of the information contained in the voltammetric peak parameters has not been very extended in literature, apart from being used to optimise the determination procedures of analytes when applying voltammetric techniques [21,22]. Some other interesting applications of peak parameters (not only in voltammetric techniques) are summarised as follows: diagnostic criteria in kinetics of surface redox processes [23]; characterization of first-order EC reactions [24]; mathematical modelling of curves to estimate peak parameters in capillary electrophoresis [25] and infrared spectroscopy [26], in order to predict optimal separation conditions and improve the interpretation of multivariate and rule induction classification models, respectively; and determination of the peak shape parameters by means of WT [27,28].

However, in the field of the electrochemical quantification, a lesser number of papers using voltammetric peak parameters can be found in the scientific literature: one of them [29] compares critically the use of peak area and peak current to determine the concentration of different metallic species in solution, while in other one [30], the authors apply ANNs based only on one parameter: peak intensity, with the aim of determining simultaneously Cd^{2+} and Pb^{2+} in mixtures measured by flow injection analysis using multiple differential pulse voltammetric detection. Actually, the ANN model obtained in the last case had the same characteristics as a MLR model, since a perceptron with a 2–2–2 topology (two neurons in the input layer: peak intensities of the two ions, two neurons in the hidden layer and two neurons in the output layer: concentrations of the two ions) with linear transfer functions was employed.

In the present work, our idea goes further: we use the five previously defined voltammetric parameters as input data, establishing a comparison amongst different chemometric techniques (MLR, PLS, PCR and ANN), in order to obtain stable and simple multivariate calibration models. The performance of each calibration model has been tested by the root-mean square error (RMS), estimated on two different test sets for all cases. This decision variable is defined by next equation:

$$\mathrm{RMS} = \sqrt{\frac{\sum_{i}(y_i - y_i^*)^2}{n}}$$

where y_i represent the measured concentration for the ions; y_i^* represent the predicted concentration for the ions; and n is the number of samples.

The innovation of this piece of research radicates in the following aspects:

- It constitutes one of the first applications of peak parameters to the direct multivariate calibration of electrochemical signals.
- No data pre-treatment is necessary, what offers certain advantages versus other chemometric strategies [19,20]: the peak parameters are directly utilized as input data by the calibration techniques, thus the chemical interpretation of the results is faster and more reliable because these parameters can be easily related to the analytes concentrations.
- The number of input data is quite low, and consequently the number of adjustable parameters as well. That is why the structure of the models obtained is very simple, avoiding the handicap of overfitting in a good way. In bibliography, it is possible to find many applications of neural networks using very simple architectures, *i.e.*, only a few variables as input data: 1–3 variables [31], 3 variables [32] and 3–6 variables [33].
- Other voltammetric peak parameter, the derivative, is used to obtain good models of prediction together with peak area and intensity, the most commonly used parameters to relate with concentration values [29,30].
- This peak parameters-based strategy can involve a fast and efficient alternative to resolve multicomponent systems in voltammetry or even in other analytical techniques such as chromatography or spectroscopy, whose signals can be represented by peak parameters.

2. Experimental

2.1. Methodology

As it has been explained yet, voltammetric peak parameters have been used here as direct inputs in different chemometric techniques in order to resolve overlapped electrochemical signals. These peak parameters can be defined as follows:

- Position (V): potential at which the current with respect to the baseline has a maximum (volts — V); also called peak potential.
- Height (I): maximum current (intensity) with respect to the baseline (ampere — A); also called peak intensity
- Peak area (S): area of the peak corrected for the baseline (arbitrary units).
- Half width (W): the difference between the peak potential, E_p, and the potential at half height, $E_{p/2}$ (volts — V).
- Derivative (D): the sum of the absolute values of the maximum and the minimum in the first derivative of the peak (arbitrary units) [34].

The values of the voltammetric peak parameters: V, I, S, W, and D, as well as the concentration of each ion used in all the mixtures studied in this work, appear in Table 1.

The voltammetric parameters have been set up in order to obtain the best sensitivity, i.e., the highest peak intensity. These instrumental conditions may have caused an increase in the half-

Table 1
Values of the voltammetric peak parameters used in the present work

Sample	*V*	*I*	*S*	*W*	*D*	$[Tl^+]$	$[Pb^{2+}]$	Sample	*V*	*I*	*S*	*W*	*D*	$[Tl^+]$	$[Pb^{2+}]$
T1	−0.494	22.937	0.276	0.111	5.742	0.10	0.00	T3L3	−0.506	189.367	2.133	0.106	59.873	0.29	0.29
T2	−0.494	51.603	0.626	0.111	12.917	0.20	0.00	T8L3	−0.499	342.900	3.992	0.111	97.627	0.77	0.29
T3	−0.497	79.307	0.966	0.114	19.887	0.30	0.00	T4L4	−0.509	247.767	2.775	0.106	79.240	0.39	0.39
T4	−0.487	107.467	1.282	0.111	26.923	0.39	0.00	T9L4	−0.504	396.667	4.593	0.111	114.767	0.86	0.38
T5	−0.494	139.000	1.680	0.116	34.340	0.49	0.00	T5L5	−0.504	329.067	3.706	0.106	103.567	0.48	0.48
T6	−0.491	171.633	2.075	0.114	42.397	0.59	0.00	T10L5	−0.504	474.833	5.473	0.111	139.633	0.94	0.47
T7	−0.492	200.600	2.424	0.113	49.963	0.68	0.00	T1L6	−0.522	235.100	2.458	0.101	85.880	0.10	0.58
T8	−0.486	231.800	2.790	0.111	56.363	0.78	0.00	T6L6	−0.504	383.900	4.329	0.106	120.133	0.57	0.57
T9	−0.489	254.933	3.065	0.111	62.943	0.87	0.00	T2L7	−0.515	299.000	3.184	0.101	105.333	0.19	0.68
T10	−0.492	291.333	3.551	0.116	73.103	0.96	0.00	T7L7	−0.504	459.500	5.153	0.106	146.300	0.66	0.66
L1	−0.525	40.973	0.410	0.096	15.640	0.00	0.10	T3L8	−0.515	362.867	3.868	0.101	128.533	0.29	0.77
L2	−0.523	74.553	0.750	0.096	28.497	0.00	0.20	T8L8	−0.506	526.067	5.691	0.101	156.000	0.75	0.75
L3	−0.522	106.233	1.064	0.096	40.463	0.00	0.30	T4L9	−0.509	426.500	4.600	0.106	148.233	0.38	0.86
L4	−0.525	139.500	1.398	0.096	53.203	0.00	0.39	T9L9	−0.504	581.067	6.526	0.106	183.733	0.84	0.84
L5	−0.520	178.433	1.790	0.098	68.223	0.00	0.49	T5L10	−0.509	489.467	5.321	0.106	167.533	0.47	0.94
L6	−0.520	209.333	2.102	0.096	80.167	0.00	0.59	T10L10	−0.501	659.567	6.974	0.101	192.367	0.93	0.93
L7	−0.520	243.100	2.441	0.096	93.090	0.00	0.68	T1L9	−0.520	328.900	3.426	0.101	119.300	0.10	0.87
L8	−0.523	280.833	2.820	0.098	107.267	0.00	0.78	T9L1	−0.499	287.900	3.414	0.111	77.050	0.87	0.10
L9	−0.520	311.833	3.125	0.096	119.500	0.00	0.87	T3L5	−0.504	255.300	2.767	0.101	85.670	0.29	0.48
L10	−0.520	341.233	3.423	0.096	131.067	0.00	0.96	T5L3	−0.504	244.200	2.786	0.111	73.120	0.48	0.29
T1L1	−0.509	65.520	0.736	0.106	20.667	0.10	0.10	T6L8	−0.509	443.000	4.871	0.106	143.800	0.57	0.76
T6L1	−0.496	215.300	2.534	0.111	54.903	0.58	0.10	T8L6	−0.499	422.000	4.655	0.101	123.700	0.76	0.57
T2L2	−0.506	127.300	1.427	0.106	40.110	0.20	0.20	T10L2	−0.499	343.600	4.058	0.111	94.070	0.95	0.19
T7L2	−0.502	274.900	3.225	0.111	75.937	0.68	0.19	T2L10	−0.520	374.000	3.906	0.101	133.700	0.19	0.95

V=peak potential (V); *I*=peak intensity (nA); *S*=area of the peaks ($\times 10^{-8}$ a.u.); *W*=half width (V); *D*=derivative ($\times 10^{-7}$ a.u.); T=Tl^+; L=Pb^{2+}; 1, 2, …, 10=0.1, 0.2, …, 1.0 mg l^{-1}.

width values (*W*). Electrochemical parameters such as pulse amplitude improve the sensitivity, but affect the width of the voltammetric peaks [35].

The chemometric techniques applied in this paper consisted of several multivariate calibration methods, such as MLR, PLS, PCR and ANNs. When applying these techniques, whose theoretical aspects can be found in [36–38], all possible combinations of the five parameters were employed as input data, in order to build the multivariate calibration models. Table 2 represents all the models studied, excepting three of them: *V*, *W* and *VW*, because the variability of the constitutive peak parameters was not high enough to discriminate the signals and, consequently to resolve the mixtures. For MLR and PLS, the final number of models was 25. Nevertheless, with respect to ANNs, the number of models increased notoriously as a consequence of all possible combinations amongst the transfer functions for every topology tested.

2.1.1. Multilinear regression (MLR)

The voltammetric peak parameters (independent variables) were mean-centred. The performance of each MLR model was tested by the RMS_{mon} error, estimated on a monitoring set of 9 mixtures in order to establish comparison with ANNs. For each combination of peak parameters, the best performing MLR models were selected and their predictive ability was further checked by an external validation set (RMS_{tst}) consisting of 8 mixtures.

2.1.2. Partial least squares regression (PLS) and principal component regression (PCR)

The voltammetric parameters were also mean-centred. The optimal number of PLS components has been chosen by crossvalidation. The number of significant components (latent variables, LVs) chosen in every model was as follows: models based on 5 parameters=from 3 to 5 LVs; models based on 4 parameters=3 and 4 LVs; models based on 3 peak parameters=2 and 3 LVs; and finally, models based on 2 parameters=2 LVs. Models based on only 1 parameter were tested at the beginning, but they were not used later since the errors obtained were rather high.

The performance of each PLS model was also tested by the RMS_{mon} error. For each combination of peak parameters, the best performing MLR models were selected and their predictive ability was checked by the RMS_{tst} error.

The MLR, PLS and PCR models were calculated by means of the software Unscrambler® ver. 7.01.

2.1.3. Artificial neural networks (ANNs)

According to the models collected in Table 2, different neural network topologies were tested: n_i–2–2, where n_i is the number

Table 2
Composition of the different multivariate calibration models tested

Number of parameters	Models*
5	*VISWD*
4	*VISW VISD VIWD VSWD ISWD*
3	*VIS VIW VID VSW VSD VWD ISW ISD IWD SWD*
2	*VI VS VD IW IS ID SW SD WD*
1	*I S D*

V=peak potential (V); *I*=peak intensity (nA); *S*=area of the peaks ($\times 10^{-8}$ a.u.); *W*=half width (V); *D*=derivative ($\times 10^{-7}$ a.u.). (*) The *VW*, *V* and *W* models were not tested because the variability of these peak parameters, and consequently the information contained in them, were not significant.

of voltammetric peak parameters (from 2 to 5). The number of hidden neurons was 2 with the aim of: 1) comparing the results with those obtained after applying a WT-based feature selection procedure as a previous step to the multivariate calibration process [20]; and 2) minimising the number of adjustable parameters, which can be calculated by the formula [38,39]: $N = (\text{input nodes} * \text{hidden nodes}) + (\text{hidden nodes} * \text{output nodes}) + \text{hidden nodes} + \text{output nodes}$. As the maximum number of input neurons can only be 5 (the biggest model, see Table 2), the maximum number of adjustable parameters would be 18. One limitation of adjusting hidden neurons is that the ratio of training set observations (31) to adjustable parameters (18 maximum) must be greater than two to reduce the risk of chance correlations and network overtraining. In our case, the maximum N value is 1.72 (31/18), what borders overfitting. However, this situation only would occur in one model, having the rest of models between 12 and 16 adjustable parameters. Thus, in most cases there is no risk of overfitting.

In this paper, we use a feedforward-type (connections must connect to the next layer) and multilayered neural networks with an improved faster back propagation (BP) algorithm (to know more about how the BP training is accomplished, please see Ref. [19]). The activation or transfer functions used here for each neuron were: linear for the input layer and all possible combinations (9) of gaussian, sigmoid and hyperbolic tangent functions for the hidden and output layers.

The program Qnet® 2000 has been used for the ANN calculations. The training process of the neural network model was stopped minimising the RMS_{mon} error, estimated on a test set of 9 mixtures. Since the starting weights are randomly generated for each set of coefficients, five ANN runs were made and the resulting RMS errors were averaged [38,39]. Taking into account all the previous considerations, the final number of neural models studied was 1175 (25 models H 9 combinations of transfer functions H 5 replicates).

For each combination of peak parameters, the best performing ANN models corresponding to low RMS_{mon} errors were selected. However, it is worth noticing that in the case of ANN this test set is used to stop the training of the network and thus it does not represent a true validation set, being rather a monitoring set. Accordingly, the predictive ability of the chosen ANN models was checked by an external validation set (RMS_{tst} error) of 8 mixtures.

2.2. Instrumentation, reagents and materials

The DPASV (Differential Pulse Anodic Stripping Voltammetry) measurements were performed with an Autolab®/PGSTAT20 (Ecochemie, Utrecht, The Netherlands) potentiostat/galvanostat, interfaced with a personal computer, and coupled to a VA 663 Stand (Metrohm, Herisau, Switzerland). The AutoLab software GPES (General Purpose Electrochemical System) was used for waveform generation and data acquisition and elaboration.

An electrochemical three electrode cell, with a platinum auxiliary electrode, a silver/silver chloride, 3 M potassium chloride reference electrode and a HMDE (Hanging Mercury Drop Electrode), was employed. All this equipment was also purchased from Metrohm.

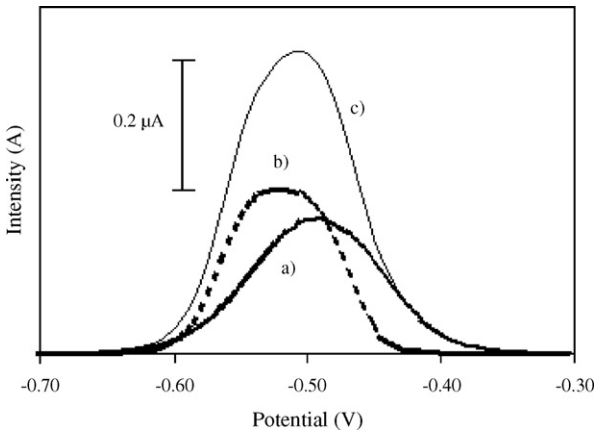

Fig. 1. Superposition of voltammograms: a) 0.7 mg l^{-1} of Tl^{+} (T7); b) 0.7 mg l^{-1} of Pb^{2+} (L7); c) mixture of 0.7 mg l^{-1} of Tl^{+} and 0.7 mg l^{-1} of Pb^{2+} (T7L7).

Analytical reagent grade chemicals were used throughout the experiments. Voltammograms were recorded at room temperature (25 ± 1 °C). All solutions were de-aerated with nitrogen when necessary for at least 10 min before performing the experiments.

A 2 M acetic acid (Panreac, Barcelona, Spain)/2 M ammonium acetate (Merck, Darmstad, Germany) buffer solution was utilized as supporting electrolyte (pH=4.8–5.0). Lead and thallium solutions were prepared from nitrate salt (Merck); stock solutions had a concentration of 250 mg l^{-1}.

The instrumental parameters used in DPASV measurements were as follows: deposition potential=−1.3 V; deposition time=120 s; rest period=20 s; initial potential=−1.3 V; end potential=0.0 V; scan rate=8.5 mV s^{-1}; pulse amplitude=0.10 V; pulse time=0.07 s; pulse repetition time=0.6 s. The drop surface was approximately 0.52 mm^2.

2.3. Software

For the statistical treatment the following software packages were used: Unscrambler® 7.01 and EXCEL® 97 Pro. Qnet® 2000, neural network software, was utilized to obtain the neural models.

2.4. Sampling

40 hardly overlapped mixtures (see Table 1) of thallium and lead, at the concentration range from 0.1 to 1.0 mg l^{-1}, were experimentally determined. Fig. 1 represents the overlapping voltammograms of the two ions and one of their mixtures. The potential peaks (E_p) of each individual analyte are situated in a few mV, which implies a very severe grade of overlapping amongst the individual signals, as the unique peak in the voltammogram of the mixtures shows. Nine out of these mixtures were used as internal test set (mon, from monitoring set for ANN), namely T2, T9, L3, L8, T1L6, T4L4, T10L5, T6L1, and T9L9, where L indicates lead, T indicates Thallium, 1 corresponds to a concentration of 0.1 mg l^{-1}, 2 corresponds to

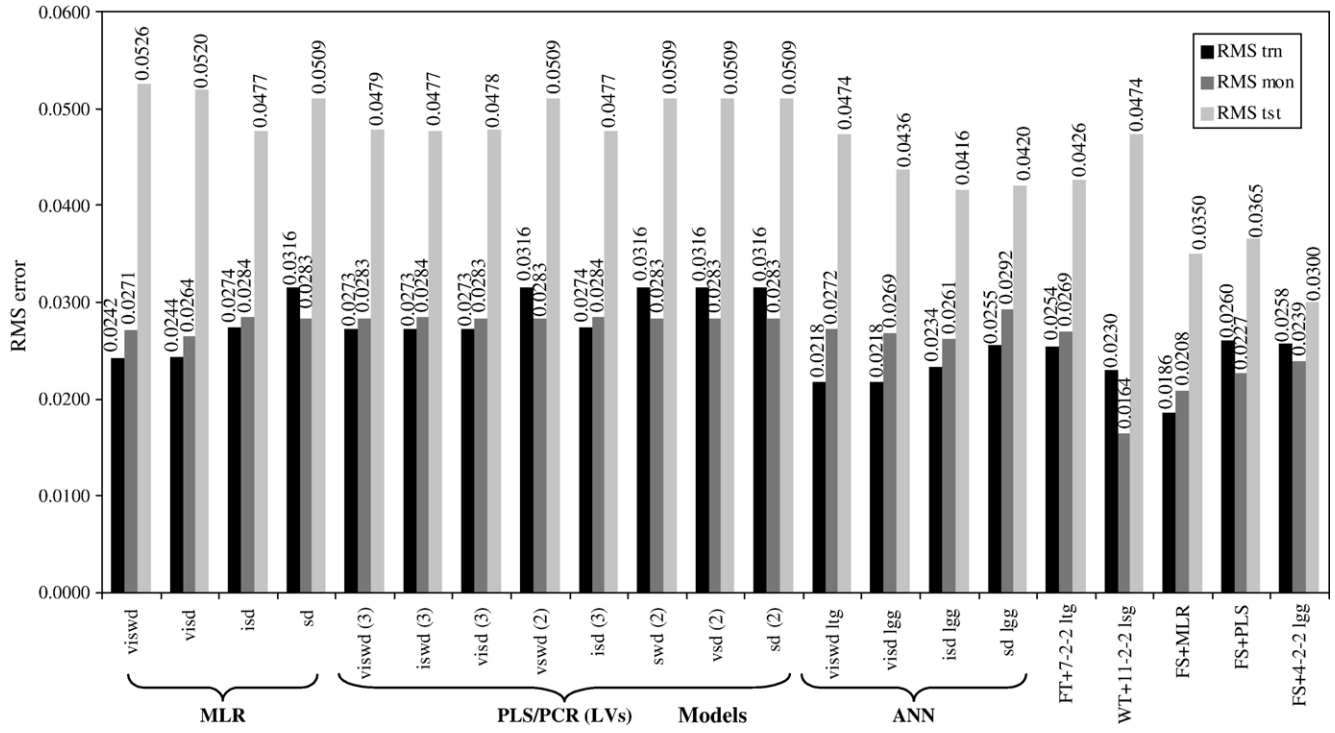

Fig. 2. RMS errors for the best multivariate calibration models based on voltammetric peak parameters. V=peak potential (V); I=peak intensity (nA); S= area of the peaks ($\times 10^{-8}$ a.u.); W=half width (V); D=derivative ($\times 10^{-7}$ a.u.); trn=training set; mon=monitoring set (first test set); tst=external test set (second test set); MLR=multilinear regression; PLS=partial least square regression; PCR=principal component analysis; ANN=artificial neural network; LVs=latent variables; FT and WT=Fourier and wavelet transform, respectively, as pre-treatment techniques[19]; X–Y–Z: neural network topologies with certain transfer functions (l=linear, g=gaussian, s=sigmoid and t=hyperbolic tangent); FS=WT-based feature selection procedure as a previous step to multivariate calibration[20].

0.2 mg l^{-1} and so on. After some time 8 additional mixtures were measured in order to obtain an external validation set (tst, test set), which correspond to T1L9, T2L10, T3L5, T5L3, T6L8, T8L6, T9L1, and T10L2, respectively. The remaining mixtures (31) were used as the training (trn) set with the aim of building the multivariate calibration models.

3. Results and discussion

In Fig. 2 different multivariate calibration models giving the best results (RMS errors) are represented. In next paragraphs, we will discuss these results for every regression technique.

3.1. Multilinear regression (MLR)

The four best MLR models appear collected in Fig. 2. As it can be seen, for models with a number of voltammetric peak parameters between three and five, the simpler the model (less number of parameters), the better the results obtained for the external test set (RMS_{tst} error); RMS_{mon} error is more or less constant and with respect to the RMS_{trn} error: the less the number of parameters, the bigger this value. However, when diminishing the structure of the models from three to two parameters, there is a characteristic change in the RMS_{tst} error trend due to the parameter cancelled now is peak intensity (I) instead of peak potential (V) or half width (W), as in the other models. As expected, peak intensity (I) is more relevant for the resolution of the system than V and W.

On one hand and according to the former ideas, a MLR multivariate calibration model requires the presence of S (peak area) and D (derivative) parameters, fundamentally. This means that S and D parameters contain relevant information to resolve the Tl^+/Pb^{2+} electrochemical system. S is directly related to I (R^2=0.99985 for Tl^+ and R^2=0.99996 for Pb^{2+}); thus, when including it in the regression models, the RMS errors decrease slightly for every set of samples (trn, mon and tst). D has something to do with the position of the maximum of the voltammetric peaks, which is definitively the keystone to resolve the system, since the difference between the maximum of the peaks for both cations is approximately 25–30 mV.

To understand why D is so important to the resolution of the overlapped signals it is necessary to remember that the derivative of a gaussian function (voltammetric signal) is a sigmoid function, which cuts the abscissa (potential) axis at the very same point

Table 3
Expressions of the calibration curves for I, S and D parameters versus the concentration of Tl^+ and Pb^{2+} ions

Ion	Peak intensity (I)	Peak area (S)	Peak derivative (D)
Tl (I)	I=(310.91±3.23) ×C+(−11.59±1.95) R^2=0.9991	S=(3.76±0.05) ×C+(−0.14±0.03) R^2=0.9985	D=(76.93±1.14) ×C+(−2.78±0.69) R^2=0.9982
Pb (II)	I=(353.02±2.38)×C+ (3.38±1.95) R^2=0.9996	S=(3.54±0.02) ×C+(0.03±0.01) R^2=0.9996	D=(135.52±0.92) ×C+(1.07±0.55) R^2=0.9996

I=peak intensity (nA); S=area of the peaks ($\times 10^{-8}$ a.u.); D=derivative ($\times 10^{-7}$ a.u.); C=concentration (mg l^{-1}).

where the maximum of the *Y* variable (intensity) is located. However, the quoted sigmoid function has other important oddity: the absolute value of the distance measured from the maximum to the minimum points delimited by the function is closely related to the intensity value ($R^2=0.9995$ for Tl^+ and $R^2=0.99998$ for Pb^{2+}) of the original gaussian function. Moreover, the information contained in *D* is more complete than that contained in the parameter *I*: for example, a MLR model based on ID gives higher errors ($RMS_{trn}=0.0663$; $RMS_{mon}=0.0564$; $RMS_{tst}=0.0733$) than a model based on SD ($RMS_{trn}=0.0316$; $RMS_{mon}=0.0283$; $RMS_{tst}=0.0509$). The parameter *V* is not possible to be related to the concentration, unlike *I*, *S* or *D*, which has a direct relationship with the concentrations of the ions [29] (see Table 3).

In order to confirm that intensity, area and derivative of the peaks are fundamental parameters, a principal component analysis (PCA) based on the voltammetric peak parameters of the signals was carried out. Fig. 3 represents the loadings plot corresponding to the first principal component for A) mean-centred and B) autoscaled peak parameters. As it can be seen, *D* and *I* have the highest values for the principal component in both cases; however, *S* only presents relevance when using autoscaled data, while with mean-centred data its relevance does not seem to be very important. The parameters *V* and *W* are clearly irrelevant.

If a model had to be selected, it would be appropriate to choose that one constituted by the parameters *ISD*, since it has the lowest RMS error values for all the data sets and is one of the simplest model at the same time.

3.2. Partial least square regression (PLS)

In the case of multivariate calibration models based on PLS, the ideas commented in the previous section are also valid here. As seen in Fig. 2, all the models with the same number of latent variables and differing exclusively in the presence or absence of *V* and/or *W* parameters give almost identical RMS error values.

Globally, the results are very similar to those obtained with MLR, mainly when the number of latent variables in PLS models agrees the number of initial peak parameters in MLR: for example, *ISD* and *ISD* (3) or *SD* and *SD* (2). Taking this statement into consideration, we would be practically talking about a multilinear regression, although in a different latent variable domain, where all the information contained in the initial peak parameters, without reducing dimensions, would be in use.

The fundamental differences between MLR and PLS regression can be found when the number of latent variables in PLS differs from the number of parameters employed to build the initial MLR models. So, for instance, for the *VISWD* (3) PLS model, based on 3 LVs, the results are a bit better than when using its counterpart of MLR (*VISWD*). The explanation to this fact may be the next one: in the PLS model, the number of variables varies from 5 (initial) to 3 LVs, what implies a reduction of dimensions; during this process, the *V* and *W* parameters presumably constitute the most important percentage of information discarded and, thus, irrelevant. Besides, at the same time, the advantage of working in a new variable space with PLS is profited.

Comparing the results obtained with the two types of regression models, mixed effects can be observed as well: the influence of removing the *V* and/or *W* parameters in the models, as well as the number of principal components. In this way, the MLR model *ISD* gives the same results as the PLS models *ISWD* (3), *VISD* (3) and *ISD* (3). The same occurs amongst the MLR model SD and the PLS models *VSWD* (2), *SWD* (2), *VSD* (2) and *SD* (2).

In order to select one PLS model as the optimal, it was decided to choose the same model as in the MLR section, since both were identical: *ISD* ≡ *ISD* (3).

3.3. Artificial neural networks (ANNs)

The results corresponding to the best ANN models are also collected in Fig. 2. As can be observed, ANN models show slightly lower RMS error values for the three sets of samples (training, monitoring and test sets) than MLR and PLS models, although the differences found with respect to the others

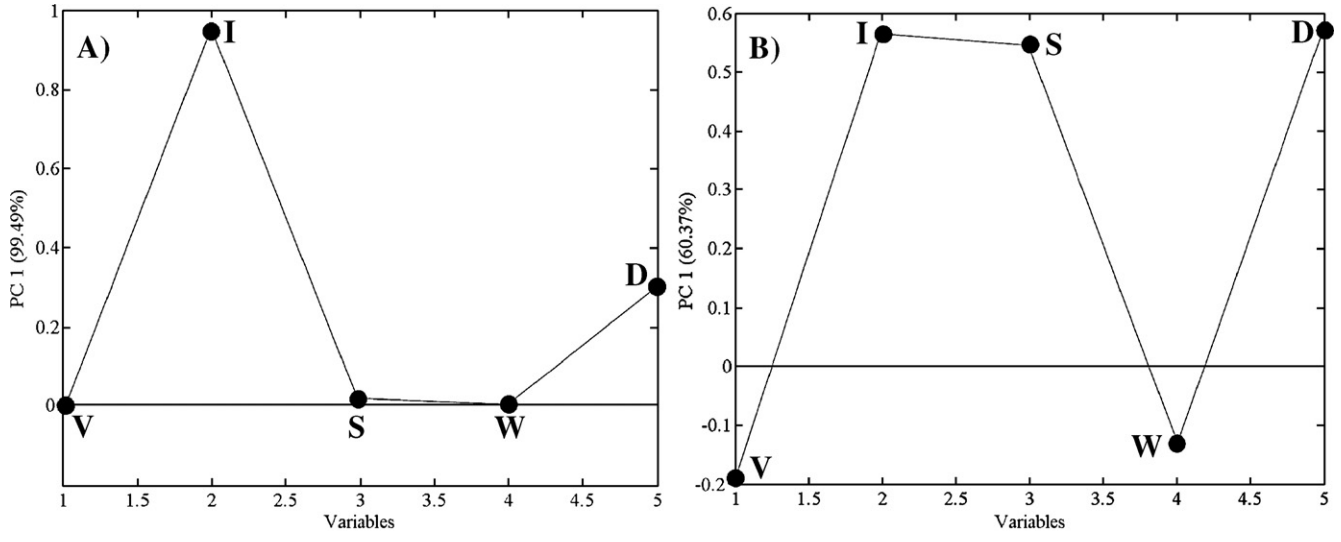

Fig. 3. Loadings plot corresponding to the first principal component for A) mean-centred and B) autoscaled peak parameters after developing a PCA based on the voltammetric peak parameters of the signals. *V*=peak potential (V); *I*=peak intensity (nA); *S*=area of the peaks ($\times 10^{-8}$ a.u.); *W*=half width (V); *D*=derivative ($\times 10^{-7}$ a.u.).

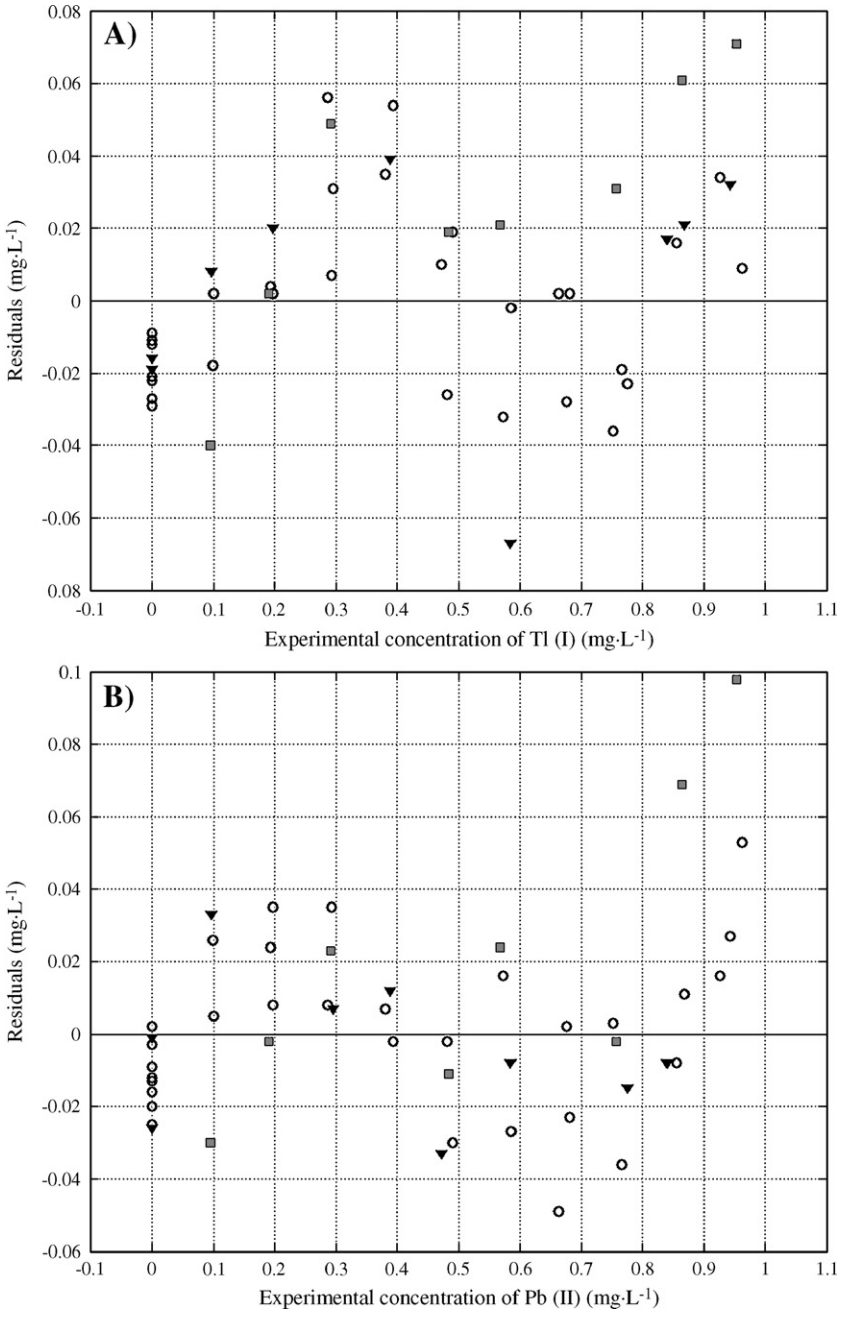

Fig. 4. Plot of the residuals versus the experimental concentrations of A) Tl^{+} and B) Pb^{2+}, respectively, for the ANN-based calibration model *ISD* lgg: -○- training set (trn); -▼- monitoring set (mon); -■- test set (tst).

chemometric techniques (MLR and PLS) are not important enough to prefer one of them instead of the others.

The removal of *V* and *W* from the most complex model (based on 5 parameters) causes a light improvement in the RMS error values for the external test set (RMS_{tst}). Nevertheless, the differences found now are not as high as in the previous chemometric techniques, since in ANN models a new factor must be taken into account: the combination of transfer functions. When removing *V* and/or *W* from a model more complex, the resultant model does not give the same RMS error values because of the different combination of transfer functions.

The best neural model selected also agrees the previous cases: *ISD* lgg ≡ *ISD* (3) ≡ *ISD* (best ANN, PLS and MLR models, respectively), although the RMS errors are lower. However, a model based on two parameters gives also rather good results and is the simplest of all tested: *SD* lgg.

Comparing the procedure described in the present work with those applied in former studies: a strategy based on FT and WT coupled to different regression methods [19] and a WT-based feature selection procedure as a previous step to the multivariate calibration process [20], the RMS errors are very similar too. More specifically, the RMS errors obtained with the use of voltammetric peak parameters are lower than when the process

Table 4
Average relative errors (%) for the optimal multivariate calibration models based on voltammetric peak parameters

Model	Ion	Error $(\%)_{trn}$	Error $(\%)_{mon}$	Error $(\%)_{tst}$
ISD≡*ISD* (3)	Tl^{+}	8.38	5.93	9.10
	Pb^{2+}	4.21	5.26	12.29
ISD lgg	Tl^{+}	5.29	6.28	10.77
	Pb^{2+}	5.43	7.25	8.15

I=peak intensity (nA); *S*=area of the peaks ($\times 10^{-8}$ a.u.); *D*=derivative ($\times 10^{-7}$ a.u.); trn=training set; mon=monitoring set; tst=test set.

described in Ref. [19] is carried out and slightly worse than when the procedure reported in Ref. [20] is applied. The last case was possible thanking to the advantage that WT provides when developing a feature selection process (FS). Nevertheless, the simplicity and fastness (low computation time) of the strategy that has been reported here is clearly evident.

The plots of the residuals for both ions are reported in Fig. 4. These residuals were calculated from the optimal ANN-based (*ISD* lgg) calibration model (remember that this model was almost equivalent to the optimal PLS and MLR models: *ISD* and *ISD* (3), with slightly better RMS errors). The trend observed in the figure is, in general, satisfactory.

The average relative errors (%) corresponding to the different sets of samples obtained with the optimal calibration models of every chemometric technique appear in Table 4. The error values for MLR and PLS models are identical, since PLS models used a number of LVs equal to their dimensions, what turns them into multilinear regressions. It has to be noticed that average relative errors of the test sets in the case of ANNs were very similar than those belonging to PLS and MLR. Considering the case of the ion Tl^{+}, the average relative errors were slightly higher, while with the ion Pb^{2+} ANNs reduced sensibly this value (almost 4%). This fact implies that Tl^{+} concentrations are predicted better with PLS or MLR, and Pb^{2+} concentrations with ANNs. In every model, one of the ions is predicted better than the other one [19,20].

Finally and summarising the results described previously, all the chemometric techniques applied predicted the concentrations of Tl^{+} and Pb^{2+} in a similar way. Besides, the error values were also comparable with those belonging to some studies previously published [11,13,19,20]. It has not to be forgotten that the experimental conditions used in this work are different from those described in the quoted references (see Refs. [11] and [13]), resulting in a harder overlapping between the signals of both ions. We recommend the use of MLR and PLS when a strong component of linearity between dependent and independent variables exists; otherwise, if the non-linearity relationship is rather significant, ANN calibration models are advisable. Moreover, pre-treatment techniques should be used when relevant information is necessary to be extracted from the measured data.

4. Conclusions

According to the former discussion, independent of the chemometric technique applied, the results were similar in all cases, although slightly better for ANN-based calibration models. In this way, any of these methods can be chosen in order to resolve this problem.

A study of the relevance of the voltammetric parameters to resolve the binary system was also carried out. The parameters *V* and *W* were irrelevant, since models differing in the presence or absence of these parameters gave identical results. The combination of the parameters *I*, *S* and *D* offered lower RMS errors for all the sets of samples; this means that they contained the most relevant information of the voltammetric signals. Moreover, models only composed of *S* and *D* gave slightly higher RMS errors for all the sets of samples than other more complex models.

In general, it can be affirmed that the results do not differ significantly from those obtained with other strategic and statistical procedures, based mainly on mathematical pre-treatments as a previous step to the multivariate calibration process, although the errors values in some of these cases are lower.

Acknowledgements

Financial support from Junta de Andalucía is acknowledged. We also thank Ministerio de Educación y Ciencia of Spain and European Community (FEDER) for the help given through the Project CTQ 2004-03708.

References

[1] B. Raspor, I. Pizeta, M. Branica, Anal. Chim. Acta 285 (1994) 103–111.
[2] H.N.A. Hassan, M.E.M. Hassouna, I.H.I. Habib, Talanta 46 (1998) 1195–1203.
[3] M.C. Antunes, J.E.J. Simão, A.C. Duarte, R. Tauler, Analyst 127 (2002) 809–817.
[4] M. Esteban, C. Arino, J.M. Díaz-Cruz, M.S. Díaz-Cruz, R. Tauler, Trends Anal. Chem. 19 (2000) 49–61.
[5] H. Martens, T. Naes, Multivariate Calibration, Wiley, Chichester, 1989.
[6] R.G. Brereton, Analyst 125 (2000) 2125–2154.
[7] P. Geladi, Chemom. Intell. Lab. Syst. 60 (2002) 211–224.
[8] K. Bessant, S. Saini, J. Electroanal. Chem. 489 (2000) 76–83.
[9] J. Saurina, S.H. Cassou, E. Fabregas, S. Alegret, Anal. Chim. Acta 405 (2000) 153–160.
[10] Y. Ni, L. Wang, S. Kokot, Anal. Chim. Acta 439 (2001) 159–168.
[11] M.C. Ortiz, J. Arcos, L. Sarabia, Chemom. Intell. Lab. Syst. 34 (1996) 245–262.
[12] A. Henrion, R. Henrion, G. Henrion, F. Sholz, Electroanalysis 2 (1990) 309–312.
[13] A. Herrero, M.C. Ortiz, Talanta 46 (1998) 129–138.
[14] R.M. de Carvalho, C. Mello, L.T. Kubota, Anal. Chim. Acta 420 (2000) 109–121.
[15] E. Cukrowska, L. Trnkova, R. Kizek, J. Havel, J. Electroanal. Chem. 503 (2001) 117–124.
[16] A. Cladera, J. Alpízar, J.M. Estela, V. Cerdà, M. Catasús, E. Lastres, L. García, Anal. Chim. Acta 350 (1997) 163–169.
[17] R.G. Brereton, Analyst 112 (1987) 1635–1657.
[18] C. Bessant, S. Saini, Anal. Chem. 71 (1999) 2806–2813.
[19] J.M. Palacios-Santander, A. Jiménez-Jiménez, L.M. Cubillana-Aguilera, I. Naranjo-Rodríguez, J.L. Hidalgo-Hidalgo-de-Cisneros, Microchim. Acta 142 (2003) 27–36.
[20] M. Cocchi, J.L. Hidalgo-Hidalgo-de-Cisneros, I. Naranjo-Rodríguez, J.M. Palacios-Santander, R. Seeber, A. Ulrici, Talanta 59 (2003) 735–749.
[21] M.M. Ghoneim, A. Tawfik, A. Radi, Anal. Bioanal. Chem. 374 (2002) 289–293.
[22] F. Ricci, C. Gonçalves, A. Amine, L. Gorton, G. Palleschi, D. Moscone, Electroanalysis 15 (2003) 1204–1211.
[23] J.J. Calvente, M.L. Gil, M.D. Sánchez, R. Andreu, F. de Pablos, Electrochim. Acta 45 (2000) 3087–3097.

[24] M.-H. Kim, L. Yan, R.L. Birke, M.-Z. Czae, Electroanalysis 15 (2003) 1541–1553.
[25] M. Farková, E.M. Peña-Méndez, J. Havel, J. Chromatogr., A 848 (1999) 365–374.
[26] B.K. Alsberg, M.K. Winson, D.B. Kell, Chemom. Intell. Lab. Syst. 36 (1997) 95–109.
[27] D.A. Sadler, P.R. Boulo, J.S. Soraghan, D. Littlejohn, Spectrochim. Acta, Part B 53 (1998) 821–835.
[28] D.A. Sadler, D. Littlejohn, P.R. Boulo, J.S. Soraghan, Spectrochim. Acta, Part B 53 (1998) 1015–1030.
[29] C. Locatelli, G. Torsi, Electroanalysis 10 (1998) 904–907.
[30] J. Alpízar, A. Cladera, V. Cerdà, E. Lastres, L. García, M. Catasús, Anal. Chim. Acta 340 (1997) 149–158.
[31] Y. Zhou, A. Yan, H. Xu, K. Wang, X. Chen, Z. Hu, Analyst 125 (2000) 2376–2380.
[32] S.M. Scott, D. James, Z. Ali, W.T. O'Hare, F.J. Rowell, Analyst 128 (2003) 966–973.
[33] A. Guiberteau Cabanillas, T. Galeano Díaz, N.M. Mora Díez, F. Salinas, J.M. Ortíz Burguillos, J.-C. Viré, Analyst 125 (2000) 909–914.
[34] General Purpose Electrochemical System (GPES), Autolab® Electrochemical Instruments Manual, Eco Chemie, Utrecht, The Nederlands, 1995, p. 74.
[35] A.M. Bond, Modern Polarographic Methods in Analytical Chemistry, Marcel Dekker, New York, 1980, p. 250.
[36] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. de Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics (Part A and B), Data Handling in Science and Technology, vol. 20A–B, Elsevier, Amsterdam, 1998.
[37] J. Zupan, J. Gasteiger, Neural Networks for Chemists: An Introduction, VCH, Weinheim, 1992.
[38] F. Despagne, D.L. Massart, Analyst 123 (1998) 157R–178R.
[39] J.W. Kauffman, P.C. Jurs, J. Chem. Inf. Comput. Sci. 41 (2001) 408–418.