



Multivariate linear regression with variable selection by a successive projections algorithm applied to the analysis of anodic stripping voltammetry data



Paola D. Marreto^a, Alessandro M. Zimer^a, Ronaldo C. Faria^a, Lucia H. Mascaro^a,
Ernesto C. Pereira^a, Wallace D. Fragoso^b, Sherlan G. Lemos^{b,*}

^a Laboratório Interdisciplinar de Eletroquímica e Cerâmica (LIEC), Universidade Federal de São Carlos (UFSCar), CP 676, 13565-905, São Carlos, SP, Brazil

^b Departamento de Química, Universidade Federal da Paraíba (UFPB), CP 5093, 58051-970, João Pessoa, PB, Brazil

ARTICLE INFO

Article history:

Received 19 October 2013

Received in revised form 5 February 2014

Accepted 7 February 2014

Available online 19 February 2014

Keywords:

Anodic stripping voltammetry

Mercury film electrode

Chemometrics

Multivariate calibration

Variable selection

ABSTRACT

Multivariate linear regression aided by a successive projections algorithm (SPA-MLR) was applied in the evaluation of anodic stripping voltammetry data obtained in the simultaneous determination of metals under conditions where there were significant complications due to interference processes such as the formation of intermetallic compounds and overlapping peaks. Using simulated data, modeled from complex interactions experimentally observed in samples containing Cu and Zn, as well as Co and Zn, it was demonstrated that SPA-MLR selected variables that allow chemical interpretation. This feature was used to make inferences about the underlying electrochemical processes during the simultaneous determination of four metals (Cu, Pb, Cd, and Co) in a concentration range where all responses were complicated by interference processes ($10\text{--}100\text{ ng mL}^{-1}$). Additionally, the analytical performances of MLR models for quantitative predictions were excellent despite the complexity of the system under study.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

In electroanalytical chemistry, the traditional approach is the use of univariate models, i.e., calibration curves constructed from a single characteristic of the sample. In voltammetry, for example, the intensity of the peak current is the most commonly used variable. A different approach called multivariate modeling, which is less sensitive to the presence of interfering substances, uses more than one variable simultaneously. This is equivalent, for example, to construct a calibration model with the current intensity measured at various potentials in the voltammogram. Multivariate linear regression (MLR) is a multivariate natural expansion of univariate linear regression and is the simplest procedure to perform a multivariate calibration. However, this technique is not efficient if there is significant collinearity in the data matrix (as for voltammetric data), and if the number of variables is greater than the number of samples. The most common way to avoid this problem is the use of latent variables or latent structures methods. These methods use linear combinations of the original variables as new variables.

Commonly, a small number of orthogonal or almost orthogonal latent variables can be obtained and used for multivariate calibration.

The selection of predictor variables prior to multivariate regression is a practice that can provide significant improvement in the prediction results when compared to the use of full data (spectra or voltammogram), mainly by discarding variables not related to the analytical response and which only incorporate noise into the regression model [1–3]. Another feature that has been highlighted as an advantage of variable selection is preservation of the original variables domain when the calibration is based on MLR. Consequently, it is easy to propose the physical interpretation of mathematical models, in contrast to what is obtained with methods based on latent structures such as partial least squares (PLS), and principal component regression (PCR). In practice, this advantage has not been greatly explored mainly due to focus on the predictive ability of the models and their comparison with different methods [4]. Additionally, there is little demand for the physical interpretation of variables obtained using spectroscopy techniques because the important spectral ranges are commonly known [5].

Voltammetric measurements have been associated with multivariate calibration to provide simultaneous determinations in a multicomponent system [6–15]. Although this is a growing trend,

* Corresponding author. Tel.: +55 83 3216 7438; fax: +55 83 3216 7438.

E-mail addresses: sherlan03@yahoo.es, sherlan@quimica.ufpb.br (S.G. Lemos).

there are fewer successful cases described in the literature compared with spectrometric data. An important difference between these two techniques is that, in voltammetric results, interference processes can introduce new voltammetric peaks such as those related to intermetallic compounds in the simultaneous determination of metallic ions by anodic stripping voltammetry (ASV). Since this kind of interference is related to the metal concentrations themselves, new signals must be included in the calibration model, which necessarily becomes a multivariate model. It is, however, very difficult to obtain information about the signals of intermetallic species. For this reason, many researchers prefer to build multivariate calibration models with all voltammograms, using latent structure methods and including many signals without correlation with the metal concentration [16].

Wang [17], in his famous book, asserts that interference processes in ASV usually occur when metallic thin films are used as the working electrode. These substrates offer many advantages over bulk metal electrodes such as an improved scope for different cell configurations and for chemical modification of their surface, lower cost (requires only minute quantities of the metal to assemble the film), a larger surface-to-volume ratio, and mechanical stability (in comparison with mercury drops, for example) [18]. However, due to the larger surface-to-volume ratio, metallic films are more susceptible to the formation of intermetallic compounds. In this case, the common approaches for eliminating or correcting such interference include removing the interfering response by adding a masking substance [19,20] or the single standard addition method [21–23].

In this work, we intended to evaluate MLR coupled to variable selection performed by the successive projections algorithm [24] (SPA-MLR) in multivariate calibration for voltammetric data using simulated voltammograms and real measurements in the simultaneous determination of metals by anodic stripping voltammetry. The focus was on observing if variable selection could reveal variables apparently not related to a specific analyte, denoting unknown electrochemical phenomena such as coupled reactions, the formation of intermetallic compounds, matrix effects, etc. It was expected that this guided soft modeling semi-empirical approach would provide a very useful qualitative interpretation of the analyzed electrochemical system in addition to quantitative predictions in the same way as hard modeling [25,26].

First, simulated voltammograms were employed in order to explore and discuss the performance of SPA-MLR under a variety of conditions of intermetallic compound formation and peak overlap. Two scenarios were studied. The first one involved the simultaneous determination of Cu and Zn considering the formation of an intermetallic compound CuZn. Pb was also introduced in this simulation in order to evaluate whether the selection algorithm would select variables not related to the analytes, once we considered that there is no interaction of Pb with Cu and Zn. It is important to point out that Cu-Pb intermetallic effects have been reported in ASV [27,28]. However, no intermetallic formation was considered in the present simulations. The second scenario involved the simultaneous determination of Cu, Zn and Co, considering the formation of the intermetallic compounds CuZn and CoZn.

SPA-MLR was also submitted to the analysis of a real system containing five metals that interact to varying degrees in a single sample. In the present work, the simultaneous DPASV analysis of Cu, Co, Pb, Cd, and Zn in water samples with a mercury thin-film electrode (MTFE) was investigated. All analytes were studied in the range from 10 to 100 ng mL⁻¹. Thus, several sorts of interactions are expected such as intermetallic compounds between Cu and Zn, Cu and Cd, and between Cu and Co, to name a few. They result in current-concentration relationships that are difficult to represent using a simple univariate regression model.

2. SPA-MLR background

The SPA is a variable selection algorithm that uses sets of calibration and validation data containing instrumental responses (\mathbf{X}) and parameter values measured by a reference method (\mathbf{y}). It initially defines the \mathbf{X}_{CAL} matrix ($K_C \times J$), where rows correspond to K_C samples of the calibration set and columns correspond to J variables, which in the context of this work are the intensities of current at each potential of the full voltammogram. From a \mathbf{X}_{CAL} column matrix, arbitrarily chosen and named \mathbf{x}_0 , SPA determines which of the other columns has the largest projection in the subspace S_0 orthogonal to \mathbf{x}_0 . This column is named \mathbf{x}_1 and can be interpreted as containing the largest amount of information not included in \mathbf{x}_0 . In the next iteration, the SPA uses \mathbf{x}_1 as the new reference column, and proceeds as above to select \mathbf{x}_2 . The algorithm follows with projections until a certain number of variables potential minimally collinear with each other is selected.

The maximum number of variables that can be selected is K_C , since the dimension of the column space of \mathbf{X}_{CAL} is reduced by one after each iteration, i.e., one degree of freedom is removed. Therefore, after K_C iterations, all column vectors of \mathbf{X}_{CAL} have been projected on the origin of the space and \mathbf{X}_{CAL} will become a matrix of null rank. Several variable chains are built upon the selection of each one of J variables as the initialization vector \mathbf{x}_0 in successive projections, and varying the length N of the chains, typically between 2 and K_C . The best variable chain is selected by building a MLR model for each chain, and validated with a validation samples set constituted by new samples that did not enter the calibration set. The chain of choice is one that corresponds to the MLR model that has the lowest root mean square error of validation.

The next stage of SPA-MLR is necessary because the construction of the variable chains takes place solely based on minimizing the collinearity between the variables and does not take into account the correlation between each variable and the response, i.e., concentration of the analyte contained in the vector \mathbf{y} . Thus, a procedure that eliminates the variables has been selected but not correlate with the concentrations is performed in order to obtain a simpler model.

3. Experimental

3.1. Simulated data

Data simulation was based on the approach used by Stromberg and Gorodovych [29], who modeled the Cu-Zn system in which the intermetallic compound is apparently insoluble. Eq. 1 shows the expression used for simulating responses characteristic of redox species in a thin-layer [30]:

$$i = \frac{n^2 F^2 \nu V C}{RT} \frac{\exp\left[\frac{nF}{RT}(E - E^0)\right]}{\left\{1 + \exp\left[\frac{nF}{RT}(E - E^0)\right]\right\}} \quad (1)$$

In this equation, n is the number of electrons involved in the reaction, F is the Faraday constant, ν is the scan rate, V is the volume of the solution and R the gas constant, E is the applied potential and E^0 is the formal potential of the species. Then, it is possible to calculate the current, i , for a concentration C of the analyte, which in anodic stripping voltammetry is represented by the concentration of the mercury electrode- C_{Hg} -given by eq. 2.

$$C_{\text{Hg}} = \gamma C \quad (2)$$

In eq. 2, γ is the accumulation coefficient and is dependent on deposition time, mass transport of the analyte to the electrode surface, concentration of the analyte in the bulk solution, C , and electrode area. As stated previously, the behavior of the Cu-Zn system has been characterized and can be related to the concentrations

Download English Version:

<https://daneshyari.com/en/article/186134>

Download Persian Version:

<https://daneshyari.com/article/186134>

[Daneshyari.com](https://daneshyari.com)