

Research Article

Online Monitoring of Copper Damascene Electroplating Bath by Voltammetry: Selection of Variables for Multiblock and Hierarchical Chemometric Analysis of Voltammetric Data

Aleksander Jaworski, Hanna Wikiel, and Kazimierz Wikiel

Technic Inc., 47 Molter St., Cranston, RI 02910, USA

Correspondence should be addressed to Aleksander Jaworski; ajaworski@technic.com

Received 5 August 2017; Accepted 19 November 2017; Published 14 December 2017

Academic Editor: Gerd-Uwe Flechsig

Copyright © 2017 Aleksander Jaworski et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The Real Time Analyzer (RTA) utilizing DC- and AC-voltammetric techniques is an in situ, online monitoring system that provides a complete chemical analysis of different electrochemical deposition solutions. The RTA employs multivariate calibration when predicting concentration parameters from a multivariate data set. Although the hierarchical and multiblock Principal Component Regression- (PCR-) and Partial Least Squares- (PLS-) based methods can handle data sets even when the number of variables significantly exceeds the number of samples, it can be advantageous to reduce the number of variables to obtain improvement of the model predictions and better interpretation. This presentation focuses on the introduction of a multistep, rigorous method of data-selection-based Least Squares Regression, Simple Modeling of Class Analogy modeling power, and, as a novel application in electroanalysis, Uninformative Variable Elimination by PLS and by PCR, Variable Importance in the Projection coupled with PLS, Interval PLS, Interval PCR, and Moving Window PLS. Selection criteria of the optimum decomposition technique for the specific data are also demonstrated. The chief goal of this paper is to introduce to the community of electroanalytical chemists numerous variable selection methods which are well established in spectroscopy and can be successfully applied to voltammetric data analysis.

1. Introduction

Electrochemically deposited copper from an acidic bath is a commonly used method of on-chip production of interconnects for microelectronics [1]. The strict requirements of chip manufacturing demand rapid and void-free filling of high-aspect ratio topologies of sub-100 nm features (as small as 18 nm, currently). Rapid superconformal filling is achieved by using multicomponent plating solutions that contain inorganic and organic constituents of various chemical characteristics present at concentrations which differ by several orders of magnitude. The narrow tolerances allowed for the final product require accurate and reliable control of the plating solution with similar or even narrower tolerances.

Accurate and prompt concentration monitoring and control of multicomponent electroplating baths are indispensable in satisfying process specifications for the manufacturing of electronic components while minimizing production

costs. The Real Time Analyzer, utilizing numerous voltammetric techniques, is an in situ, online monitoring system that provides a complete chemical analysis of electrometallization solutions. The fully computerized instrumentation requires no specially trained chemical operators and practically eliminates the need for a chemical analytical laboratory.

Typically used organic additives include (i) suppressors (like polyethers) that inhibit the rate of copper deposition at the tops of the trenches and vias by increasing the copper ion reduction overpotential by surface absorption interaction with chloride ion [2, 3]; (ii) accelerators, disulfide molecules, which facilitate the reduction process of copper, most probably by anion-induced adsorption of copper complex(es) at the metallic surface [4–6]; and (iii) levelers, high molecular weight amines or amides, that act as secondary suppressors to control the grain size of the deposited copper and inhibit overplating [7]. In combination, these additives can achieve accelerated, bottom-up electrodeposition of copper

in submicron inlaid features, which permits void-free interconnect wiring in damascene structures. This paper presents a description of RTA data analysis using the example of a suppressor. Although the presentation deals with suppressor analysis, the approach presented is general and is utilized by the RTA for determining the concentration of all plating bath components.

The organic additives undergo significant changes of concentration during bath usage [4, 8, 9]. Therefore, their concentrations should be closely and accurately monitored and replenished in order to maintain their concentrations at the level corresponding to the optimum plating performance.

Ni and Kokot [10] explored the question of whether chemometrics methods enhance the performance of electro-analytical methods and provided evidence for a strong affirmative reply to this question. Despite numerous undisputable benefits, the application of chemometrics to electroanalysis has not been used nearly as widely as in spectroscopy or chromatography.

Wikiel et al. [11] and Jaworski et al. [12–16] extensively studied the application of soft modeling techniques for the development of analytical models aimed at accurately and robustly predicting the concentrations of deliberately added bath constituents. Also, Wikiel et al. [17] and Jaworski et al. [18] employed various soft modeling techniques for determinant analyses implemented online as early fault detection routines for industrial electroplating solutions.

Specific waveforms are developed to produce voltammograms having regions which show linear dependence of the current on the concentration of the analyte of interest, while being practically immune to varying concentrations of all other bath constituents. Sometimes, the voltammograms recorded for a single waveform contain several portions (ranges of points of voltammogram) meeting this objective. For such a waveform, its data can be divided into meaningful blocks in order to improve the interpretability. The ability of building a multivariate analytical model utilizing information contained in each of the ranges (blocks) of the voltammogram leads to increased diversity within the data as compared to single-range-based data sets. The diversity among the data results in a greater robustness of the calibration model calculated based on that data.

Hierarchical Principal Component Regression (HPCR) [12, 19], Hierarchical Partial Least Squares (HPLS) [12, 19], Consensus PCR (CPCR) [12, 19], and Multiblock PLS (MBPLS) [12, 19] methods provide tools for handling variables arranged into meaningful blocks, which can be decomposed and subsequently (HPCR, CPCR) or simultaneously (HPLS, MBPLS) regressed against the dependent variables (concentrations).

The chief goal of this presentation is to show that the variable selection methods whose applications are well established in spectroscopy can also be transferred to electroanalytical data. Specifically, this objective is achieved by an introduction of a rigorous, multistep procedure for selecting the blocks of the voltammogram to be subsequently used for analytical model developments and a choice of proper data decomposition technique in order to compress

the multivariate voltammetric data and reasonably extract information.

Wikiel et al. [11] and Jaworski et al. [12] introduced a technique coupling Simple Modeling of Class Analogy- (SIMCA-) based modeling power [20] and Least Squares Regression (LSR) for the selection of portions of voltammograms for further factor and regression analyses. The aim of this paper is to extend the criteria for determination of optimum ranges of voltammograms by a novel application of techniques such as Uninformative Variable Elimination- (UVE-) PLS [21, 22], UVE-PCR [23], and PLS-Variable Importance in the Projection (VIP) [24, 25] along with Interval PLS (IPLS) [26], Interval PCR (IPCR), and Moving Window PLS (MWPLS) [27] in electroanalysis. These variable selection techniques have been used in spectroscopy (predominantly near infrared and infrared [22, 26, 27]), chromatography, and mass spectrometry, but with the exception of IPLS [28] they have not been applied to electroanalysis. In recent years, the use of variable selection methods in trace analysis of metals by anodic stripping voltammetry (ASV) surfaces sporadically [28, 29], sometimes being combined with Multivariate Linear Regression on aligned ASV peaks rather than chemometric data compression techniques [29]. The marginal utilization of variable selection techniques in electroanalysis has persisted despite promising results in 1999 [30] obtained by application of genetic algorithm as a variable selection method in the multivariate analysis with PLS of several polarographic and stripping voltammetric data sets, where different interferences were present.

Modern chemometrics is a mature scientific discipline presenting the researcher with a vast number of powerful data decomposition techniques. Although some techniques appear more advanced than the others because of their mathematical complexity, the most suitable methods should be properly chosen depending on the kind of data to be analyzed in order to develop a sound and robust analytical model.

2. Experimental

Voltammetric experiments were performed utilizing the Real Time Analyzer (Technic, Inc., Cranston, USA), a fully computer-control electroanalytical system. Measurements were conducted inside a compact flow-through electrochemical cell (electrode compartment of the Multi-Task Electrochemical Probe (MTEP)) submersed in the temperature-controlled ($25 \pm 0.2^\circ\text{C}$) plating solution. The volume of the inner cell compartment was about 20 mL. The solution to be analyzed was circulated inside the probe using a software-controlled diaphragm pump (KNF Neuberger, Baltherswil, Switzerland).

The electrochemical cell was a classical three-electrode system with a working electrode made of platinum (Johnson-Matthey) wire (1 mm diameter; 10 mm length), an auxiliary platinum (Johnson-Matthey) foil electrode forming a cylinder around the working electrode, and an in situ generated reference electrode made of copper metal deposited onto a platinum (Johnson-Matthey) wire, immediately before the measurement.

All inorganic chemicals were of Analytical Grade (J.T. Baker, Phillipsburg, NJ). A proprietary organic additive system (Enthone, West Haven, CT) was used in this study.

All calculations were performed in the MATLAB R2012b (The MathWorks, Inc., Natick, MA) environment. The procedure for PLS and scaling routines were taken from the PLS Toolbox 6.5.4. (Eigenvector Research, Inc., Manson, WA, <http://www.eigenvector.com>). The procedure for MWPLS was taken from iToolbox (University of Copenhagen, Denmark, <http://www.models.life.ku.dk/iToolbox>). All other procedures were written by the authors.

3. Results and Discussion

3.1. Composition of Training and Validation Sets. The subject of this investigation was the ViaForm[®] copper plating bath which consisted of six components: copper (II) ion (from copper sulfate), sulfuric acid, chloride ion, suppressor, accelerator, and leveler present for the target concentrations of 0.785 M, 0.820 M, 1.50 mM, 7.00 mL L⁻¹, 7.00 mL L⁻¹, and 0.76 mL L⁻¹, respectively.

The specific objective was to create a calibration model for suppressor in the presence of varying concentration of accelerator and leveler. The concentrations of suppressor, accelerator, and leveler were varied linearly on four levels within the ranges of 5.00 to 9.00 mL L⁻¹, 5.00 to 9.00 mL L⁻¹, and 0.38 to 1.13 mL L⁻¹, respectively. The training set was composed as a 5-level-3-component linear orthogonal array exploring uniformly distributed 25 combinations of suppressor, accelerator, and leveler concentrations. Additionally, the training set was augmented by the voltammetric data recorded for three standard solution containing suppressor, accelerator, and leveler concentrations at the low, target, and high limits. The inorganic bath components, copper (II) ion, sulfuric acid, and chloride ion, were held constant at their target levels. The voltammetric data were recorded for each of the 28 solutions in triplicate resulting in 84 samples ($I = 84$). The composition of the entire training set is presented in Table 1.

In order to assess the predictive abilities, the calibration model was externally validated on the validation set. The external validation set consisted of 27 data sets obtained for 9 solutions containing suppressor, accelerator, and leveler varied linearly on three levels within the ranges of 5.50 to 8.50 mL L⁻¹, 5.50 to 8.50 mL L⁻¹, and 0.47 to 1.04 mL L⁻¹, respectively. The validation set was a 3-level-3-component linear orthogonal array exploring uniformly distributed 9 combinations of suppressor, accelerator, and leveler concentrations. The composition of the validation set is presented in Table 2.

Waveform design is a preliminary step of the plating bath analysis utilizing voltammetry [11]. This procedure aims to obtain waveforms which are bath specific and are designed to produce a current response that changes linearly with the concentration of the analyte. It is important in waveform development to minimize the influence of the concentration changes of components other than the analyte on the current response of the voltammogram of the particular analyte.

TABLE 1: Composition of the training set solutions.

Calibration solution #	Suppressor mL L ⁻¹	Accelerator mL L ⁻¹	Leveler mL L ⁻¹
C1	5.00	5.00	0.38
C2	5.00	6.00	0.57
C3	5.00	7.00	0.76
C4	5.00	8.00	0.94
C5	5.00	9.00	1.13
C6	6.00	5.00	0.57
C7	6.00	6.00	0.76
C8	6.00	7.00	0.94
C9	6.00	8.00	1.13
C10	6.00	9.00	0.38
C11	7.00	5.00	0.76
C12	7.00	6.00	0.94
C13	7.00	7.00	1.13
C14	7.00	8.00	0.38
C15	7.00	9.00	0.57
C16	8.00	5.00	0.94
C17	8.00	6.00	1.13
C18	8.00	7.00	0.38
C19	8.00	8.00	0.57
C20	8.00	9.00	0.76
C21	9.00	5.00	1.13
C22	9.00	6.00	0.38
C23	9.00	7.00	0.57
C24	9.00	8.00	0.76
C25	9.00	9.00	0.94
C26	5.00	5.00	0.38
C27	7.00	7.00	0.76
C28	9.00	9.00	1.13

TABLE 2: Composition of the validation set solutions.

Validation solution #	Suppressor mL L ⁻¹	Accelerator mL L ⁻¹	Leveler mL L ⁻¹
V1	5.50	5.50	0.47
V2	5.50	7.00	0.75
V3	5.50	8.50	1.04
V4	7.00	5.50	0.75
V5	7.00	7.00	1.04
V6	7.00	8.50	0.47
V7	8.50	5.50	1.04
V8	8.50	7.00	0.47
V9	8.50	8.50	0.75

In some cases, it is possible to obtain a waveform producing voltammograms whose several portions can be utilized for calibration calculation. Figure 1 presents a waveform whose current response linearly varies only with the changing concentration of suppressor.

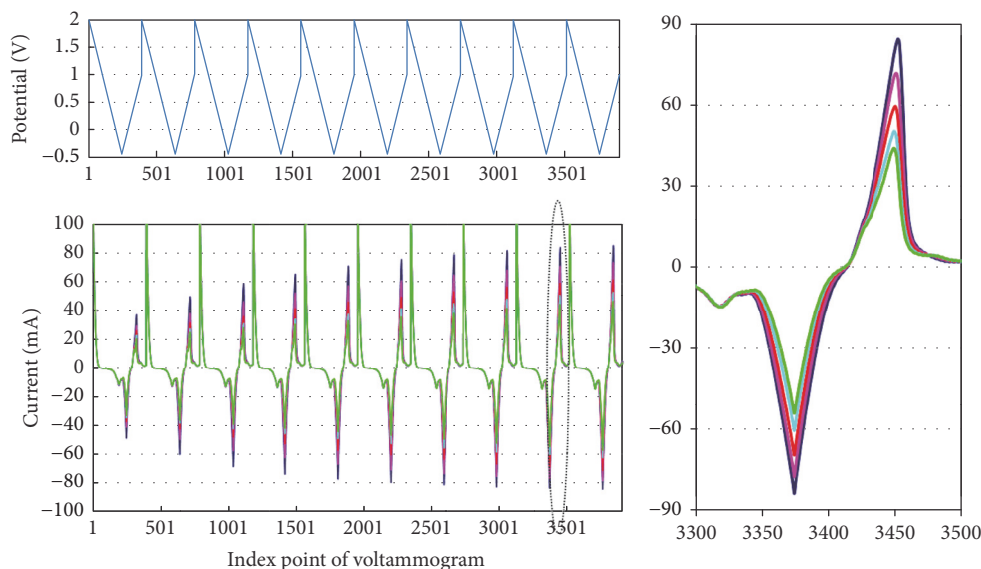


FIGURE 1: Multicyclic staircase voltammograms ($\Delta E = 10$ mV, $\Delta t = 1$ ms) of the training set corresponding to various concentrations of suppressor: 5.00 (dark blue, soln. *C1* in Table 1), 6.00 (magenta, *C8*), 7.00 (red, *C15*), 8.00 (light blue, *C19*), and 9.00 (green, *C21*) mL L⁻¹. Magnified portion corresponding to the ninth cycle.

The voltammetric data throughout this paper are predominantly considered numerical input for chemometric treatment aiming to select specific sections of the voltammograms which are most informative for subsequent calibration calculation. Therefore, the abscissa of the voltammograms in this paper is the index point of the voltammogram rather than the applied potential as is usually used. The values of the index points of the voltammogram are transformed values of the applied potential. The index points of voltammograms are a linear function of the voltammetric current sampling time. For the voltammogram of Figure 1, the current is sampled at the end of each step of duration of $\Delta t = 1$ ms. For instance, the 3350th point of the voltammogram is sampled at 3.350 s of the duration of voltammetric measurement when the applied potential is -208.7 mV versus $E_{\text{Cu}^{2+}/\text{Cu}}$ during the forward potential sweep of 9th CV cycle. The index points of the voltammogram are independent variables which are necessary for a unique identification of variables and relevant for subsequent chemometric selection. The applied potential values are not unique identifiers of independent variables, as the same potential values are applied several times during multicyclic voltammogram. For instance, the same potential value of -208.7 mV versus $E_{\text{Cu}^{2+}/\text{Cu}}$ is also applied at 3.398 s of the voltammograms of Figure 1 corresponding to 3398th index point of voltammogram during the reverse potential sweep of 9th CV cycle. The relationship between applied potential values and index point of voltammograms/time is presented in upper left corner of Figure 1. In order to present the entire voltammograms in Figure 1, scaling was applied which prevents the detailed quantitative visual analysis of the voltammograms. However, one can notice cyclical pattern among these voltammograms. Each of the voltammograms consists of ten distinctive cycles. Figure 1 also shows the

magnified portion of voltammograms focusing on the ninth of the ten cycles.

The recorded current corresponds to the effect of suppressor on copper ion reduction (index points 3340–3400) and on copper metal oxidation (index points 3420–3460). Examination of the voltammograms shown in Figure 1 enables estimation that the current response in some portions of these voltammograms is linearly dependent on concentration of suppressor, while being independent of concentration variations of other bath constituents (see composition on *C1*, *C8*, *C15*, *C19*, and *C21* solutions in Table 1). Analogous conclusions can be drawn while visually analyzing the individually magnified last five cycles of voltammograms of Figure 1.

3.2. Determination of the Optimum Ranges for a Calibration

3.2.1. SIMCA Modeling Power and LSR. Wiekil et al. [11] and Jaworski et al. [12] introduced a technique which proved to be helpful for selecting the range(s) of voltammograms to be taken for further decomposition and regression analysis. This method was employed [12] for selecting applicable ranges of various voltammograms to be subsequently analyzed by hierarchical and multiblock decomposition/regression techniques. This paper focuses on the applicable points of voltammograms of a single waveform (Figure 1). The method of determination of the optimum ranges of voltammograms presented in this paper is significantly improved and extended over that presented in [11–13] by introducing a two-step (whole voltammogram, then optimization within individual blocks) SIMCA-based selection by confirming the ranges by several independent approaches to exclude artifacts.

The initial stage of determining of the most promising ranges of the voltammogram to be taken for the calibration

calculation utilizes two independent procedures applied for each index point of that voltammograms within the training set data:

- (i) correlation calculation based on the univariate LSR,
- (ii) SIMCA -based on calculation of modeling power [20].

The LSR-based method provides information about which range(s) of the voltammogram shows the greatest correlation with the concentration of the component to be calibrated. It also determines the range where the current responses depend only on changes in concentration of the component of interest. In this method, the regression equation for each index point of the training set voltammograms of Figure 1 is calculated by the least squares method. The regression equation obtained is employed to self-predict concentrations for the training set. The self-predicted concentrations are correlated with the corresponding actual concentrations within the training set. The squared correlation coefficients, r_j^2 , are calculated for each j th point of the voltammogram.

The SIMCA-based method gives information about noise-to-signal ratio for each point within the chosen range. This method [20] utilizes Principal Component Analysis (PCA) calculated residuals to obtain the modeling power, R , for each j th point of the voltammogram:

$$R_j = 1 - \frac{rv_j(\text{error})}{rv_j(\mathbf{x})}, \quad (1)$$

where $rv_j(\text{error})$ is the square root of the residual variance for the number of factors of F for the j th point of the voltammogram:

$$rv_j^2(\text{error}) = \frac{\sum_{i=1}^I e_{i,j}^2}{(I - F - 1)}, \quad (2)$$

where $e_{i,j}$ is the element of the matrix of residuals and I is the number of samples of the training set (in the example discussed throughout the paper, $I = 84$).

$rv_j(\mathbf{x})$ is the square root of the meaningful variance for j th point of the voltammogram:

$$rv_j^2(\mathbf{x}) = \frac{\sum_{i=1}^I x_{i,j}^2}{(I - 1)}, \quad (3)$$

where $x_{i,j}$ is the element of the training set matrix \mathbf{X} .

The modeling power, as implemented in the initial stage of the method for selection of optimum ranges, provides information about noise-to-signal ratio for each point of the voltammogram. As R_j approaches 1, this feature becomes highly relevant; conversely, as it approaches 0, the feature approaches zero utility in the model [20].

Figure 2 presents the parameters r_j^2 and R_j , calculated for the entire voltammograms (all points of the voltammograms) of the waveform of Figure 1 recorded for the training set. It can be seen that values of both parameters r_j^2 and R_j , tend to change abruptly with sharp oscillations between extremes

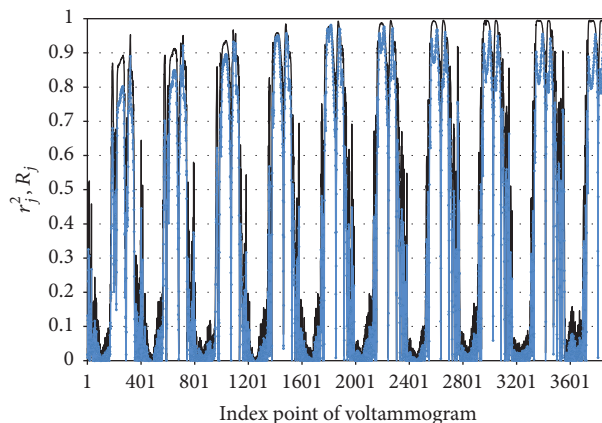


FIGURE 2: Squared coefficient of correlation-for-suppressor-concentration (r^2) (solid black line) and SIMCA modeling power (R) ($F = 1$) (solid blue line with diamonds) calculated for entire voltammograms of waveform presented in Figure 1.

related to useful and useless values for both modeling power and regression. R_j parameters were calculated based on the PCA decomposition using the number of factors ($F = 1$) as there is expected to be only one predominant source of variance in the selected ranges, namely, the changing concentration of suppressor. The preliminary selection of ranges of the voltammogram (Table 3) was conducted based on the criteria that the points of the voltammogram within each range should change monotonously and that every value of r_j^2 and R_j , within the range must not fall below 0.95 and 0.80, respectively. Also, each selected range was continuous, including all points of the voltammogram within the limits of the range.

The subsequent step (second step of the SIMCA modeling power) is to determine the optimum ranges of voltammograms based solely on the modeling power technique. The voltammograms of the training set for each of the $K = 14$ preliminarily selected ranges (Table 3) are then analyzed individually. In other words, separate modeling power analyses (each requiring another PCA decomposition) are applied individually for each of the ranges rather than for all of the points of the voltammogram (as the data of Figure 2 was obtained). Figure 3 presents modeling power values calculated individually for each of $K = 14$ ranges of Table 3 as compared to modeling power values calculated for the whole voltammogram.

One can notice in Table 3 that the average R values calculated within each range obtained by individual decomposition of each of the ranges (Figure 3) are systematically higher than average R values corresponding to the same points of the voltammogram but obtained by a decomposition of the entire voltammogram (Figure 2). This has an obvious explanation, as the PCA decomposition of the entire voltammogram needs also to model the variances within the portion of the voltammograms which have no utility for the suppressor calibration (e.g., those showing very low values of squared correlation coefficient) (Figure 2). The lower values of average R obtained for the entire voltammogram decomposition

TABLE 3: Preliminary estimation of the ranges of points of staircase voltammogram, variables j , of Figure 1 selected based on results of Figure 2. The potentials are expressed versus $E_{\text{Cu}^{2+}/\text{Cu}}$ reference electrode.

Data range index, k	Signal	Range of voltammogram, variables j	Width, J	E_{begin} , mV	E_{End} , mV	Average R (see (1)), whole scan decomposition	Average R (see (1)), individual range decomposition
(1)	Reduction	1401–1430	30	−268	−338	0.9414	0.9804
(2)	Oxidation	1476–1496	21	122	322	0.9204	0.9309
(3)	Reduction	1785–1826	42	−197	−287	0.9592	0.9663
(4)	Oxidation	1865–1888	24	102	333	0.9078	0.9085
(5)	Reduction	2173–2219	47	−168	−268	0.9379	0.9542
(6)	Oxidation	2256–2279	24	102	333	0.8871	0.9005
(7)	Reduction	2562–2612	51	−208	−248	0.9110	0.9455
(8)	Oxidation	2645–2670	26	82	333	0.8599	0.8895
(9)	Reduction	2952–3005	54	−137	−228	0.8867	0.9375
(10)	Oxidation	3038–3062	25	102	343	0.8518	0.8858
(11)	Reduction	3344–3399	56	−148	−198	0.8703	0.9351
(12)	Oxidation	3428–3454	27	92	353	0.8447	0.8755
(13)	Reduction	3734–3793	60	−134	−168	0.8564	0.9280
(14)	Oxidation	3819–3846	28	92	363	0.8436	0.8712

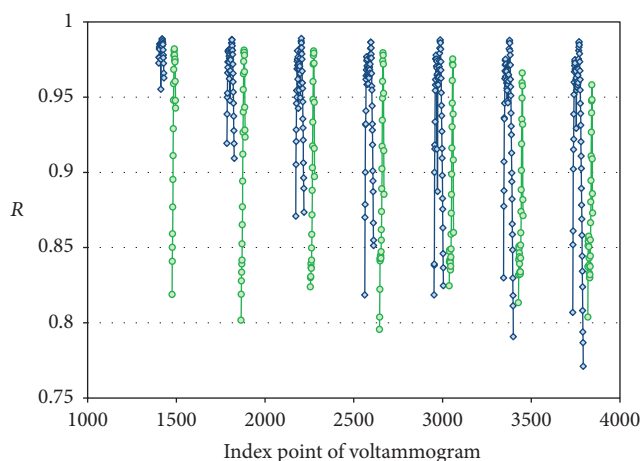


FIGURE 3: SIMCA modeling power (R) ($F = 1$) calculated individually and independently from other ranges for each of K ranges (Table 2). Ranges corresponding to reduction and oxidation signals are denoted with blue diamonds and green circles, respectively.

are the result of the compromise taken by decomposing unselected data (ranges) and therefore incorporating unnecessarily additional numerical noise. However, the R_j values for some j variables (usually for those close to the beginning or end of the range) obtained via individual decomposition of the preselected ranges may be lower than corresponding R_j values obtaining via decomposition of entire voltammetric data (e.g., see Figure 3; several last variables of 13th range are of lower R_j than 0.80). The individual selection of the ranges of voltammogram is shown in Table 4. This secondary selection was achieved by individually analyzing modeling power values for each of $K = 14$ ranges and by selecting points of the voltammograms corresponding to the values of R which must not be lower than 0.85.

As the values of modeling power, calculated via PCA decomposition with a single factor, presented in Figure 3 approach one, it means that there is single predominant, orthogonal variance within that data. This variance is caused by the varying concentration of suppressor, as the squared correlation coefficients calculated by LSR for these ranges of voltammograms are also approaching unity (Figure 2). This predominant variance is a focal point of the calculation of the analytical model aiming to predict the concentration of suppressor. Other, relatively small orthogonal variances are also present in the data. They may be worth of exploring for the final tuning of the analytical model but it should be determined by cross- and external validations whether they would be worth incorporating into the analytical model optimized for suppressor-caused variance only. The trade-off may be an unnecessary incorporation of numerical noise which lowers the predictive ability of the model. In order to avoid such a scenario, the final tuning of the model needs to be conducted based on external validation.

3.2.2. Uninformative Variable Elimination. The uninformative variables (index points of voltammogram) increase the bias and imprecision of the latent variables [21, 22]. While introducing UVE-PLS, Centner et al. [21] pointed out “in the original PLS method all variables are used; PLS is a so-called full-spectrum method. However, one can wonder whether it is useful to include all variables, because some of them may be noisy and/or contain nonrelevant information.” Elimination of uninformative variables leads to more parsimonious and more robust models and to better prediction [21, 22]. In the UVE method the random (and therefore uninformative) variables are appended to the original data matrix $\mathbf{X}^{(I \times J)}$ resulting in matrix $\mathbf{Z}^{(I \times 2J)}$ to calculate a reliability criterion for each original and each added random variable j and to retain only the experimental variables for which the value of the

TABLE 4: Individual (range by range) selection of the ranges of points of staircase voltammogram, variables j , of Figure 1 determined based on results of Figure 3. The potentials are expressed versus $E_{\text{Cu}^{2+}/\text{Cu}}$ reference electrode.

Data range index, k	Signal	Range of voltammogram, variables j	Width, J	E_{begin} , mV	E_{End} , mV
(1)	Reduction	1401–1430	30	–268	–338
(2)	Oxidation	1478–1496	19	142	322
(3)	Reduction	1785–1826	42	–197	–287
(4)	Oxidation	1871–1888	18	162	333
(5)	Reduction	2173–2219	47	–168	–268
(6)	Oxidation	2263–2279	17	172	262
(7)	Reduction	2563–2612	50	–158	–248
(8)	Oxidation	2655–2670	16	182	333
(9)	Reduction	2955–3002	48	–168	–228
(10)	Oxidation	3048–3062	15	202	343
(11)	Reduction	3345–3395	51	–158	–238
(12)	Oxidation	3441–3454	14	222	353
(13)	Reduction	3735–3786	52	–148	–238
(14)	Oxidation	3834–3846	13	242	363

reliability criterion is larger than the values obtained for the random variables. The appended artificial random variables are multiplied by a small constant (1×10^{-10}) in order to have a negligible influence on the model while retaining the same random variation within artificial variables. The goal of UVE-PLS and UVE-PCR is not variable selection in the sense that one tries to find the best subset of variables for fitting or prediction of a model but elimination of those variables that are useless. Therefore, UVE-PLS (UVE-PCR) is not an equivalent but a supplementary technique to the coupled LSR and SIMCA modeling power.

Because the latent variables are liner combinations of the original ones, the PLS and PCR models can be expressed as

$$\mathbf{y} = \mathbf{Z} \times \mathbf{b}^* + \mathbf{e}, \quad (4)$$

where \mathbf{b}^* is the regression vector described for the PCR by the following equation:

$$\mathbf{b}^* = \mathbf{B} \times \boldsymbol{\beta}, \quad (5)$$

where $\boldsymbol{\beta}$ is the vector of Inverse Least Squares (ILS) regression coefficients for PCR:

$$\boldsymbol{\beta} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}. \quad (6)$$

For each j th variable the reliability regression criterion for UVE-PCR [23] and UVE-PLS [21, 22] is defined as

$$c_j = \frac{\bar{b}_j^*}{s(b_j^*)}, \quad (7)$$

where \bar{b}_j^* is a mean and $s(b_j^*)$ is a standard deviation from the vector of I b_{ij}^* coefficients obtained by (leave-one-out) jackknifing ($i = 1, \dots, I$).

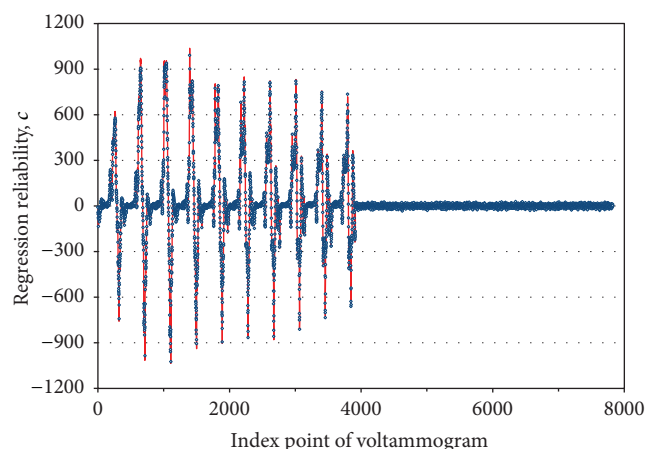


FIGURE 4: Regression reliability, c (see (7)), calculated by UVE-PLS (blue diamonds) and UVE-PCR (red solid line) (both for $F = 1$) for $J = 3911$ points of mean centered voltammogram augmented by J artificial random variables.

The UVE-PLS and UVE-PCR (number of latent variables for decomposition, $F = 1$) were applied to the all J points of the voltammograms of Figure 1 and augmented J variables corresponding to random noise data. The results obtained for regression reliability are shown in Figure 4. The performance of UVE-PLS and UVE-PCR is equivalent.

For each variable j , the greater the absolute value of regression reliability c , the more informative the j th variable for the model. By visual comparison of data in Figures 2 and 4, it can be seen that the elevated absolute values of regression reliability overlap with modeling power and r^2 approaching unity. The cut off criterion is determined by the maximum absolute value of regression reliability calculated

TABLE 5: Extreme values of parameters calculated for individually selected ranges of points of staircase voltammogram, variables j , (Table 4) using UVE-PLS, UVE-PCR, PLS-VIP, IPLS, IPCR, and MWPLS for $F = 1$.

Data range index, k	Range of voltammogram/ variables j	UVE-PLS Min Abs(c)	UVE-PCR Min Abs(c)	Min PLS-VIP	IPLS Max RMSECV	IPCR Max RMSECV	MWPLS Max RMSECV
(1)	1401–1430	615	606	1.47	.350	.350	.313
(2)	1478–1496	331	334	1.47	.356	.357	.334
(3)	1785–1826	424	423	1.47	.245	.245	.268
(4)	1871–1888	281	281	1.47	.274	.275	.301
(5)	2173–2219	335	334	1.47	.224	.225	.237
(6)	2263–2279	250	251	1.47	.276	.278	.276
(7)	2563–2612	285	285	1.48	.286	.286	.213
(8)	2655–2670	231	232	1.48	.327	.328	.252
(9)	2955–3002	237	241	1.48	.131	.131	.174
(10)	3048–3062	225	227	1.47	.221	.223	.232
(11)	3345–3395	230	230	1.49	.162	.162	.162
(12)	3441–3454	224	226	1.47	.227	.229	.214
(13)	3735–3786	222	222	1.49	.237	.237	.142
(14)	3834–3846	225	226	1.47	.317	.317	.198

for the random variables (right half of the Figure 4) which are 32.6 and 31.7 for UVE-PLS and UVE-PCR, respectively.

Table 5 lists in columns three and four minimum absolute values of regression reliability calculated for individually selected ranges of variables j (Table 4), for UVE-PLS and UVE-PCR, respectively. The same minimum value of 222 was obtained for range of $k = 13$ with both methods, UVE-PLS and UVE-PCR. As this value lays significantly above the cut off level, both methods UVE-PLS and UVE-PCR confirm the correctness of choosing the ranges of Table 4 for building of an analytical model.

3.2.3. *Variable Importance in the Projection.* PLS-VIP is a combined measure of how much a variable contributes to describe the two sets of data: the dependent (\mathbf{y}) and the independent variables (\mathbf{X}) [24, 25], which in the studied case correspond to concentration of suppressor and voltammetric current values, respectively. For the number of factors $F = 1$, the expression describing VIP value for the j th variable is given by

$$VIP_j = \sqrt{J \sum_{j=1}^J \left(\frac{w_j}{\|\mathbf{w}\|} \right)^2}, \quad (8)$$

where w_j is the PLS weight value for variable j and $\|\mathbf{w}\|$ is the Frobenius norm of the weight vector \mathbf{w} defined by the following expression:

$$\|\mathbf{w}\| = \sqrt{\sum_{j=1}^J w_j^2}. \quad (9)$$

The weights in a PLS model reflect the covariance between the independent and dependent variables and the

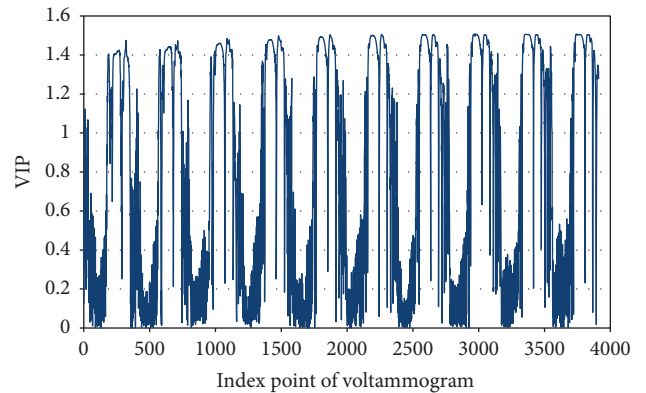


FIGURE 5: Variable Importance in the Projection by PLS (PLS-VIP) (see (8)) calculated for $J = 3911$ points of autoscaled voltammograms for $F = 1$.

inclusion of the weights is what allows VIP to reflect not only how well the dependent variable is described but also how important that information is for the model of independent variables. Since the average of squared VIP scores equals unity, values smaller than one indicate nonimportant variables.

The PLS-VIP (number of latent variables for decomposition, $F = 1$) was applied to all points of the autoscaled voltammograms of Figure 1 and autoscaled dependent variables \mathbf{y} (concentrations). The results obtained for VIP are shown in Figure 5 and Table 5. All index points of the voltammogram (variables j) preselected with modeling power and LSR (Table 4) are also of highest importance in the projection based on PLS-VIP analysis, with the minimum value of VIP of 1.47 obtained for numerous individually selected ranges (Table 5). As this value lies significantly above the cut off level of one, the application of PLS-VIP also confirms the

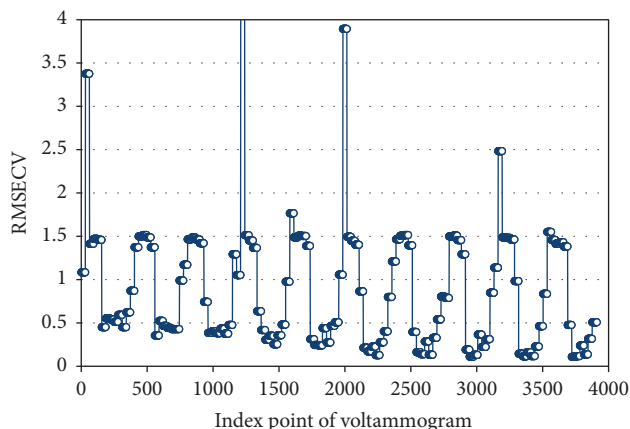


FIGURE 6: RMSECV calculated with IPLS for 125 nonoverlapping intervals of width of 31 variables and 1 interval of 36 variables of mean centered voltammograms for $F = 1$.

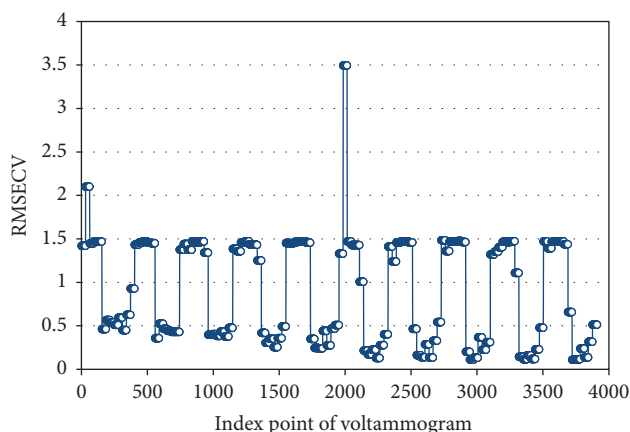


FIGURE 7: RMSECV calculated with IPCR for 125 nonoverlapping intervals of width of 31 variables and 1 interval of 36 variables of mean centered voltammograms for $F = 1$.

correctness of choosing the ranges of Table 4 for building of an analytical model.

3.2.4. Interval Partial Least Squares and Interval Principal Component Regressions. In IPLS [26] and IPCR the voltammograms are divided into a number of nonoverlapping, consecutive intervals and PLS and PCR models are developed for each of these intervals. The aim is to find intervals which give better predictions than the predictions obtained when using the full voltammogram. The comparison of interval performance is mainly based on the Root Mean Square Error of Cross-Validation (RMSECV). The RMSECV values obtained for the training set voltammograms with IPLS and IPCR for the arbitrarily chosen interval of the width of 31 points of the voltammograms (variables j) is presented in Figures 6 and 7, respectively. The points of the voltammograms were divided into 125 nonoverlapping consecutive intervals and an addition last interval with a width of 36 to accommodate all $J = 3911$ index points of voltammogram ($125 \times 31 + 36 = 3911$).

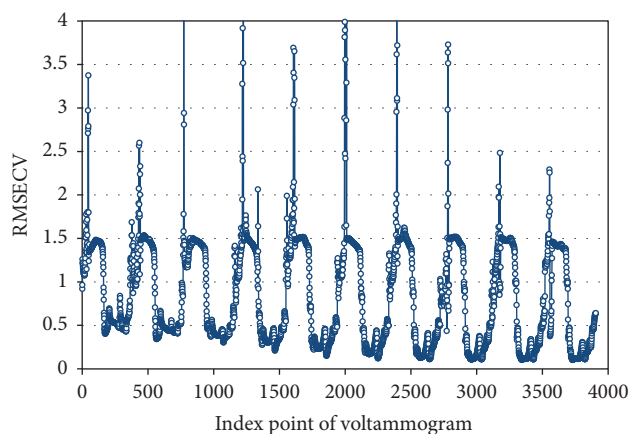


FIGURE 8: RMSECV calculated with MWPCR for a moving one-step-at-a-time interval of 31 points of mean centered voltammograms for $F = 1$.

It can be seen in Table 5 that the ranges of points selected for calibration for the voltammograms in Table 4 correspond to the lowest values of RMSECV in Figures 6 and 7. However, there are two problems with using IPLS and IPCR for variable selection related to the specificity of voltammetric data of Figure 1. First, some of the selected ranges in Table 4 (especially that for oxidation processes) are very narrow ($k = 14$, $J = 13$, 0.33% of total number of variables J). The second problem is related to the fact that these narrow ranges are neighboring points of voltammograms of very poor correlation and have no utility to the model (Figures 2, 4, and 5). These two facts may lead to a situation where there is insufficient representation of “proper” variables in a selected interval to obtain a RMSECV value lower than that obtained for cross-validation of entire voltammogram.

In order to address the issues outlined in the paragraph above, MWPLS was implemented. This technique allows a sufficient representation of “proper” variables neighboring on both sides, the variable for which the RMSECV is calculated. The results obtained with MWPLS are presented in Figure 8. One can notice in Table 5 that even the maximum values of RMSECV obtained with the MWPLS for the individually selected ranges of Table 4 correspond to the minor values of the graph of Figure 8. All three methods, IPLS, IPCR, and MWPLS, confirm the selection ranges in Table 4.

There must be an agreement about the outcome of the selection of variables for all methods presented (LSR, two-step SIMCA modeling power, UVE-PLS (UVE-PCR), PLS-VIP, IPLS, IPCR, and MWPLS) in order to accept these variables for further model development. As Andersen and Bro [31] pointed out, it is always instructive to compare the results from several types of variable selection to assess whether they complement each other.

3.3. Selection of the Most Suitable Data Decomposition Method. Multivariate data decomposition techniques are employed in chemometrics in order to compress vast amount of data while extracting the significant information. The techniques commonly used can be divided into three major groups:

(i) two-way techniques: PCA and PLS; (ii) hierarchical and multiblock techniques: HPCA [19], HPLS [19], CPCA [19], and MBPLS [19]; and (iii) multiway techniques: Generalized Rank Annihilation Method (GRAM) [32], PARAllel FACtor Analysis (PARAFAC) [32], Multilinear PLS (N-PLS) [32], and Tucker models [32]. There is also a hybrid method called Multiway PCA (MPCA) [33] described as “poor man’s” multiway technique [34] as it relies upon a two-way PCA and rearrangement by unfolding the original multiway data.

The structure of the voltammetric data of the training set to be decomposed is described in Table 4. It consists of fourteen ($K = 14$) blocks, each of them containing $I = 84$ samples. The number of variables of each block J varies from 13 to 52 depending on the block number k . As the data consists of several blocks, the two-way methods, PCA and PLS, are not applicable. By extensive selection, this data could be arranged into a three-way array of the dimensions ($I \times J_{\min} \times K$), where J_{\min} would correspond to that of the narrowest ($k = 14$) range variables $J_{k=14} = 13$. For instance, in order to fit such a three-way array every fourth variable would need to be extracted from the widest range of $J = 52$ points. Such a selection would mean a loss of 75% of the data for the widest range already at the preliminary stage of the data arrangement (prior to the actual analysis).

Qualitatively, the data of Table 4 can be divided into two distinct groups, each consisting of seven ranges. The division is based on the two electrode processes investigated: reduction and oxidation. Although quantitatively different (Figure 1), the qualitative difference among the data of various blocks investigating, for instance, oxidation, may not be predominant enough to consider these groups as members of different slab, a layer (submatrix) of a three-way array. Application of multiway techniques for the higher number of factors for insufficiently dissimilar slabs may result in some factors to be highly correlated.

One may consider application of MPCA for decomposition of the multiblock training data of Table 4 following their unfolding into two-way matrix \mathbf{X} of the dimension ($I \times \sum_{k=1}^K J_k$). However, even after columnwise autoscaling, the variables will not be equally (or controllably) represented. The data of the wider blocks will have a greater representation than those of the narrower ones.

The hierarchical and multiblock data decomposition techniques [12, 19] calculate for each of K blocks the block scores and block loadings using regular two-way decomposition techniques. Subsequently, all block scores are combined into the super scores prior to the regression. The hierarchical and multiblock data decomposition techniques are the most suitable to be employed for treatment of voltammetric data defined in Table 4. Application of these techniques to the voltammetric data for suppressor avoids making unnecessary compromises or risking the accuracy and soundness of the data representation, which otherwise may be needed when arranging data in three-way arrays. The scaling employed in this paper provides equal representation of each of $K = 14$ blocks, regardless of the width of that blocks (J_k). Before applying CPCA, HPCA, HPLS, and MBPLS, the data is block-scaled by dividing by the square root of the number of

variables in the block in order to equalize the representation of various blocks [19]:

$$\mathbf{X}_k = \frac{\mathbf{X}_k^0}{\sqrt{J_k}}. \quad (10)$$

Westerhuis et al. [19] derived an algebraic proof of the equivalence of the CPCA and MBPLS to the standard PCA and PLS, respectively, when the same variable scaling was applied for these methods. Westerhuis et al. [19] recommended utilizing the standard PCA and PLS methods which require less computation and give a better estimation for the scores in cases where there are missing data because correlations among all variables can be used for estimation of the scores instead of only the correlations among the variables in the specific block.

3.4. Validation of Analytical Models. Calibration calculations were conducted using comparatively CPCR, HPCR, HPLS, and MBPLS methods for the training set data ($I = 84$ samples). The predictive ability of the calibration models was verified on the external validation set containing voltammograms recorded for $I_V = 27$ different samples. Prior to being projected on the calibration model, the validation set was scaled using the scaling parameters of the training set. The optimum number of factors (F) was determined for each of the methods based on the analysis of Predictive Residual Sum of Squares (PRESS) values calculated for the residual concentrations of the external validation set employing the following expression:

$$\text{PRESS}_f = \sum_{i=1}^{I_V} (y_{V,i}^O - \hat{y}_{V,f,i}^O)^2, \quad (11)$$

where $y_{V,i}^O$ describes the original (not centered) actual concentration of i th sample in the validation set, while $\hat{y}_{V,f,i}^O$ denotes the rescaled predicted concentration using f -factor decomposition of i th sample of the validation set. The results obtained are presented in Table 6.

Additionally, the predictive performance of the CPCR, HPCR, HPLS, and MBPLS analytical models was comparatively assessed by determining the values of the following two derived parameters: mean no-sign relative-to-target error of prediction and mean no-sign relative error of prediction (Table 6). The mean no-sign relative-to-target error of prediction (MNSRTEP) is defined by the equation

$$\text{MNSRTEP} = \sum_{i=1}^{I_V} \frac{|\hat{y}_{V,i}^O - y_{V,i}^O|}{I_V y_t^O} \times 100\%, \quad (12)$$

where y_t^O is the target-level concentration. In the case of the specific bath composition in this investigation the target concentration of accelerator was 7.00 mL L^{-1} .

The mean no-sign relative error of prediction (MNSREP) is described by the following equation:

$$\text{MNSREP} = \sum_{i=1}^{I_V} \frac{|\hat{y}_{V,i}^O - y_{V,i}^O|}{I_V y_{V,i}^O} \times 100\%. \quad (13)$$

TABLE 6: r^2 , PRESS, MNSRTEP, and MNSREP calculated for external validation set for various calibration models for different number of factors. Selected optimum parameters are bolded.

Decomposition and regression Technique	# of factors, F (latent variables)	r^2	PRESS	MNSRTEP	MNSREP
CPCR	1	0.9948	0.5475	1.6381	1.7168
CPCR	2	0.9971	0.1972	1.0765	1.0820
CPCR	3	0.9971	0.1859	1.0152	1.0019
HPCR	1	0.9948	0.5470	1.6375	1.7161
HPCR	2	0.9972	0.1969	1.0761	1.0779
HPCR	3	0.9972	0.1892	1.0393	1.0302
HPLS	1	0.9954	0.5056	1.6003	1.6746
HPLS	2	0.9978	0.1855	1.0824	1.1048
HPLS	3	0.9967	0.3397	1.4312	1.5598
MBPLS	1	0.9949	0.5423	1.6323	1.7103
MBPLS	2	0.9969	0.2039	1.0997	1.1111
MBPLS	3	0.9890	0.6846	1.8698	2.0860

While examining the results in Table 6, one can see based on PRESS, MNSRTEP, and MNSREP values that the vast majority of the total variance has already been captured by the first factor. The parameters r^2 , PRESS, MNSRTEP, and MNSREP have practically the same values for $F = 1$ for all three techniques: CPCR, HPCR, and MBPLS. Although the values obtained for HPLS are different than that for the other methods, the difference is not significant. The introduction of a second factor/latent variable introduces slight improvement of the values of parameters PRESS, MNSRTEP, and MNSREP, along with improvement of correlation. The incorporation of the third factor/latent variable into the decomposition worsens the values of all parameters: r^2 , PRESS, MNSRTEP, and MNSREP for MBPLS and HPLS. Usually, the optimum number of factors/latent variables should be determined based on the first local minimum of the $PRESS_f$ value. In the specific case discussed, the improvement of the (already good) values of parameters PRESS, MNSRTEP, and MNSREP by the addition of second factor justifies using of number of factors $F = 2$ for final tuning of the analytical model. Therefore, our conclusion is to employ two factors/latent variables for building of the analytical model, regardless of the chosen chemometric technique for data decomposition. The selected optimum parameters are presented in bold in Table 6.

The performance of all four techniques is equivalent to the exception of the HPLS and MBPLS results for $F = 3$, which are worse than that of the other methods. On the other hand, the performances of HPLS and MBPLS for $F = 3$ for take-one-out cross-validatory calculation within the training set (Table 7) are superior (especially for MBPLS) as compared to other methods for the same number of factors/latent variables used for decomposition. Usually, a combination of better cross-validatory (or self-prediction) performance with worse predictive performance of external data is evidence of overfitting of the model. The PLS-based models tend to be more overfit than the PCA-based ones. Again, by keeping the number of factors low, one can significantly reduce the

TABLE 7: r^2 and PRESS results for take-one-out cross-validation within the training set for $F = 3$.

Decomposition and regression Technique	# of factors, F (latent variables)	r^2	PRESS
CPCR	3	0.9929	1.2435
HPCR	3	0.9929	1.2375
HPLS	3	0.9938	1.0798
MBPLS	3	0.9964	0.6294

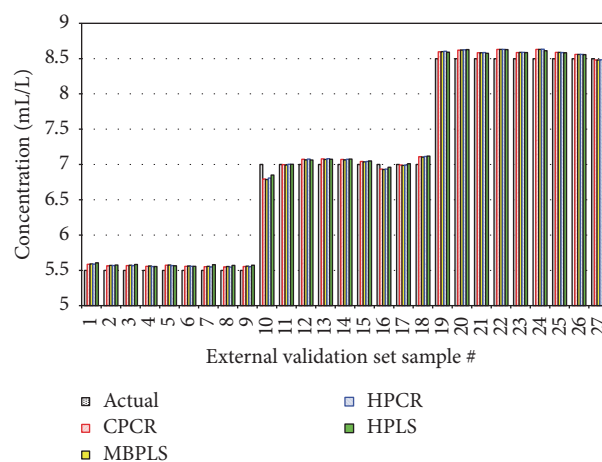


FIGURE 9: Actual (black) and predicted with CPCR (red), MBPLS (yellow), HPCR (blue), and HPLS (green) concentrations of suppressor for external validation set for $F = 2$.

probability of such occurrence (see the practically equivalent performance of all techniques for $F = 1$ and for $F = 2$ in Table 6).

Figure 9 presents a comparison of the actual concentrations of suppressor with those predicted by CPCR, MBPLS, HPCR, and HPLS for the external validation set of 27 samples for $F = 2$ demonstrating equivalent performance

of all techniques and providing evidence for the absence of artifacts in the selected voltammetric data.

4. Conclusions

This paper has introduced several variable selection techniques (UVE-PLS, UVE-PCR, PLS-VIP, IPLS, IPCR, and MWPLS) to the field of electroanalysis which are well established for analysis of spectroscopic data. Specifically, the rigorous, multistep procedure of selecting the blocks of the voltammogram to be used subsequently for analytical model development based on LSR, two-step SIMCA modeling power, UVE-PLS, UVE-PCR, PLS-VIP, MWPLS, IPLS, and IPCR focused on individual blocks was introduced. To ensure feasibility of the model, several variable selection methods were utilized comparatively to verify that their results do not contradict each other. Detailed criteria for choosing the decomposition technique proper for the data in order to compress the multivariate data and reasonably extract information were presented. The optimization of the number of factors based on external validation and cross-validation was presented. As a general recommendation, a few variable selection methods need to be implemented concurrently for the investigated multivariate data set, as their consistent-to-each-other performance provides an evidence for an absence of artifacts in that data. All the methodology for data selection can be automated and ultimately can lead to a fully automatic data selection process that would reduce the time and effort of method development significantly.

Electroanalytical chemists would substantially benefit from utilization of the variable selection methods allowing them to focus only on relevant portions of voltammetric responses while eliminating those uninformative. Otherwise, the incorporation of the irrelevant variables (both random and systematic) into a multivariate model leads to less precision (higher variance due to imbedded error). By analogy to other disciplines of analytical chemistry, variable selection should become a routine step of multivariate data pretreatment in electroanalytical chemistry by utilizing existing chemometric techniques. This paper proves that existing variable selection methods can be transferred to electroanalytical data.

Abbreviations

ASV:	Anodic Stripping Voltammetry
CPCR:	Consensus Principal Component Regression
CV:	Cyclic Voltammetry
GRAM:	Generalized Rank Annihilation Method
HPCR:	Hierarchical Principal Component Regression
HPLS:	Hierarchical Partial Least Squares
IPCR:	Interval Principal Component Regression
IPLS:	Interval Partial Least Squares
LSR:	Least Squares Regression
MBPLS:	Multiblock Partial Least Squares
MNSREP:	Mean No-Sign Relative Error of Prediction
MNSRTEP:	Mean No-Sign Relative-to-Target Error of Prediction
MPCA:	Multiway Principal Component Analysis

MTEP:	Multi-Task Electrochemical Probe
MWPLS:	Moving Window Partial Least Squares
N-PLS:	Multilinear Partial Least Squares
PARAFAC:	Parallel Factor Analysis
PCA:	Principal Component Analysis
PCR:	Principal Component Regression
PLS:	Partial Least Squares
PLS-VIP:	Variable Importance for Prediction by Partial Least Squares
PRESS:	Predictive Residual Sum of Squares
RMSECV:	Root Mean Square Error of Cross-Validation
SIMCA:	Simple Modeling of Class Analogy
UVE-PCR:	Uninformative Variable Elimination by Principal Component Regression
UVE-PLS:	Uninformative Variable Elimination by Partial Least Squares.

Conflicts of Interest

Dr. Aleksander Jaworski, Dr. Hanna Wkiel, and Dr. Kazimierz Wkiel declare that there are no conflicts of interest regarding the publication of the paper.

Acknowledgments

The authors would like to thank Dr. Allan Reed for his comments on the manuscript.

References

- [1] P. C. Andricacos, C. Uzoh, J. O. Dukovic, J. Horkans, and H. Deligianni, "Damascene copper electroplating for chip interconnections," *IBM Journal of Research and Development*, vol. 42, no. 5, pp. 567–573, 1998.
- [2] J. P. Healy, D. Pletcher, and M. Goodenough, "The chemistry of the additives in an acid copper electroplating bath. Part I. Polyethylene glycol and chloride ion," *Journal of Electroanalytical Chemistry*, vol. 338, no. 1-2, pp. 155–165, 1992.
- [3] J. G. Long, P. C. Searson, and P. M. Vereecken, "Electrochemical characterization of adsorption-desorption of the cuprous-suppressor-chloride complex during electrodeposition of copper," *Journal of The Electrochemical Society*, vol. 153, no. 4, pp. C258–C264, 2006.
- [4] J. P. Healy, D. Pletcher, and M. Goodenough, "The chemistry of the additives in an acid copper electroplating bath. Part II. The instability 4,5-dithiaoctane-1,8-disulphonic acid in the bath on open circuit," *Journal of Electroanalytical Chemistry*, vol. 338, no. 1-2, pp. 167–177, 1992.
- [5] J. P. Healy, D. Pletcher, and M. Goodenough, "The chemistry of the additives in an acid copper electroplating bath. Part III. The mechanism of brightening by 4,5-dithia-octane-1, 8-disulphonic acid," *Journal of Electroanalytical Chemistry*, vol. 338, no. 1-2, pp. 179–187, 1992.
- [6] T. P. Moffat, D. Wheeler, and D. Josell, "Electrodeposition of copper in the SPS-PEG-CL additive system I. Kinetic measurements: influence of SPS," *Journal of The Electrochemical Society*, vol. 151, no. 4, pp. C262–C271, 2004.
- [7] T. P. Moffat, D. Wheeler, S.-K. Kim, and D. Josell, "Curvature enhanced adsorbate coverage model for electrodeposition,"

- Journal of The Electrochemical Society*, vol. 153, no. 2, pp. C127–C132, 2006.
- [8] K. Dietz, “Organic additives in copper plating bath (part I),” *CircuTree*, vol. 13, pp. 22–26, 2000.
- [9] K. Dietz, “Organic additives in copper plating bath (part II),” *CircuTree*, vol. 13, pp. 54–58, 2000.
- [10] Y. Ni and S. Kokot, “Does chemometrics enhance the performance of electroanalysis?” *Analytica Chimica Acta*, vol. 626, no. 2, pp. 130–146, 2008.
- [11] K. Wikiel, A. Jaworski, H. Wikiel, and D. Hazebrouck, “Method and apparatus for real time monitoring of industrial electrolytes,” *The US Patent and Trademark Office*, vol. 7, pp. 270–733, 2007.
- [12] A. Jaworski, K. Wikiel, and H. Wikiel, “Application of multi-block and hierarchical PCA and PLS models for analysis of AC voltammetric data,” *Electroanalysis*, vol. 17, no. 15–16, pp. 1477–1485, 2005.
- [13] A. Jaworski, H. Wikiel, and K. Wikiel, “Copper ECD process control by means of electroanalysis combined with multi-way data decomposition chemometric techniques,” *ECS Transactions*, vol. 6, no. 19, pp. 39–52, 2008.
- [14] A. Jaworski, H. Wikiel, and K. Wikiel, “Application of multiway chemometric techniques for analysis of AC voltammetric data,” *Electroanalysis*, vol. 21, no. 3–5, pp. 580–589, 2009.
- [15] A. Jaworski, H. Wikiel, and K. Wikiel, “Copper ECD process of ULSI circuits controlled by electroanalysis combined with multi-way chemometrics. Transfer of the PARAFAC/ILS calibration,” *ECS Transactions*, vol. 16, no. 22, pp. 11–26, 2009.
- [16] A. Jaworski, H. Wikiel, and K. Wikiel, “Multi-way standardization of an AC voltammetric analyzer for electrometallization baths,” *Analytica Chimica Acta*, vol. 656, no. 1–2, pp. 42–50, 2009.
- [17] A. Jaworski, H. Wikiel, and K. Wikiel, “Method and apparatus for real time monitoring of electroplating bath performance and early fault detection,” *US Patent*, 2007.
- [18] A. Jaworski, H. Wikiel, and K. Wikiel, “Controlling copper ECD processes by early fault detection and diagnosis using in-situ electrochemical sensor coupled with pattern recognition chemometrics,” *ECS Trans*, vol. 25, no. 3, pp. 199–217, 2009.
- [19] J. Westerhuis, T. Kourti, J. MacGregor, and J. Chemom, “Analysis of multiblock and hierarchical PCA and PLS models,” *Journal of Chemometrics*, vol. 12, no. 5, pp. 301–321, 1998.
- [20] M. Sharaf, D. Illman, and B. Kowalski, “Chemometrics,” in *Chemical Analysis. A Series of Monographs on Analytical Chemistry and Its Applications*, P. J. Elving and J. D. Winefordner, Eds., vol. 82, Wiley, New York, NY, USA, 1986.
- [21] V. Centner, D.-L. Massart, O. E. de Noord, S. de Jong, B. M. Vandeginste, and C. Sterna, “Elimination of uninformative variables for multivariate calibration,” *Analytical Chemistry*, vol. 68, no. 21, pp. 3851–3858, 1996.
- [22] V. Centner, “Multivariate Approaches: UVE-PLS,” in *Comprehensive Chemometrics*, S. Brown, R. Tauler, and B. Walczak, Eds., Elsevier, Amsterdam, Netherlands, 2009.
- [23] V. Centner, J. Verdú-Andrés, B. Walczak et al., “Comparison of multivariate calibration techniques applied to experimental NIR data sets,” *Applied Spectroscopy*, vol. 54, no. 4, pp. 608–623, 2000.
- [24] S. Wold, E. Johansson, and M. Cocchi, *3D QSAR in Drug Design; Theory, Methods and Applications*, ESCOM, Leiden, The Netherlands, 1993.
- [25] I. G. Chong and C. H. Jun, “Performance of some variable selection methods when multicollinearity is present,” *Chemometrics and Intelligent Laboratory Systems*, vol. 78, pp. 103–112, 2005.
- [26] L. Nørgaard, A. Saudland, J. Wagner, J. P. Nielsen, L. Munck, and S. B. Engelsen, “Interval partial least-squares regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy,” *Applied Spectroscopy*, vol. 54, no. 3, pp. 413–419, 2000.
- [27] J.-H. Jiang, R. J. James, H. W. Siesler, and Y. Ozaki, “Wavelength interval selection in multicomponent spectral analysis by moving window partial least-squares regression with applications to mid-infrared and near-infrared spectroscopic data,” *Analytical Chemistry*, vol. 74, no. 14, pp. 3555–3565, 2002.
- [28] L. Pinto and S. G. Lemos, “Comparison of different PLS algorithms for simultaneous determination of Cd(II), Cu(II), Pb(II), and Zn(II) by anodic stripping voltammetry at bismuth film electrode,” *Electroanalysis*, vol. 26, no. 2, pp. 299–305, 2014.
- [29] P. D. Marreto, A. M. Zimer, R. C. Faria et al., “Multivariate linear regression with variable selection by a successive projections algorithm applied to the analysis of anodic stripping voltammetry data,” *Electrochimica Acta*, vol. 127, pp. 68–78, 2014.
- [30] A. Herrero and M. Cruz Ortiz, “Qualitative and quantitative aspects of the application of genetic algorithm-based variable selection in polarography and stripping voltammetry,” *Analytica Chimica Acta*, vol. 378, no. 1–3, pp. 245–259, 1999.
- [31] C. M. Andersen and R. Bro, “Variable selection in regression—a tutorial,” *Journal of Chemometrics*, vol. 24, no. 11–12, pp. 728–737, 2010.
- [32] A. Smilde, R. Bro, and P. Geladi, *Multi-Way Analysis in Chemistry*, Wiley, Chichester, UK, 2004.
- [33] S. Wold, P. Geladi, K. Esbensen, and J. Öhman, “Multi-way principal components-and PLS-analysis,” *Journal of Chemometrics*, vol. 1, no. 1, pp. 41–56, 1987.
- [34] B. M. Wise, N. B. Gallagher, S. W. Butler, D. D. White Jr., and G. G. Barna, “A comparison of principal component analysis, multiway principal component analysis, trilinear decomposition and parallel factor analysis for fault detection in a semiconductor etch process,” *Journal of Chemometrics*, vol. 13, no. 3–4, pp. 379–396, 1999.

