Simultaneous Spectrophotometric Determination of Heavy Metal Ions Using Several Chemometrics Methods: Effect of Different Parameters of Savitzky-Golay and Direct Orthogonal Signal Correction Filters

Rouhollahi, Ahmad; Tavakoli, Hossein

Department of Chemistry, Faculty of Sciences, K. N. Toosi University of Technology, Tehran, I. R. IRAN

Nayebi, Shahzad; Ghasemi, Jahanbakhsh*[+]

Department of Chemistry, Faculty of Sciences, Razi University, Kermanshah, I. R. IRAN

Noroozi, Mohammad; Hashemi, Majid

Researches and Petroleum Engineering Center of Kermanshah, Kermanshah, I. R. IRAN

ABSTRACT: *Direct orthogonal signal correction(DOSC) and Savitzky-Golay filters (SGF) were applied as preprocessing methods on the original and first derivative absorbance data. Principle component regression (PCR), partial least squares (PLS) and iterative target transformation factor analysis (ITTFA), were used in spectrophotometric simultaneous determination of heavy divalent metal ions, lead, zinc, mercury and cadmium, using 4-(2-pyridylazo) resorcinol (PAR) as metallochromic indicator. The optimum values of the parameters for DOSC and SGF were obtained according to REP, R^2 and RMSEP functions for calibration and prediction sets. The concentration for Hg^{2+}, Pb^{2+}, Zn^{2+} and Cd^{2+} ions in calibration set were varied between 0 - 12.24, 0 - 9.81, 0 - 0.87 and 0 - 3.96 ppm, respectively. The experimental calibration set was composed of 35 sample solutions and the 7 solutions as prediction set using a simple lattice (4,4) mixture design. The absorption spectra were recorded from 450 to 600 nm and absorbance data were autoscaled. The effect of pH on the sensitivity and selectivity was studied in the range of 1.00 - 11.00 and pH=7.50 was chosen according to net analyte signal (NAS) as a function of pH.*

* *To whom correspondence should be addressed.*
+ *E-mail: jahan.ghasemi@gmail.com*

INTRODUCTION

Several techniques such as polarography [1], mass spectrometry [2], inductively coupled plasma atomic emission [3], voltammetry [4], liquid chromatography [5], etc. have been used for the simultaneous determination of lead, zinc, mercury and cadmium in aqueous media. The simultaneous determination of these ions by the common UV-visible spectrometry techniques and conventional metallochromic indicators in aqueous solution is difficult because, generally, the absorption spectra overlap in this region and the superimposed curves are not suitable for quantitative evaluation.

Under computer-controlled instrumentation, derivative techniques and multivariate calibration methods are playing very important roles in the multicomponent analysis of mixtures by ultraviolet (UV)-visible molecular absorption spectrophotometry [6,7]. The main advantage of multicomponent analysis using multivariate calibration is the speed of the method of determination for components of interest in a mixture, as the separation step could be avoided.

THEORY

Multivariate Calibration Methods

Nowadays quantitative spectrophotometry has been greatly improved by the use of a variety of multivariate statistical methods; particularly principle component regression (PCR), partial least squares regression (PLS) and iterative target transformation factor analysis (ITTFA). The theory and application of these methods in spectroscopy have been discussed by several workers [8-14] and here we describe them briefly.

Principal Component Regression (PCR)

Principal component regression consists of two steps; in the first step, as is in principal component analysis (PCA), it does decomposition of spectral matrix into a set of eigenvectors, loadings and scores, and then regress the obtained scores against the concentrations as a calibration step (as we do in inverse least squares) [9].

Partial Least Squares (PLS)

Partial least squares modeling is one of the most powerful multivariate statistical tools that has been applied to the quantitative analysis [11-13]. PLS is a quantitative spectral decomposition technique that is closely related to principal component regression (PCR). However, in PLS the decomposition is performed in a slightly different fashion. PLS actually uses the concentration information during the decomposition process. Thus, the eigenvectors and scores calculated using PLS is quite different from those of PCR. The main idea of PLS is to get as much concentration information as possible into the first few loading and score vectors [8, 9].

Iterative Target Transformation Factor Analysis (ITTFA)

In the absence of good candidate targets to be tested by target transformation factor analysis (TTFA), one defines an initial target, which is gradually improved until the TTFA test passes. The target is modified by using chemical knowledge (e.g. in our study non-negativity) and resubmitted in an iterative way until the tested target is considered to match one of the true factors satisfactorily. In other word, ITTFA can give a priori estimates of the transformation matrix to rotate the abstract factor analysis solutions to good approximations of real peak profiles when there are no data points unique to the individual components [14].

Preprocessing
Direct Orthogonal Signal Correction (DOSC)

Orthogonal signal correction (OSC) [15] is generally applied to spectroscopic data (X) to remove from the spectra as much as possible the variation that is unrelated (i.e. orthogonal) to y, the vector of the parameter which has to be modeled (e.g. the concentration). After the OSC correction, a new PLS or PCR model can be built and this model will be less complex than the model built with the original, uncorrected data. Algorithms to perform an orthogonal signal correction have been proposed by *Wold* [16], *Sjoblom* [17], *Wise* and *Gallagher* [18] and *Fearn* [19].

A similar method called direct orthogonalization has been developed by *Andersson* [20]. All of these methods find an approximate solution to the problem set out, i.e. finding a subspace of X that is orthogonal to y and accounts for the largest possible proportion of X-variance. Recently, the exact solution was found independently by *Westerhuis et al.* [15]. The corresponding algorithm was coined direct orthogonal signal correction (DOSC). The first step in this algorithm is a decomposition of X (the

spectral matrix) into two orthogonal parts, one part related to **y** and another part that is orthogonal to it. This is carried out by projecting (or regressing) y onto X. In this way, one decomposes y into $\hat{y}$; the part of y lying in X-space and f , the residual that is unrelated to X, i.e.

$$y = \hat{y} + f \tag{1}$$

For spectral data, X is generally of less than full column rank. The column rank of a matrix is the number of independent columns (wavelengths). Since in UV-vis spectra, neighboring wavelengths are not independent, but highly correlated, the rank of a spectral matrix is less than its number of columns.

Next, X is projected onto $\hat{y}$ giving $\hat{X}$ and E, the residual part of X that is orthogonal both to $\hat{y}$ and y, i.e.

$$X = \hat{X} + E \tag{2}$$

Principal component analysis (PCA) or singular value decomposition (SVD) is applied to E in order to find a small number of principal components T corresponding to the largest singular values. This T is a basis for the low-dimensional subspace that accounts for the maximum of variance of E, the part of X that is unrelated to y.

The DOSC-corrected spectra of the calibration data can now be written as:

$$X_{DOSC} = X - TP' \tag{3}$$

where TP' is the 'orthogonal' part removed from the original spectra with P the loading matrix:

$$P = XT(T'T)^{-1} \tag{4}$$

T can be expressed as linear combinations of X:

$$T = XR \tag{5}$$

Here, R is the matrix of weights of the original variables in the principal orthogonal directions, which can be obtained via X^+, the Moore_ Penrose generalized inverse of the original data **X:**

$$R = X^+T \tag{6}$$

Given weights R and loadings P, one can directly obtain corrected spectra for new data:

$$X_{DOSC} = X - XRP' \tag{7}$$

Derivation and Smoothing Using the Savitzky-Golay Filter

By calculating first or higher order of derivatives, baseline drifts are eliminated and also small spectral differences are enhanced. These derivations are performed by using the Savitzky-Golay filter, which is a moving window averaging method: a window is selected where the data are fitted by a polynomial of a certain degree. The central point in the window is replaced by a linear combination of itself and some number of nearby neighbors [7, 22, 23].

Selection of optimum pH

The effect of pH on the sensitivity and selectivity was studied according to the net analyte signal (NAS, a portion of the signal of each component that is orthogonal to the rest signal of other components) for each component in a first-order system [25, 26]. The NAS is defined as:

$$NAS = \left(I - R_n R_n^+\right) r_n \tag{8}$$

where I is the identity matrix, R_n the matrix of pure spectra of all constituents except the *n*th analyte, R_n^+ the pseudoinverse or general inverse of R_n and r_n is the spectrum of the analyte. The NAS is a vector and is related to the regression vector as depicted in the following equations:

$$c = rb + e \tag{9}$$

$$b = \frac{NAS}{\|NAS\|_2} \tag{10}$$

where $\|NAS\|_2$ designates the square root of the sum of squares of each element in the vector b, c the analyte concentration and e the error vector. Sensitivity (SEN) and selectivity (SEL) were calculated using following equations:

$$SEN = \frac{1}{\|b\|_2} = \|NAS\|_2 \tag{11}$$

$$SEL = \frac{1}{\|b\|_2 \|r_n\|_2} = \frac{\|NAS\|_2}{\|r_n\|_2} \tag{12}$$

Figures of merit

There are several functions to evaluate and assess the correctness and also the validity of the calibration model

and also the effects of the preprocessing methods. In this study we have applied several objective function to check the efficiency of the calibration and preprocessing methods. The most common of these are root mean squared error of prediction (RMSEP), square of the correlation coefficient (R^2), relative error of prediction (REP) and predictive residual error sum of squares (PRESS) which were calculated for each component as follows:

$$RMSEP = \left[1/n \sum_{i=1}^{n} (\hat{x}_i - x_i)^2 \right]^{0.5}$$

$$R^2 = \sum_{i=1}^{n} (\hat{x}_i - \bar{x})^2 \Big/ \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$REP(\%) = \frac{100}{\bar{x}} \left[\frac{1}{n} \sum_{i=1}^{n} (\hat{x}_i - x_i)^2 \right]$$

$$PRESS = \sum_{i=1}^{n} (\hat{x}_i - x_i)^2$$

where x_i is the true concentration of the analyte in the sample i, represents the estimated concentration of the analyte in the sample i, the mean of true concentration in the prediction set, and n the total number of samples used in the prediction sets

In this work we report the simultaneous spectro-photometric determination of lead, zinc, mercury and cadmium with 4-2-pyridylazo resorcinol PAR using PCR, PLS and ITTFA multivariate calibration methods. The results obtained by these methods with and without preprocessing using DOSC, derivation and smoothing by the Savitzky-Golay filter are compared and discussed. We apply two versions sof derivative spectra a) without removing of the noisy part of the derivative spectra, named as DERV and b) with removing the noisy part named as CORRDERV.

EXPERIMENTAL
Reagent

All chemical were of analytical-reagent grade and deionized water was used throughout. Stock solutions of 500 ppm of lead, zinc, mercury and cadmium were prepared from their nitrate salts. A stock solution of PAR 8.48×10^{-3} M in water was prepared by dissolving solid reagent samples. Buffer solutions (pH 1.00-11.00) were prepared by mixing sufficient amount of HCl, and KCl, (0.2 M) for pH=1.00 and 2.00, potassium hydrogen phthalate and NaOH, (0.1 M) for pH= 2.00 to 4.00,

tris (hydroxymethyl) aminomethane and HCl (0.1M) for pH= 7, 8, 9 and finally sodium bicarbonate (0.05 M) and NaOH (0.1 M) for pH=10 [27]. The pH-meter was calibrated with at least two buffer solutions at pH= 2.00 and 9.00.

Apparatus

Electronic absorption measurements were carried out on a CECIL 9000 spectrophotometer (slit width 0.2 nm and scan rate 500 nm/min) using 1.00 cm quartz cells. A Metrohm 692 pH-meter furnished with a combined glass-saturated calomel electrode was used for pH measurements.

Computer hardware and software

Absorption spectra were digitized and stored at wavelengths from 460 to 600 nm in steps of 0.5 nm and then transferred in text format to a Pentium II computer for subsequent manipulation by corresponding programs. The data pretreatment was done in workspace of MATLAB for windows (Mathworks, Version 6.5). PLS program for calibration-prediction and experimental design was written in MATLAB according to the algorithm described by *Martens* and *Naes* [9] and PLS routine of PLS-Toolbox (Eigenvector Company, Version 2.1). ITTFA program for calibration-prediction according to the algorithm described by *Gemperline* [14] was written in MATLAB and DOSC m-file achieved from reference [30].

Procedure

In a 5 ml calibrated flask, known amounts of the standard solutions of each cation and 2.0 ml of PAR solution (8.48×10^{-5} M) were placed and diluted to the mark with appropriate buffer solution. The final concent-ration of lead, mercury, zinc and cadmium cations were varied between 0.00 - 9.81, 0.00 - 0.87, 0.00 - 12.24, and 0.00 - 3.96 ppm respectively according to a simplex lattice (4,4). Finally after 30 minutes the spectra of all prepared solutions were recorded on spectrophotometer. The autoscaled spectra were used for subsequent analysis.

RESULTS AND DISCUSSION

The pH has a drastic effect on the resolution, shape and also intensity of the spectra of the complexes of cited metal ions with PAR. So the influence of the pH of the
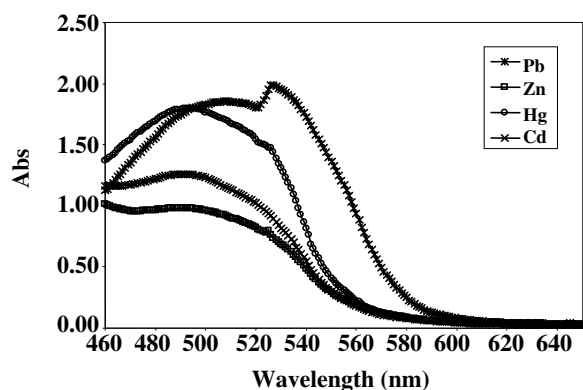
Fig. 1: The effect of pH on the absorption spectra of the PAR and lead, zinc, mercury and cadmium complexes in different pH-value.

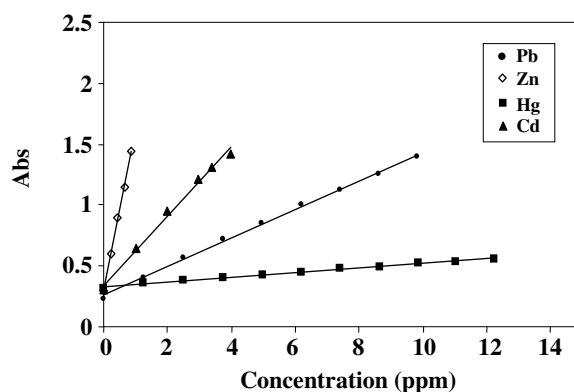*Fig. 2: Analytical curve for univariate determination of lead, zinc, mercury and cadmium complexes at optimum experimental conditions.*

medium on the absorption spectra, and in consequence on the sensitivity and selectivity of the determination of metal complexes, was studied over the range 1.0 - 11.0 and the result at pH=7.5 is shown in Fig. 1.

This pH=7.5 is the optimum value according to variations of the SEN and SEL values against pH. This value is selected as the optimum value, to compromise between the sensitivity and selectivity for all four metal ions and also to reduce the effect of the uncomplexed PAR absorbance. A reagent concentration of 8.48×10^{-5} M was chosen to ensures a sufficient reagent excess relative to all metal ions concentrations.

Univariate Calibration

To evaluate the linear behavior of each metal complex and also the linear dynamic range for each metal ion an individual calibration curve was constructed using the absorbance values of the resulting complexes at their $\lambda_{max's}$ (508, 489, 493 and 491 nm for lead, zinc, mercury and cadmium respectively) vs. metal ion concentrations. The linear ranges, as depicted in Fig. 2, are 0.0 - 9.8, 0.0 - 0.7, 0.0 - 12.24 and 0.0 - 3.96 ppm for of lead, zinc, mercury and cadmium, respectively. Linear regression results; line equations and R^2 are:

Abs. $= 0.1171\ C_{Pb} + 0.2612$ ($R^2 = 09976$),

Abs. $= 1.3108 C_{Zn} + 0.3046$ ($R^2 = 0.9992$),

Abs. $= 0.2853 C_{Hg} + 0.3398$ ($R^2 = 0.9936$) and

Abs. $= 0.0192 C_{Cd} + 0.3291$ ($R^2 = 0.9950$) for lead, zinc, mercury and cadmium, respectively.

Mixture Design

Multivariate calibration methods require a suitable

experimental design of the standard belonging to the calibration set in order to provide good prediction ability. A mixture design for four component mixtures was used for calibration [11, 12, 30]. The calibration set, table 1, contains 35 standard solutions and the pH of all solution was adjusted to 7.5 just before recording. The spectra was recorded between 400 and 650 nm, which implies working with 401 experimental data points per spectrum (as the spectra are digitized every 0.5 nm interval). This region was selected for analysis, because this region contains the suitable spectral information from the component mixtures of interest. The compositions of the calibration mixtures were selected according to a (4,4) simplex lattice design. For model assessment, seven test mixtures inside the linear range of each metal ion were used. The concentrations of each cation in test mixtures were inside the calibrations space by inspection of their scores in the first PC vs. second PC's plot, (Fig. 3).

Optimum Conditions for Calibration and Preprocessing Methods

The cost functions for different calibration and preprocessing methods rise from the deviation of the calculated concentration in calibration or predication set by their analytical values. Here we use several types of equations which indicate and stress on different aspects of random and bias errors.

Selection of Optimum Number of Factors for Calibration Methods

To select the number of factors in PCR or PLS algorithm, in order to model the system without over

Table 1: The 35 designed sample solutions by using (4, 4) simplex lattice design.

| No. of Solution	Lead (II) ppm	Zinc (II) ppm	Mercury(II) ppm	Cadmium(II) ppm
1	9.80	0.00	0.00	0.00
2	0.00	0.87	0.00	0.00
3	0.00	0.00	12.24	0.00
4	0.00	0.00	0.00	3.96
5	4.93	0.43	0.00	0.00
6	4.89	0.00	6.13	0.00
7	4.93	0.00	0.00	1.97
8	0.00	0.43	6.17	0.00
9	0.00	0.43	0.00	1.98
10	0.00	0.00	6.17	1.97
11	2.46	0.64	0.00	0.00
12	2.44	0.00	9.19	0.00
13	2.47	0.00	0.00	2.95
14	0.00	0.22	0.00	2.97
15	0.00	0.00	3.09	2.96
16	0.00	0.21	9.21	0.00
17	4.91	0.21	3.08	0.00
18	4.93	0.22	0.00	0.99
19	4.91	0.00	3.08	0.98
20	0.00	0.43	3.09	0.99
21	2.46	0.43	3.09	0.00
22	2.47	0.43	0.00	0.99
23	2.45	0.21	6.15	0.00
24	2.45	0.00	6.15	0.98
25	0.00	0.21	6.17	0.98
26	0.00	0.22	3.09	1.97
27	2.46	0.00	3.09	1.97
28	2.47	0.22	0.00	1.98
29	7.35	0.21	0.00	0.00
30	7.38	0.00	3.09	0.00
31	7.37	0.00	0.00	0.98
32	0.00	0.65	0.00	0.99
33	0.00	0.00	9.21	0.98
34	0.00	0.65	3.09	0.00
35	2.46	0.22	3.09	0.98
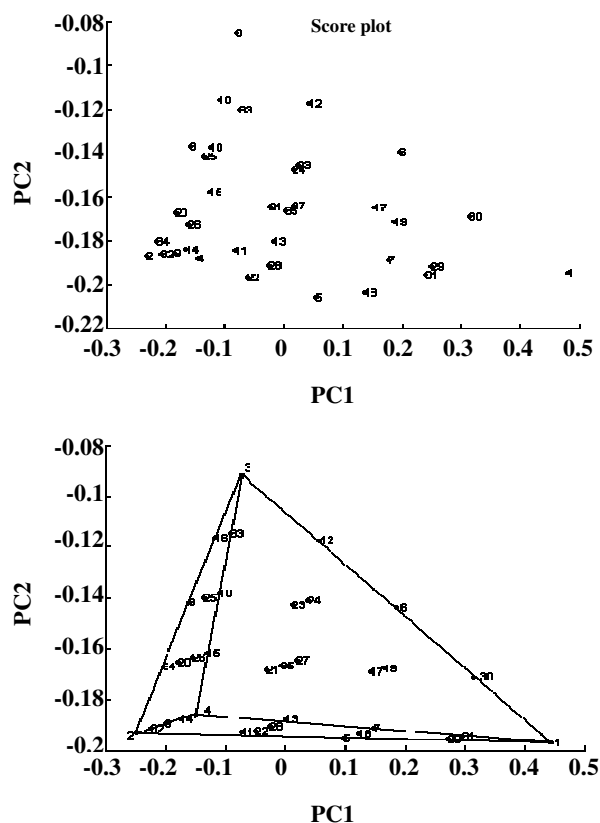
***Fig. 3: Score plot of calibration set solutions as PC2 vs. PC1
a) original , b) after DOSC.***

fitting of the concentration data, a cross-validation method, leaving out one sample at a time, was used [10]. Given the set of 35 calibration spectra, the PLS or PCR calibration on 34 spectra were performed, and using this calibration the concentration of the compounds in the sample left out during calibration was predicted. This process was repeated 35 times until each calibration sample had been left out once. The predicted concentration of the compounds in each sample was compared with the known concentration of the compound in this reference sample and PRESS was calculated. The PRESS was calculated in the same manner each time a new factor was added to the PLS or PCR model.

One reasonable choice for the optimum number of factors would be that number which yielded the minimum PRESS. However, for PCR and PLS models using the number of factors (h*) that yields a minimum in PRESS usually lead to some over fitting. A better criterion for selecting the optimum number of factors involves the comparison that PRESS from model should not be significantly greater than PRESS from the model with h*

factors. The F statistic was used to make the significance determination. *Haaland* and *Thomas* [10] empirically determined that an F-ratio probability of 0.75 is a good choice.

For PLS and PCR models, we selected the number of factors for the first PRESS values with the F-ratio probability, which drops below 0.75. The number of factors which are selected as the optimum numbers for the calibration models were 5, 6, 12 and 11 in PLS and 4, 7, 6, 7 in PCR for Pb, Zn, Hg and Cd, respectively.

We used a different tactic for selection of the optimum number of factors in ITTFA. For ITTFA calibration, the predicted concentration profile of each compound in samples was compared with the known concentration profile of the compound in the reference samples and each time the relative error of prediction (REP) was calculated. The REP was calculated in the same manner each time a new factor was added to the ITTFA model. The number of component which yielded the minimum REP, was selected as optimum number in ITTFA without and with application of preprocessing methods. The REP value reached to a minimum at numbers 4, 6, 4 and 7 for Pb, Zn, Hg and Cd, respectively.

The effect of the DOSC and first derivation of spectra (DERV) and corrected first derivative of spectral data (CORDERV) on the PCR, PLS and ITTFA calibration methods were evaluated by using commonly RMSEP, and the correlation coefficients of the line was obtained by plotting the calculated versus analytical concentration in the calibration set. The RMSEP and linear correlation coefficients are listed in table 2. As it can be seen from the data in table 2 it is easy to conclude that DOSC-PLS, DERV-PLS, CORRDERV-PLS and PLS show distinct superiority with respect to PCR and ITTFA with and without preprocessing methods.

Tuning of DOSC

The parameters which have the main role on the efficiency of the DOSC filter as a preprocessing method are tolerance and the number of components. Here we used REP and correlation coefficient of the plot of analytical versus calculated concentrations of the metal ions. As precious experiences have shown the number of components are rarely more than two DOSC components. So we check the effect of the tolerance value at just one and two DOSC components.

Table 2: The RMSEP, REP, R², PRESS R and number of factors for prediction set.

Original	Component	REP	RMSEP	PRESS	R^2	R
PCR	Pb(4)*	3.9424	0.0693	0.0336	0.9175	0.9579
	Zn(7)	9.9384	0.0306	0.0065	0.8890	0.9428
	Hg(6)	31.655	0.6971	3.4021	0.7617	0.8728
	Cd(7)	17.676	0.1991	0.2776	0.6366	0.7978
PLS	Pb(5)	3.6103	0.0635	0.0282	0.9359	0.9674
	Zn(6)	8.7978	0.0271	0.0051	0.8899	0.9434
	Hg(12)	4.6393	0.1021	0.0730	0.9426	0.9709
	Cd(11)	5.2283	0.0589	0.0242	0.9557	0.9776
ITTFA	Pb(4)	7.8185	0.1375	0.1323	0.8420	0.9176
	Zn(6)	37.1170	0.1144	0.0917	0.5026	0.7090
	Hg(4)	42.1490	0.9282	6.0318	0.5027	0.7090
	Cd(7)	23.1760	0.2611	0.4773	0.6590	0.8118
DOSC	Component	REP	RMSEP	PRESS	R^2	R
PCR TOL=1E-3 NOCOMP=2	Pb(4)	4.5798	0.0805	0.0454	0.9517	0.975
	Zn(7)	8.2287	0.0253	0.0045	0.9184	0.9583
	Hg(4)	2.7925	0.0615	0.0264	0.9843	0.9921
	Cd(7)	16.198	0.1825	0.2331	0.5566	0.7460
PLS TOL=1E-3 NOCOMP=2	Pb(4)	4.5015	0.0791	0.0438	0.9854	0.9926
	Zn(6)	6.8671	0.0211	0.0031	0.9230	0.9607
	Hg(4)	2.7402	0.0603	0.0254	0.9850	0.9925
	Cd(11)	7.9200	0.1393	0.1358	0.8509	0.9224
ITTFA TOL=1E-3 NOCOMP=2	Pb(6)	23.8860	0.4201	1.2355	0.6358	0.7974
	Zn(6)	39.4850	0.1217	0.1037	0.4995	0.7067
	Hg(4)	49.1830	1.0832	8.2129	0.3774	0.6143
	Cd(10)	36.6520	0.4129	1.1939	0.4471	0.6687
DERV	Component	REP	RMSEP	PRESS	R^2	R
PCR	Pb(4)	5.4797	0.0964	0.0650	0.9039	0.9508
	Zn(4)	16.378	0.0505	0.0179	0.8883	0.9425
	Hg(4)	38.292	0.8433	4.9784	0.6343	0.7965
	Cd(8)	1.8018	0.0203	0.0029	0.9487	0.9740
PLS	Pb(5)	4.23	0.0744	0.03875	0.9380	0.9685
	Zn(8)	1.196	0.0036	0.0001	0.9887	0.9943
	Hg(8)	3.381	0.0744	0.0388	0.9851	0.9855
	Cd(8)	1.8018	0.0203	0.0028	0.9487	0.9740

Table 2

	Component	REP	RMSEP	PRESS	R^2	R
ITTFA	Pb(6)	23.8860	0.4201	1.2355	0.6358	0.7974
	Zn(6)	39.4850	0.1217	0.1037	0.4995	0.7067
	Hg(4)	49.1830	1.0832	8.2129	0.3774	0.6143
	Cd(10)	36.6520	0.4129	1.1939	0.4471	0.6687
CORDERV	Component	REP	RMSEP	PRESS	R^2	R
PCR	Pb(4)	5.7874	0.1017	0.0725	0.8894	0.9431
	Zn(4)	15.8610	0.0489	0.0167	0.8992	0.9483
	Hg(4)	38.5780	0.8496	5.0530	0.6321	0.7951
	Cd(4)	24.4570	0.2755	0.5315	0.5635	0.7507
PLS	Pb(6)	2.1111	0.0371	0.0096	0.9693	0.9845
	Zn(8)	1.9061	0.0058	0.0002	0.9957	0.9978
	Hg(7)	7.1972	0.1585	0.1758	0.9819	0.9909
	Cd(7)	3.3590	0.0378	0.0100	0.9549	0.9772
ITTFA	Pb(5)	25.4720	0.4480	1.4050	0.6421	0.8013
	Zn(5)	39.2150	0.1209	0.1023	0.4766	0.6904
	Hg(4)	51.5200	1.1346	9.0119	0.3443	0.5867
	Cd(5)	46.2450	0.5210	1.9006	0.3129	0.5593

The best values of the tuning of the DOSC by changing of the tolerance value at two levels of DOSC components are shown in table 2, for all four metal ion complexes. As it is very clear from the REP and R values for PLS calibration method, the best results are obtained at 2 DOSC components and the tolerance values between 0.01 - 0.001.

The effect of the DOSC parameters on the required number of significant factors in the PLS calibration model were studied. It can be concluded that the PRESS values are reduced by using two DOSC components in the model building. It means the application of DOSC does not results in a more efficient model but it just leads to a more simple and interpretable model. So, the number of significant factors by DOSC-PLS model and two DOSC components are 5, 5, 7 and 8 for lead, mercury, zinc and cadmium, respectively. The effects of the DOSC components and tolerance value on the PCR and ITTFA calibration models were also studied and the two DOSC components and 0.01-0.001 tolerance values are obtained. The numbers of significant factors required in the PCR and ITTFA models were 7, 7, 9 and 11 and 8, 9, 10 and 11 for lead, mercury, zinc and cadmium respectively.

Optimization of Savitzky - Golay filter

The fundamental idea of Savitzky-Golay algorithm is to fit a different polynomial to the data surrounding each data point. The number of data points and polynomial degree must be optimized by user. We use error value estimator (REP) and correlation coefficient (r) to optimize the Savitzky - Golay conditions. The most two important factors which have to be minimized are degree of polynomial and window size (the number of data points to be smoothed or derivative). The tuning process was performed as a one-at-a-time manner, that is, each time one variable was kept fix and the other one was lefts to change freely. First at a predefined degree of polynomial the number of data points changed and vice versa. The results of the optimization of the Savitzky-Golay filter are shown in table 2. The number of optimum data point (window size) and the order of the polynomial are 13 and 2 respectively for all calibration models and all metal ion complexes.

Smoothing After First Derivation of Spectra

To avoid enhancing the noise, which is a consequence of derivation, sometimes spectra are first smoothed.

The effect of smoothing was investigated and found that using Savitzky-Golay filer as soothing filter does not have satisfactory effects in the present work. This is probably due to the fact that by the application of smoothing process some useful information may be removed from the spectral data and lead to the increasing the error or at least let it remains constant. So we used first order derivation spectra obtained by Savitzky-Golay filter without any more smoothing for the calibration models. The results of effect of the smoothing on the calibration models by using REP and R are shown in table 2.

Comparison of the Calibration Models in Predicting of the Test Set Solutions.

To choose the best model according to its ability in prediction of the concentrations of the test set solutions, according to the effect of the number of DOSC components, number of significant factors, derivation, smoothing and   type of the calibration were used in model building by evaluation of the REP %, RMSEP, PRESS, R^2, and R values. The results of this assessment are presented in table 2. As it is shown in table 2 the most versatile model to predict the concentrations of test solutions is DOSC-PLS with 2 DOSC components, 0.001 tolerance and numbers of significant factors of 4, 6, 4 and 11 for lead, zinc, mercury and cadmium, respectively, on the original spectral data.

CONCLUSIONS

The proposed method is a spectrophotometric simultaneous determination of some heavy metal ions using PAR as a metallochromic indicator with several chemometrics calibration and with two preprocessing methods. The effect of several aspects of the combination of calibration and preprocessing methods on prediction sets were studied by evaluation of the different assessors. The results showed that the most versatile method for prediction of the concentration of four metal ions form their mixture spectral data is DOSC-PLS with 4, 5, 4 and 11 significant factors for lead, zinc, mercury and cadmium, respectively. The most important finding of this study is that, the application of the DOSC on different preprocessed or original spectral data does not produce a more efficient model but it produces a simple and more interpretable model.

REFERENCE

[1] Copeland, L.C., Griffith, F.S., *Anal. Chem.*, **22**, 1269 (1950).

[2] Mykytiuk, A. P., Russell, D. S., Sturgeon, R. E., *Anal. Chem.*, **52**, 1281 (1980).

[3] Carroll, S. A., O'Day, P., Piechowski, M., *Environ. Sci. Technol.*, **32**, 956 (1998).

[4] Pei, J., Tercier-Waeber, M. L., Buffle, J., *Anal. Chem.*, **72**, 161 (2000) .

[5] Ichinoki, S., Yamazaki, M. *Anal. Chem.*, **57**, 2219 (1985).

[6] O'Haver, T.C., Green, G. L., *Anal. Chem.*, **48**, 312 (1976).

[7] Ghasemi, J., Niazi, A., *Microchem. J*., **68**, 1 (2001).

[8] Geladi, P., Kowalski, B. R., *Anal. Chim. Acta,* 1**85**, 1 (1986).

[9] Martens, H., Naes, T., "Multivariate Calibration" John Wiley, New York (1991).

[10] Haaland, D. M., Thomas, E. V., *Anal. Chem.*, **60**, 1193 (1988).

[11] Ghasemi, J., Shahabadi, N., Seraji, H. R., *Anal. Chim. Acta,* **510**, 121 (2004).

[12] Ghasemi, J., Ahmadi, Sh., Torkestani, K., *Anal. Chim. Acta*, **487**, 181 (2003).

[13] Luypaert, J., Heuerding, S., de Jong, S., Massart, D. L., *J. Pharm. Biomed. Anal*., **30**, 453 (2002).

[14] Gemperline, P. J., *J. Chem. Inf. Comput. Sci.*, **24**, 206 (1984).

[15] Westerhuis, J. A., de Jong, S., Smilde, A. K., *Chemom. Intell. Lab. Sys*., **56** , 13 (2001).

[16] Wold, S., Antti, H., Lindgren, F., Ohman, J., *Chemom. Intell. Lab. Sys*., **44**, 175 (1998).

[17] Sjoblom, J., Svensson, O., Josefson, M., Kullberg, H., Wold, S., *Chemom. Intell. Lab. Sys.,* **44**, 229 (1998).

[18] Wise, B. M., Gallagher, N. B., http:// www. eigenvector. com/matlab/osc.html. (1999).

[19] Fearn, T., *Chemom. Intell. Lab. Sys*., **50**, 47(2000).

[20] Andersson, C. A., *Chemom. Intell. Lab. Sys*., **47**, 51 (1999).

[21] Savitzky, A., Golay, M. J. E., *Anal. Chem.*, **36**, 1927 (1964).

[22] Press, W. H., Teukolsky, S. A., Wetterling W.T., Flanerry B. P., "Numerical Recipes In C: The Art of Scientific Computing", Cambridge University

Press., Chap. 14, Statistical Description of Data (1992).

[23] Gorry P. A., *Anal. Chem.*, **62**, 570 (1990). and Signal Processing Toolbox: sgolayfilt command.

[24] Lorber, A., *Anal. Chem.*, **58**, 1167 (1988).

[25] Weast, R. C., "CRC Handbook of Chemistry and Physics", 1th Ed. © CRC Press, Inc. (1988).

[26] Eriksson, L., Johansson, E., Wold N. K., Wold, S., Multi- and Megavariate Data Analysis, Umetrics, Umea, Sweden (2001).

[27] http://www-its.chem.uva.nl/ research/ pac/ Software/ osc.zip

[28] Massart, D. L., Vandeginste, B. G. M., Buydens, L. M. C., de Jong, S., Lewi, P.J., Smeyers-Verbeke, J., "Handbook of Chemometrics and Qualimetrics" Part A, Elsevier, Amsterdam (1997).

[29] Edward V. Thomas and David, M., *Anal. Chem.*, **62**, 1091 (1990).

[30] http://www.eigenvector.com/MATLAB/OSC.html