



International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)

www.iasir.net

Impact Evaluation of Uncorrected Background on the Minimum Uncertainty of the "White" Mixtures Analysis Using Tikhonov Regularization and the Partial Least Squares Method

J. Dubrovkin
Computer Department
Western Galilee College
2421, Acre
Israel

Abstract: The paper reports the Tikhonov regularization and the partial least squares-based analysis of "white" mixtures which spectra were disturbed by uncorrected various-profile background. The dependences of the minimum relative mean-square error of analysis on the background norm were obtained by different methods based on estimating the optimal regularization parameter and compared to each other. The upper limit of the minimum uncertainty of the multicomponent quantitative analysis was established. In the case of low background, the partial least-squares method was found to yield the best predictions performance. If the background intensity increases, the prediction uncertainty of this method also increases being larger than the minimal achievable value.

Keywords : uncertainty, quantitative spectrochemical analysis, Tikhonov regularization, partial least squares.

I. Introduction

Spectrochemical analysis of multicomponent mixtures with known qualitative composition ("white" mixtures) is one of the common methods used in analytical chemistry. For example, it is employed in pharmaceuticals to evaluate drug composition. The most popular model of the spectrum of "white" mixtures is the ideal one, which takes into account homoscedastic (noise variance is constant) and uncorrelated noise [1, 2]. However, in a more realistic situation, a number of factors should be included: the presence of unknown impurities in the sample solution, errors in preparing the reference pure-component solutions, and the perturbation of the mixture spectrum caused by the impact of unstable instrumental and external parameters such as light scattering and/or temperature [3, 4]. The distortions that are introduced into the ideal model by the above factors can be significantly corrected by such well-known preprocessing methods as the polynomial approximation of a background [5-7], the least squares fitting to the average spectrum (multiplicative scatter correction) [8], the derivative spectrometry [9, 10], and the wavelet transform [11]. An exhaustive review of the mathematical algorithms for the background correction in the multivariate calibration-based quantitative spectrochemical analysis of multicomponent mixtures can be found in [4, 12, 13]. In particular, the authors point out that complete elimination of background from a mixture spectrum is impossible.

Earlier we have shown theoretically [14] that the errors of quantitative analysis that employs the spectra perturbed by a randomly varying polynomial baseline could be significantly reduced by including the noise covariance matrix into the least-squares estimator. The polynomial coefficients used in the work were normal random numbers with zero means. As an extension of this study, here we numerically calculate the minimum relative mean-square errors (RMSE) of the Tikhonov regularization and the partial least squares-based analysis of mixtures whose spectra contain random background with non-zero mean. The impact of the Lorentzian, Gaussian and polynomial background on the best achievable metrological characteristics of the analysis is evaluated. For comparison, the corresponding RMSE values are also determined using simple preprocessing methods (the centering, the 1st-order differentiation and the algebraic background correction).

Standard notations of linear algebra are used throughout the paper. Bold upper-case and lower-case letters denote matrices and column vectors, respectively. Upper-case and lower-case italicized letters represent scalars. All calculations were performed and the plots were built using the MATLAB program.

II. Least Squares-Based Analysis

The spectrum of an n -component mixture can be represented as

$$\mathbf{y} = \mathbf{A}\mathbf{c} + \boldsymbol{\varepsilon}, \quad (1)$$

where $\mathbf{A} = \mathbf{K}_n \mathbf{W}$, $\mathbf{K}_n = \mathbf{K} + \mathbf{\Xi}$, $\mathbf{K}$ is the matrix of the pure-component spectra, $\mathbf{\Xi}$ is the noise matrix in which the elements of each column are normally distributed with zero mean and the same variance σ_K^2 , $\mathbf{W}$ is the diagonal matrix with unit trace, $\mathbf{c}$ is the vector of the component fractions in the mixture, and $\boldsymbol{\varepsilon}$ is the normal noise with zero mean and covariance matrix $\sigma_y^2 \mathbf{I}$, where $\mathbf{I}$ is the identity matrix.

To perform the non-regularized least-squares-based analysis of the mixture spectrum that contains unknown background $\mathbf{b}$, we will use three simple methods of background correction, namely, the centering (CN), the 1st derivative (DR), and the algebraic correction (AC) methods. These methods are based on column-wise matrix transformations and not the transformations of the matrix coordinate system [15], similar to orthogonal signal correction [12]. Using these methods, the concentration vector $\mathbf{c}_v$ can be estimates as:

$$\hat{\mathbf{c}} = [(\mathbf{K}_n^{tr} \mathbf{W})^T \mathbf{K}_n^{tr} \mathbf{W}]^{-1} (\mathbf{K}_n^{tr} \mathbf{W})^T (\mathbf{K}_n^{tr} \mathbf{W} \mathbf{c}_v + \mathbf{b}^{tr} + \boldsymbol{\varepsilon}), \quad (2)$$

where

$$\mathbf{K}_n^{tr} = \begin{cases} \text{CN: } [\bar{\mathbf{k}}_1, \bar{\mathbf{k}}_2, \dots, \bar{\mathbf{k}}_n], \\ \text{DR: } [\mathbf{k}'_1, \mathbf{k}'_2, \dots, \mathbf{k}'_n], \\ \text{AC: } [\mathbf{k}_1^{trn}, \mathbf{k}_2^{trn}, \dots, \mathbf{k}_n^{trn}], \end{cases}$$

$\bar{\mathbf{k}}_j = \mathbf{k}_j - \mathbf{1}\bar{k}_j$, $\mathbf{k}_j$ is the j -column of matrix $\mathbf{K}$, $\mathbf{1}$ is the vector of all units, $\bar{k}_j$ is the mean value of $\mathbf{k}_j$, $\mathbf{k}'_j$ is the first derivative of $\mathbf{k}_j$, $\mathbf{k}_j^{trn} = \mathbf{O}\mathbf{O}^T \mathbf{k}_j$ is vector $\mathbf{k}_j$ transformed using truncated matrix $\mathbf{O} = [\mathbf{o}_p, \mathbf{o}_{p+1}, \dots, \mathbf{o}_q]$, and $\mathbf{o}_p$ is the vector whose elements are the values of the p -order Chebyshev orthogonal polynomial [15]. Vector $\mathbf{b}^{tr}$ is obtained by the same transformation as vectors $\mathbf{k}^{tr}$. (In what follows, we use the super - and the subscripts only for clarity).

If the unknown background can be well approximated by low-order polynomials in the range of $[0, p - 1]$, then the reduction of these polynomials in the orthogonal transformation of matrix $\mathbf{K}$ and of vector $\mathbf{b}$ will eliminate this background.

In the case of “white” mixtures, matrix $\mathbf{K}$ is known *a priori*. Then, for given vectors $\mathbf{b}$ and $\mathbf{c}_v$, we will search for such matrix $\mathbf{W}$ that minimizes the relative mean squared error (RMSE) of vector $\hat{\mathbf{c}}$:

$$RMSE(\mathbf{W})_{min} = \min_{\mathbf{W}} \sqrt{\sum_{t=1}^n ((\hat{\mathbf{c}} - \mathbf{c}_v)_t)^2} / \sqrt{\sum_{t=1}^n (\mathbf{c}_v)_t^2}, \quad (3)$$

where the line over the sum indicates the average value obtained from an independent analyses.

III. Tikhonov Regularization-Based Analysis

The regularization-based estimate of the vector $\mathbf{c}_v$ is:

$$\hat{\mathbf{c}} = [(\mathbf{K}_n \mathbf{W})^T \mathbf{K}_n \mathbf{W} + \alpha^2 \mathbf{I}]^{-1} (\mathbf{K}_n \mathbf{W})^T (\mathbf{K}_n \mathbf{W} \mathbf{c}_v + \mathbf{b} + \boldsymbol{\varepsilon}), \quad (4)$$

where α is the regularization parameter, usually calculated by the L-curve method [16]. If the Euclidian norm of vector $\mathbf{b}$ ($\|\mathbf{b}\|$) is known *a priori*, then, according to the perturbation theory, the regularization parameter can be evaluated by minimizing the following expression [18]:

$$\min_{\alpha} \{\alpha\gamma + \|\mathbf{b}\|\alpha^{-1/2}\}, \quad (5)$$

where $\gamma = \|(\mathbf{A}^T \mathbf{A})^{-2} \mathbf{A}^T (\mathbf{A} \mathbf{c}_v + \boldsymbol{\varepsilon})\|$. Here, the distortions of matrix $\mathbf{A}$, which are very small as compared to the distortions of the mixture spectrum accounted for by the uncorrected background, are neglected.

Eq. (5) reaches its minimum when

$$\alpha_{min} = \{\|\mathbf{b}\|/(2\gamma)\}^{1/3}. \quad (6)$$

Next, for given vectors $\mathbf{b}$ and $\mathbf{c}_v$, we searched for matrix $\mathbf{W}$ that minimizes the RMSE-value:

$$RMSE(\mathbf{W}, \alpha_{opt})_{min} = \min_{\mathbf{W}, \alpha_{opt}} \sqrt{\sum_{t=1}^n ((\hat{\mathbf{c}} - \mathbf{c}_v)_t)^2} / \sqrt{\sum_{t=1}^n (\mathbf{c}_v)_t^2}, \quad (7)$$

where α_{opt} is the optimal value of the regularization parameter, calculated for each matrix $\mathbf{W}$ using three techniques, namely, the L-curve method, Eq. (6), and fitting to the global minimum of the function of two variables, $\min_{\mathbf{W}, \alpha} RMSE$ (the fitting mode).

IV. The Partial Least Squares-Based Analysis

The matrix of the calibration set spectra is

$$\mathbf{Y} = \mathbf{C} \mathbf{A}^T, \quad (8)$$

where $\mathbf{C}$ is the “closure” concentration matrix (in which the sum of the elements of each row is unit). For three-component mixtures, we choose matrix $\mathbf{C}$ (66 x 3), each row of the matrix containing non-coinciding combinations of three numbers taken from array $[0, 0.1, 0.2, \dots, 0.9, 1]$. This experimental design uniformly covers all mixture concentrations from 0 to 1.

The estimate of the component concentration vector $\mathbf{c}_v$ is:

$$\hat{\mathbf{c}} = \mathbf{B}_{PLS} [\mathbf{1}, \mathbf{K}_n \mathbf{W} \mathbf{c}_v + \mathbf{b} + \boldsymbol{\varepsilon}], \quad (9)$$

where $\mathbf{B}_{PLS}$ is the regression matrix, calculated by the partial least-squares algorithm (PLS) [1].

The PLS regression technique was also applied to matrix $\mathbf{K}_n^{tr}$ and vector $\mathbf{b}^{tr}$ (PLS-truncation), as described in the Section II. The RMSE-values were calculated according to Eq. 3.

V. Computer Modeling

Computer modeling was performed using the UV-spectra of Excedrin tablet components and the sums of Gaussians (Fig.1). The components of the concentration vector $\mathbf{c}_v$ (in %) of the mixtures under study are shown in Fig. 2. The signal-to-noise ratio of the mixture spectra was 300÷500. The background spectrum for these mixtures was modeled by the wings of a Lorentzian or a Gaussian band, by a second-degree polynomial, or by a single Gaussian band as an interfering component (Fig. 3). The peak full widths at half maximum (w_f), the peak positions (i_{max}) and amplitudes (I_{max}), and the coefficients (a_0, a_1 and a_2) of polynomial $a_0 + a_1 i + a_2 i^2$, where i is the number of the point, were varied in wide ranges so that several hundred cases could be studied for each model. The *RMSE* values were obtained as the mean of 100 independent calculations. In each calculation, all the background parameters (P) were changed randomly:

$$P = P_{const} * \gamma, \quad (10)$$

where P_{const} is a constant, γ is a normally distributed random variable whose mean and the standard deviation values are equal to 1 and 0. 2, respectively.

The lower and the upper limits of the orthogonal polynomial order used for the algebraic background correction were selected according to the following considerations. On the one hand, the upper limit must be large enough to provide good approximation of the spectra. On the other hand, this limit should not be too great since, in the presence of noise, the errors of calculating high- order polynomials are very large [15]. As for the lower limit, it should be as small as possible to preserve the essential spectral information, but, at the same time, large enough to suppress the background spectrum.

For data sets 1÷3 and for Excedrin tablets, the algebraic correction of the Lorentzian background was performed using orthogonal polynomials of the 3rd÷5th order. The upper, 5th-order, limit provides satisfactory approximation of the component spectra (Figs. 4a-d). For data set 4, the above considerations required choosing the upper limit of the 9th order although it provided only poor polynomial approximation (Fig. 4e). Lower 3rd-order limit allows significant suppression of the Lorentzian background (Fig. 4f) and complete suppression of the polynomial background (Fig. 4g). Thus for data set 4, the 2nd÷9th- and the 3rd÷9th-order orthogonal polynomials were used. In the former case, the polynomial background was partly suppressed (Fig. 4g). The PLS-regression was performed on the two principal components.

VI. Results and Discussion

The following conclusions can be drawn from the results obtained by computer modeling:

1. The minimum uncertainty (given by the root of the *RMSE* value, *RRMSE*) of the regularization-based least-squares multicomponent analysis is significantly larger in the presence of uncorrected background as compared to the case of the background being absent ($\|\mathbf{b}\| = 0$) [17]. This effect is most pronounced for the Gaussian background (Figs. 5a, c). The results obtained by modeling the background by three parts of a Gaussian band - the left and the right wings as well as the central peak (Figs. 3a, b) - are presented in Fig. 5. The errors were calculated separately for each part. The form of the *RRMSE*($\|\mathbf{b}\|$) function for such background is different from those for other background profiles (compare Figs. 5a, 6a, and 7a). The increase of the errors in the middle part of the background norm range (Fig. 5a) is accounted for by the impact of the interfering Gaussian band in the vicinity of the central point of the mixture spectrum (Fig. 3b).

2. In accordance with the theoretical Eq. 6, the optimal value of regularization parameter α_{opt} increases with the increase of the background norm (Figs. 5-7, c and d). At the same time, larger α_{opt} -values reduce random errors (noise) to a greater extent [18]. Therefore, the noise impact on the *RRMSE* values is significant only when the background is very low (compare plots “a” and “b” in Figs. 5-7).

The addition of the random noise to the mixture spectra of data set 4 slightly reduces the minimum uncertainty in the middle part of polynomial background norm $\|\mathbf{b}\|$ (Fig. 7b).

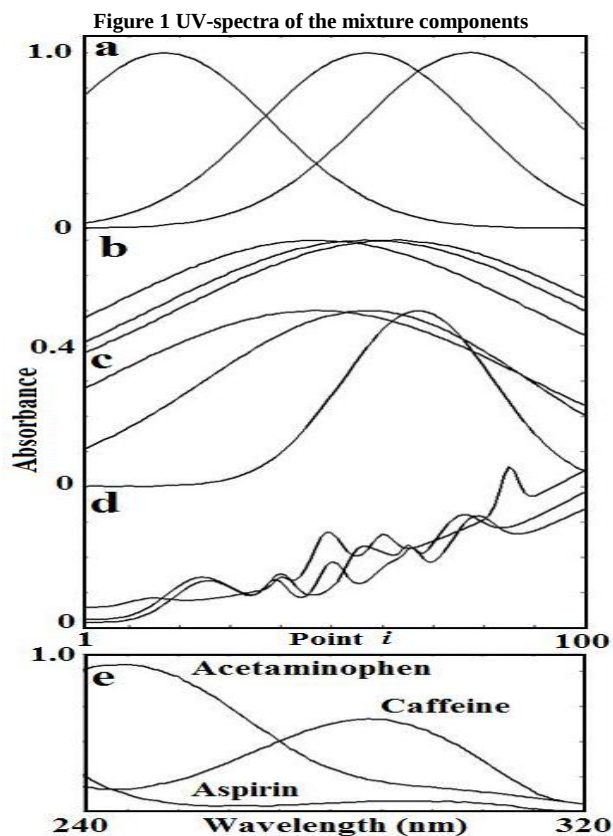
For non-regularized solutions, the relative uncertainty has the form [18]:

$$\frac{\|\delta \hat{\mathbf{c}}\|}{\|\mathbf{c}\|} \leq \text{cond}(\mathbf{A}) \frac{\|\mathbf{b}\|}{\|\mathbf{y}\|}, \quad (11)$$

where $\text{cond}(\mathbf{A})$ is the condition number of matrix $\mathbf{A}$. For data set 2, which is characterized by badly-resolved pure-component spectra, this number is very large and thus the relative errors are also large if the background is significant. In this case, the calculation in the fitting mode (Eq. 7) increases the regularization parameter and the slope of the *RRMSE*($\|\mathbf{b}\|$) dependence decreases (see Figs. 5, 6*, plots a, c (in the absence of noise); b, d (random noise)).

3. The comparison of different methods of calculating α (Figs. 8-12) shows that the use of *a priori* information on the background norm (Eq. 6) allows reducing the errors and making them close to the minimum achievable ones. This result is most pronounced for data set 2, which has large condition number. The L-curve method is not effective for evaluating the “best” value of α . In this situation, the norm of an unknown background can be

estimated by decomposing the mixture spectrum into pure-component spectra of known forms. This norm can be used for evaluating the value of α according to Eq. 6.



Spectra of Gaussians (a-d) and of Excedrin tablet components (e) obtained in water:methanol:glacial acid (69:28:3) solutions at concentrations $c = 20 \text{ mg L}^{-1}$. (Reproduced with permission from www.dissolutiontech.com)

Figure 2 Diagram of the concentration vector components

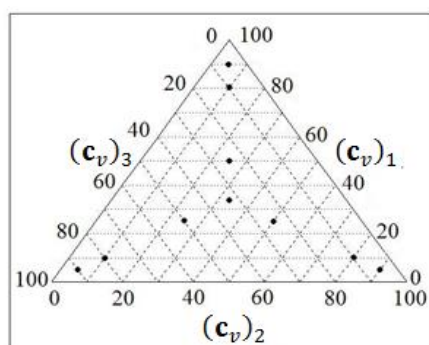
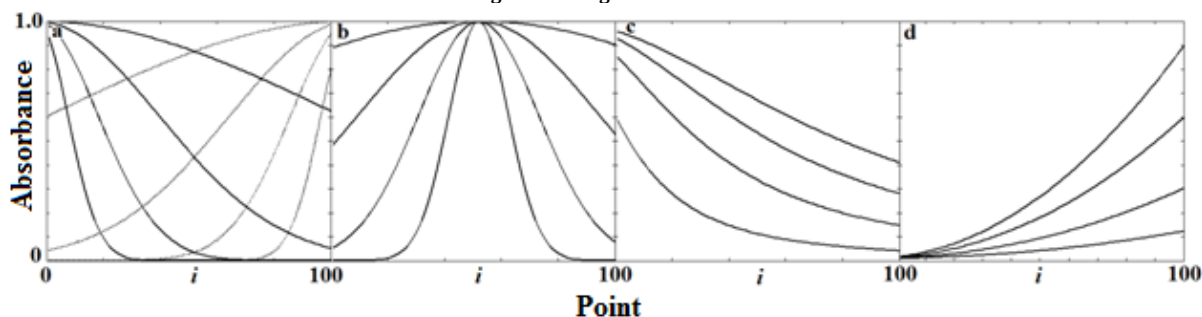
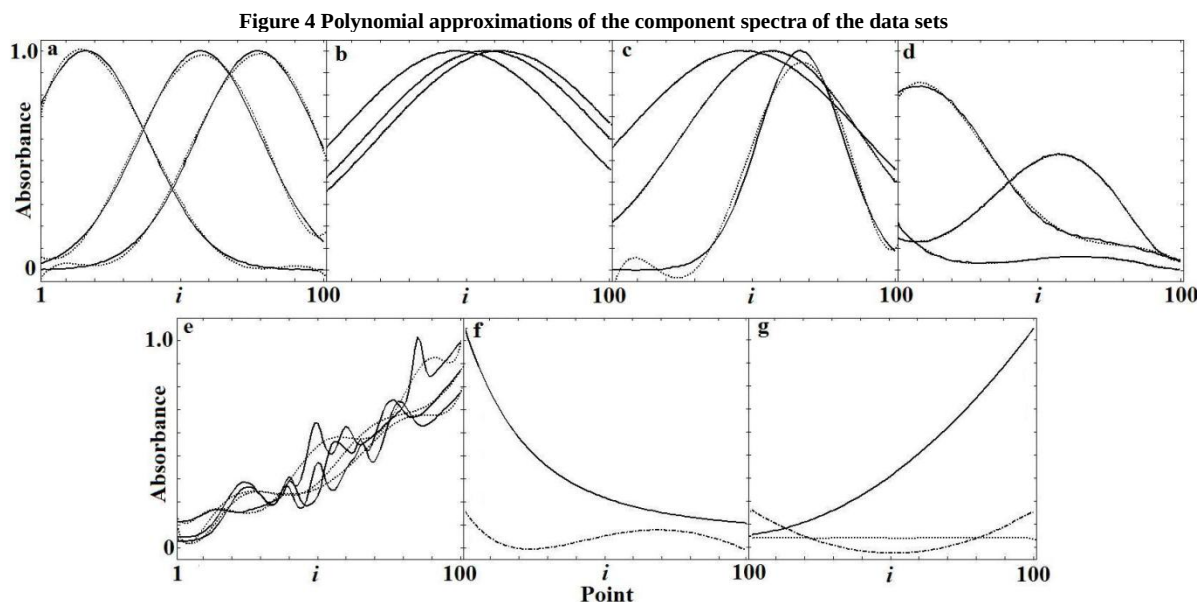


Figure 3 Background curves



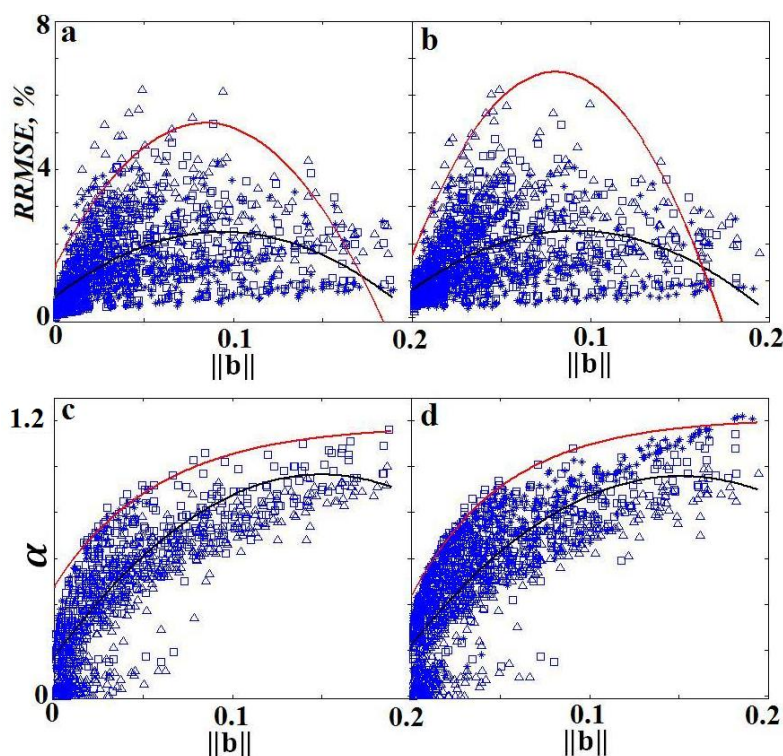
Gaussian background: (a) $i_{max} = 20$ (—), 42 (···). (b) $i_{max} = 31$; $I_{max} = 1$; w_f values of 5, 10, 20, and 50 relate to the curves in the “bottom-up” order. (c) Lorentzian background: $i_{max} = -20$, $I_{max} = 1$; w_f values of 50, 100, 150, and 200 relate to the curves in the “bottom-

up” order. (d) Polynomial background: $6 \times 10^{-4}(0.2 - 0.01i + a_2 i^2)$, a_2 values of 0.02, 0.05, 0.1, and 0.15 relate to the curves in the “bottom-up” order.



(a) - (d) data sets 1, 2, 3, and Excedrin tablet, respectively; the order of the approximating polynomial ranges from 0 to 5 (---). (e) data set 4; the order of the approximating polynomial ranges from 0 to 9 (---). Background: (f) Lorentzian (—); $i_{max} = -20$, $I_{max} = 1.67$, $w_f = 50$; the order of the approximating polynomial ranges from 3 to 5 (---). (g) Polynomial (—); $6 \times 10^{-4}(0.2 - 0.01i + 0.17i^2)$; the order of the approximating polynomial ranges from 2 to 9 (---) and from 3 to 9 (---).

Figure 5 Dependences of the relative error and of the regularization parameter on the Gaussian background norm. α -fitting procedure



$\sigma_K = 0$ (a, c) and 0.01 (b, d); $\sigma_y = 0$ (a, c) and 0.002 (b, d). Data sets: 1 (Δ), 2 (*), 3 ($\square$).

Figure 6 Dependences of the relative error and of the regularization parameter on the Lorentzian background norm. α -fitting procedure

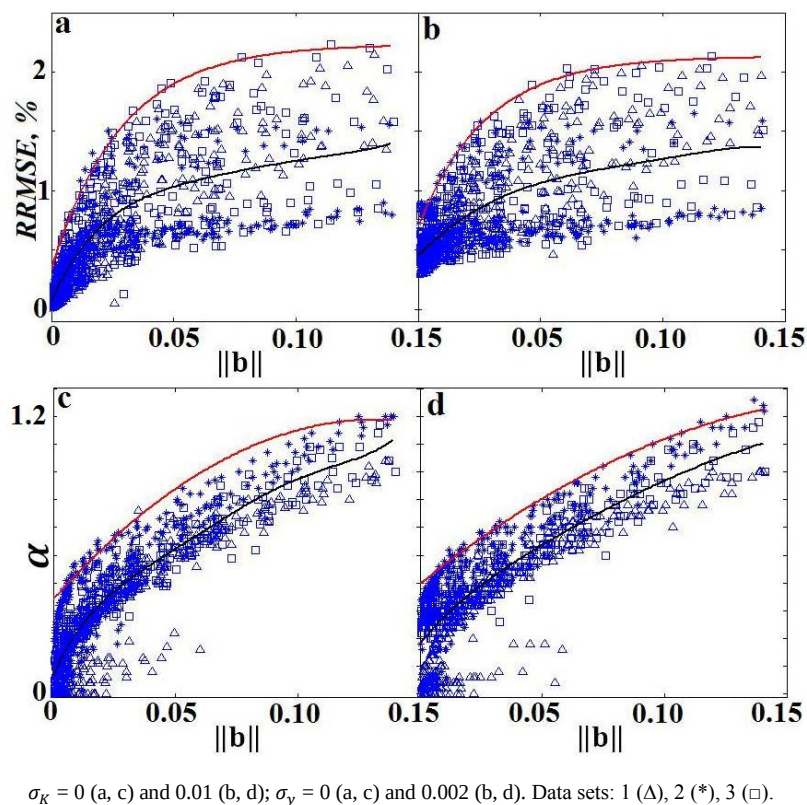


Figure 7 Dependences of the relative error and of the regularization parameter on the polynomial background norm. Data set 4. α -fitting procedure

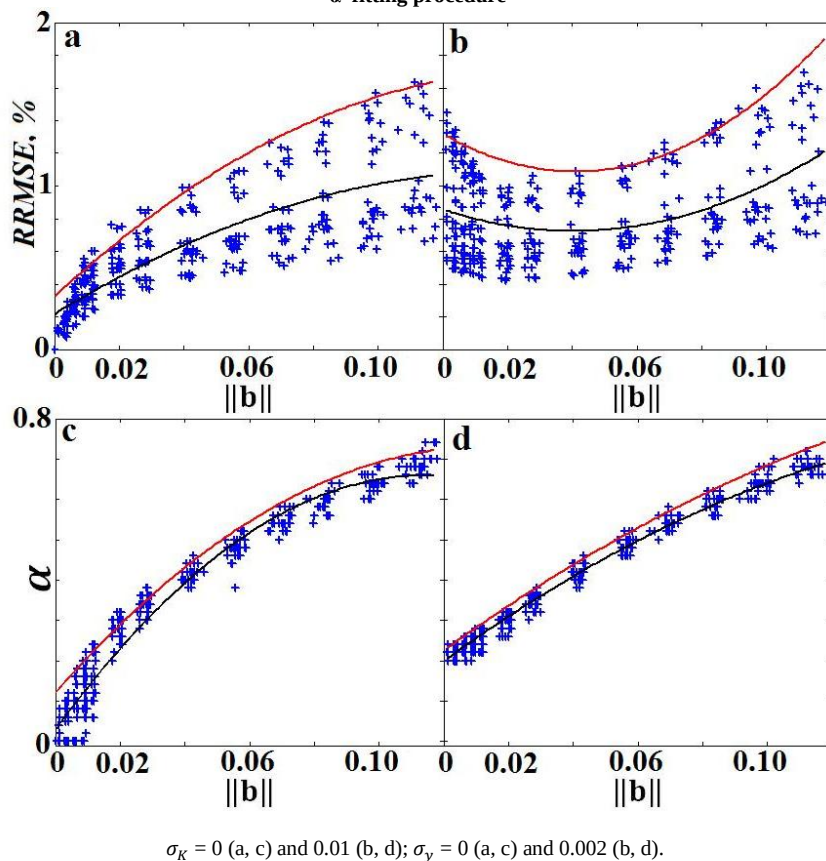
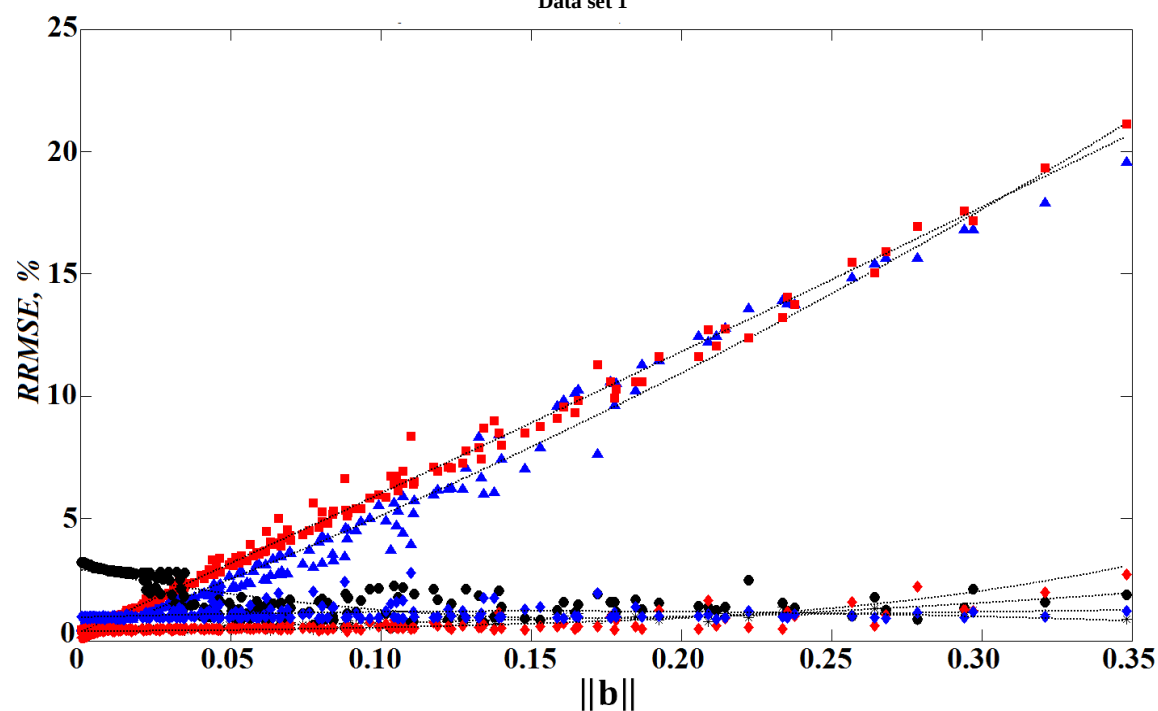
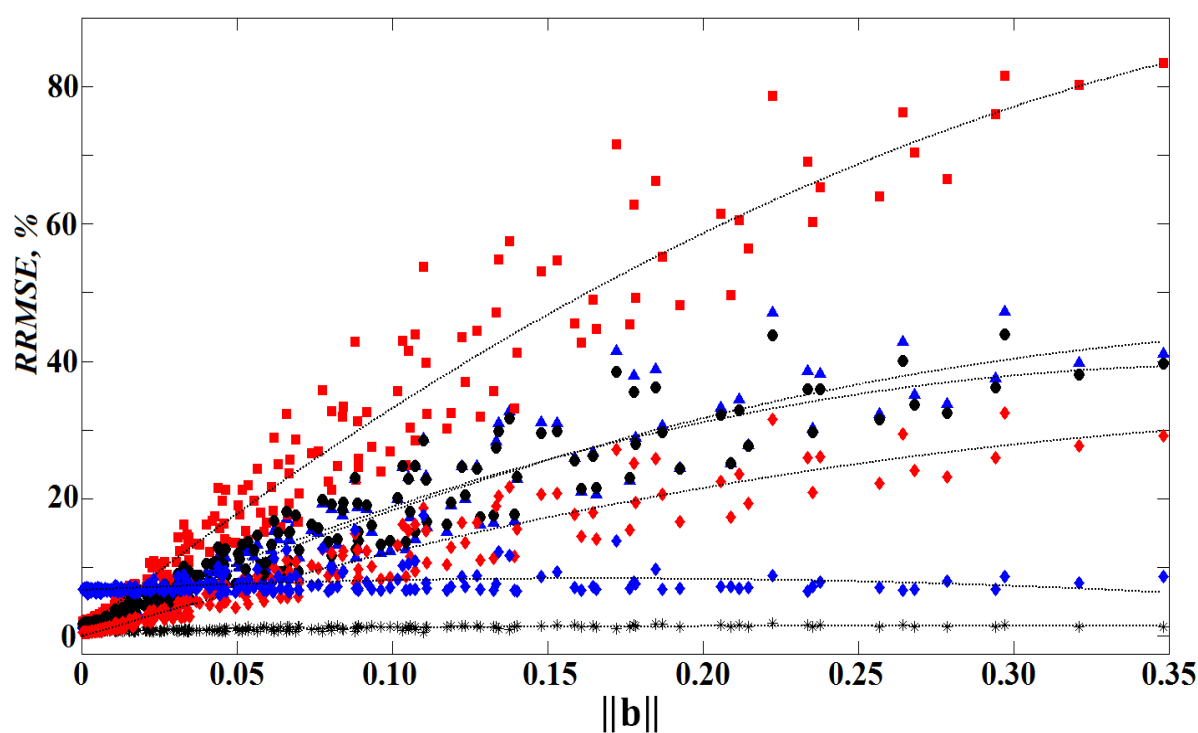


Figure 8 Comparison of the relative error dependences on the Lorentzian background norm for different methods of calculating α .
Data set 1



$\alpha = 0$ (■). Evaluation of α : the L-curve technique (▲), the perturbation theory (●), the fitting procedure (*), PLS (◆), PLS-reduction (◆).
 $\sigma_y = 0.002, \sigma_K = 0.001, c_v = [1/3, 1/3, 1/3]$.

Figure 9 Comparison of the relative error dependences on Lorentzian background norm for different methods of evaluating α .
Data set 2



$\alpha = 0$ (■) ($RRMSE/10$). Evaluation of α : the L-curve method (▲), the perturbation theory (●), the fitting procedure (*), PLS (◆), PLS-truncation (◆). $\sigma_y = 0.002, \sigma_K = 0.001, c_v = [1/3, 1/3, 1/3]$.

Figure 10 Comparison of the relative error dependences on Lorentzian background norm for different methods of evaluating α . Data set 3

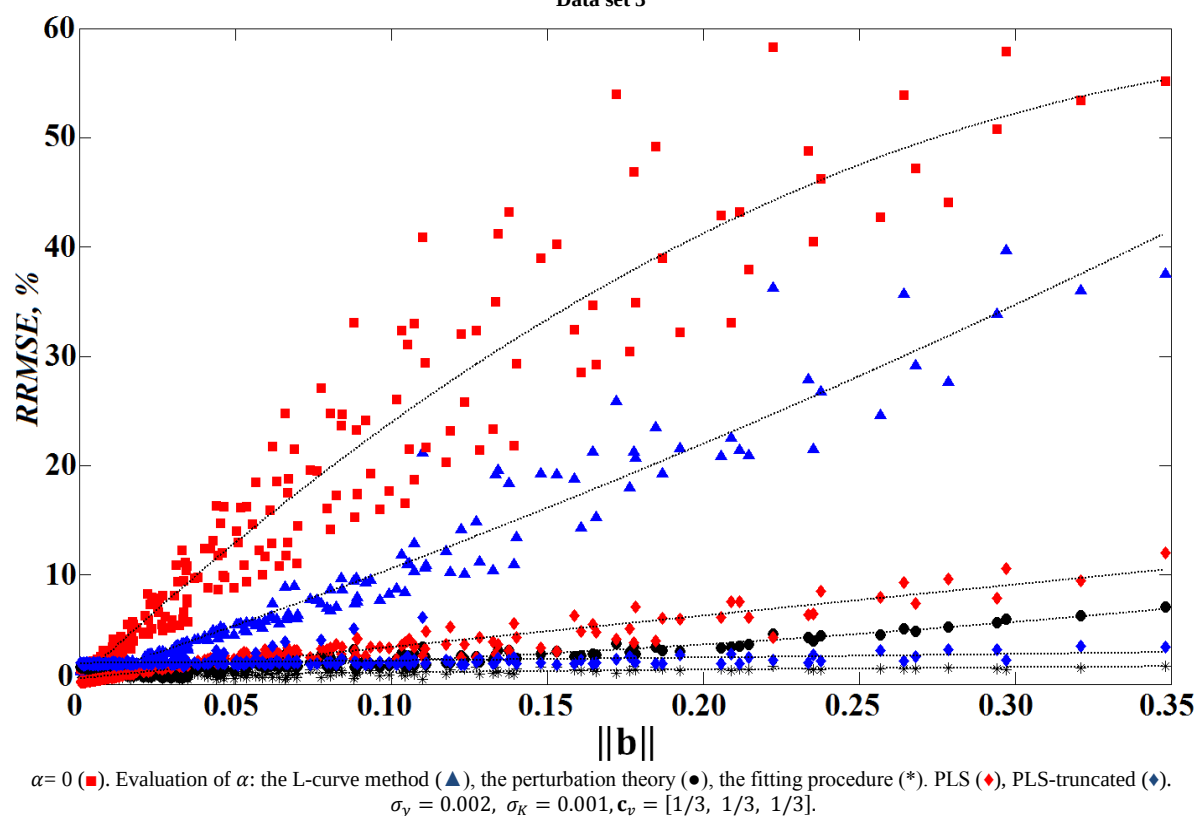


Figure 11 Comparison of the relative error dependences on the polynomial background norm for different processing methods of evaluating α . Data set 4

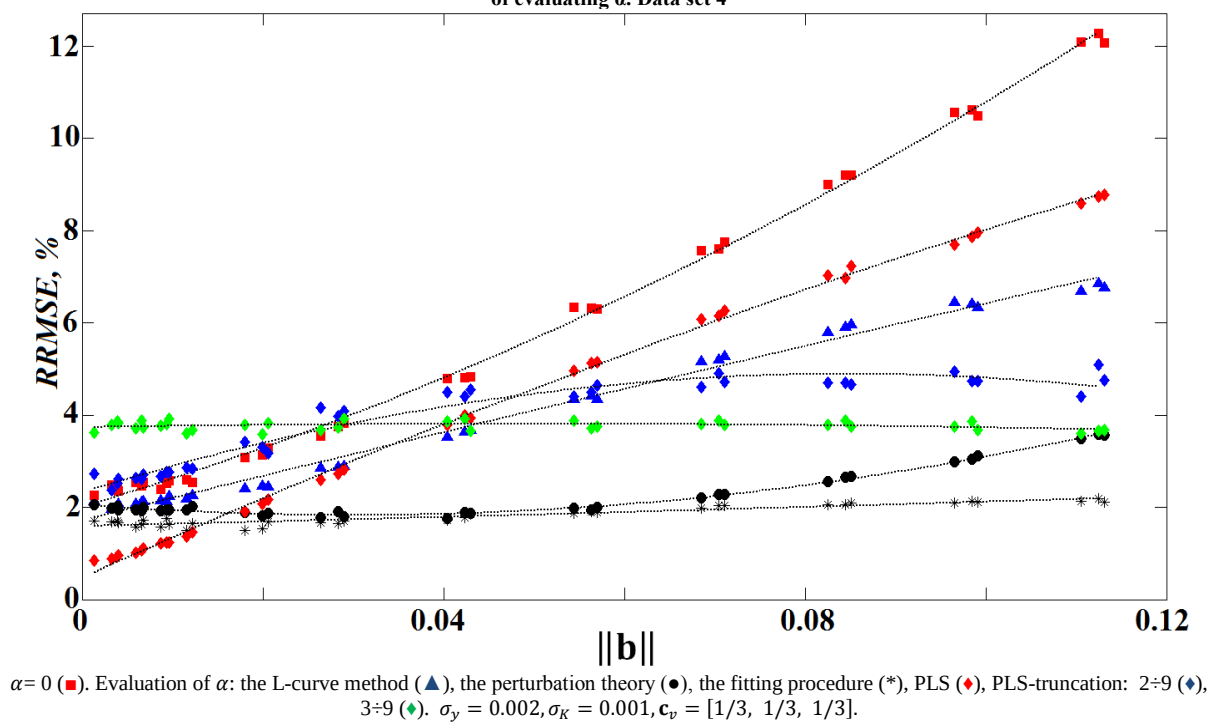
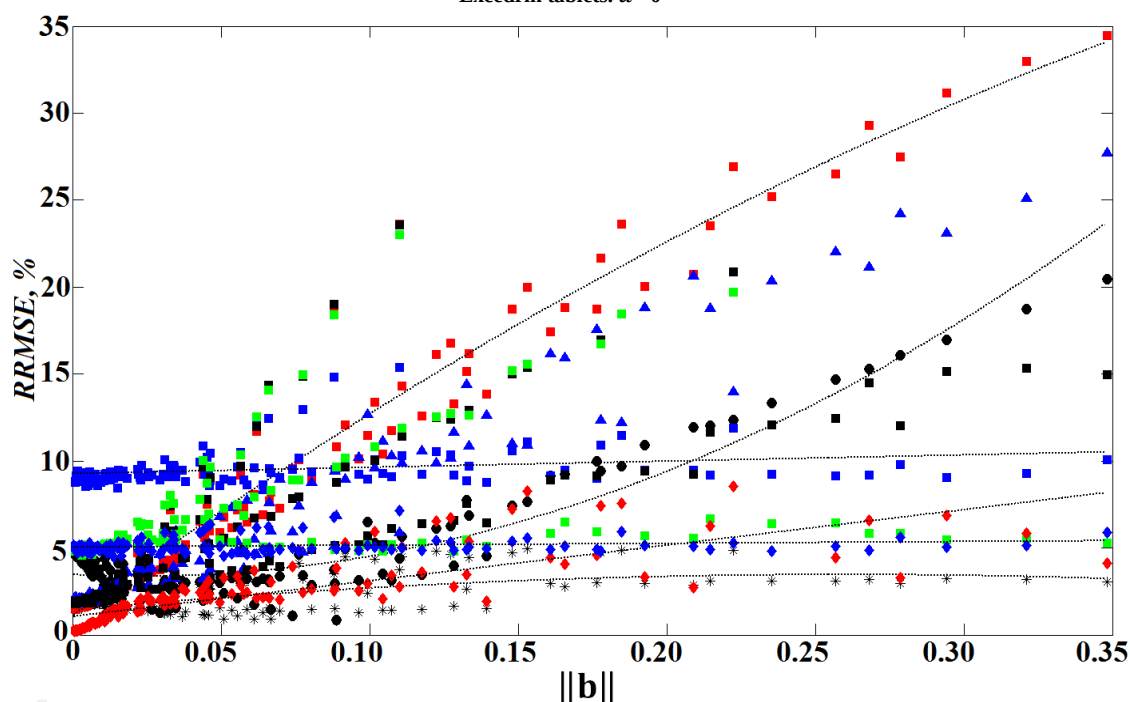


Figure 12 Comparison of the relative error dependences on Lorentzian background norm for different methods of evaluating α . Excedrin tablets. $\alpha = 0$



Processing method: no (■), CN (■), DR (■), AC (■). Evaluation of α : the L-curve method (▲), the perturbation theory (●), the fitting procedure (*), PLS (◆), PLS-truncation (◆). $\sigma_y = 0.002$, $\sigma_K = 0.001$, $c_v = [0.446, 0.108, 0.446]$ for acetaminophen, caffeine and aspirin, respectively.

4. The efficiency of simple background correction methods is low (Fig. 12). Centering the spectrum by subtracting its mean value is effective only for low-intensity backgrounds. The algebraic correction and the first-order differentiation partly suppress the background in the mixture spectrum; on the other hand, the signal-to-noise ratio decreases, which makes the preprocessing much less beneficial. The results obtained in this study are in full agreement with the experimental research [19], which showed that multiplicative scatter correction and first-order differentiation lead to lower prediction accuracies than the PLS regression performed on raw data.

5. For low and middle values of the background norm in the mixture spectra of data set 1, the errors obtained using the PLS-regression and the α -fitting method are close to each other (Fig. 8). For large norms, the PLS regression based on the polynomial reduction is advantageous over the standard PLS procedure (Fig. 8). The minimum achievable RMSE values obtained using the α -fitting method are the lowest limit of the concentration measurement uncertainty for data sets 2, 3, and 4 (Figs. 9-11). The benefits of the α -fitting method as compared to the PLS-regression are most pronounced for data set 2, which has a very large condition number. This result is explained by the fact that the PLS-regression and the regularization techniques have similar filtering properties [20]. However, while filter coefficients in the regularization method smoothly decay to zero for small singular values, the PLS coefficients may oscillate. Therefore, the RRMSE values in the former method are smaller than in the latter.

It can be concluded that the PLS regression based on the polynomial reduction significantly reduces the errors as compared to the non-reduction case, the only exception being very small values of the background norm.

References

- [1]. H. Martens, T. Næs, Multivariate Calibration, New York: Wiley, 1992.
- [2]. G. A. F. Seber and A. J. Lee, Linear Regression Analysis, New York: Wiley, 2003.
- [3]. J.-M. Roger, F. Chauchard, and V. Bellon-Maurel, "EPO-PLS external parameter orthogonalization of PLS application to temperature-independent measurement of sugar content of intact fruits", Chemom. Intell. Lab. Syst. vol. 66, 2003, pp. 191-204.
- [4]. Á. Rinan, F. van den Berg, and S. B. Engelsen, "Review of the most common pre-processing techniques for near-infrared spectra", Trends in Analytical Chemistry, vol. 28, 2009, pp. 1201-1222.
- [5]. D. M. Haaland, R. G. Easterling, "Improved sensitivity of infrared spectroscopy by the application of least squares methods", Appl. Spectrosc., vol. 34, 1980, pp. 539-546.
- [6]. R. J. Barnes, M. S. Dhanoa, and S. J. Lister, "Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra", Appl. Spectrosc., vol. 43, 1989, pp. 772-777.
- [7]. J.-C. Boulet and J.-M. Roger, "Pretreatments by means of orthogonal projections", Chemom. Intell. Lab. Syst. vol. 117, 2012, pp. 61-69.
- [8]. P. Geladi, D. MacDougall, and H. Martens, "Linearization and scatter-correction for near-infrared reflectance spectra of meat", Appl. Spectrosc., vol. 39, 1985, pp. 491-500.

- [9]. J. M. Dubrovkin, (1988). "Chemical analysis with the use of the mathematical differentiation", *Zhurnal Analiticheskoi Khimii*, vol. 43, 1988, pp. 965-979. [Russian].
- [10]. C. D. Brown, L. Vega-Montoto, and P. D. Wentzell, "Derivative preprocessing and optimal corrections for baseline drift in multivariate calibration", *Appl. Spectr.*, vol. 54, 2000, pp. 1055 -1068.
- [11]. U. Depczynski, K. Jetter, K. Molt, and A. Niemöller, "Quantitative analysis of near infrared spectra by wavelet coefficient regression using a genetic algorithm", *Chemom. Intell. Lab. Syst.*, vol. 47, 1999, pp. 179–187.
- [12]. J. Engel, J. Gerretzen, E. Szymańska, J. J. Jansen, G. Downey, L. Blanchet, and L. M. C. Buydens, "Breaking with trends in pre-processing?", *Trends in Analytical Chemistry*, vol. 50, 2013, pp. 96-106.
- [13]. Lu Xu, Yan-Ping Zhou, Li-Juan Tang, Hai-Long Wu, Jian-Hui Jiang, Guo-Li Shen, and Ru-Qin Yu, "Ensemble preprocessing of near-infrared (NIR) spectra for multivariate calibration", *Anal. Chim. Acta*, vol. 616, 2008, pp. 138-143.
- [14]. J. M. Dubrovkin and V. A. Anan'ev, "Error estimation in the quantitative spectrochemical analysis of the multicomponent mixtures in the presence of the randomly varying spectral baseline", *Bulletin of Kemerovo State University*, vol. 55, 2013, pp. 232-236. [Russian].
- [15]. C. Lanczos, *Applied analysis*, New York: Dover Publication, Inc, 1988.
- [16]. P. C. Hansen, "Analysis of discrete ill-posed problems by means of the L-curve", *SIAM Rev.*, vol. 34, 1992, pp. 561–580.
- [17]. J. M. Dubrovkin, "Estimation of the minimum uncertainty of quantitative spectroscopic analysis", *Bulletin of Slovak Spectroscopic Society*, vol. 19, 2012, pp. 4-7.
- [18]. V. Voevodin and Yu. A. Kuznetsov, *Matrices and calculations*. Moscow: Nauka, 1984. [Russian].
- [19]. A. Bogomolov and A. Melenteva, "Scatter-based quantitative spectroscopic analysis of milk fat and total protein in the region 400-1100 nm in presence of fat globule size variability", *Chemom. Intell. Lab. Syst.*, vol. 126, 2013, pp. 129-139.
- [20]. E. Andries and J. H. Kalivas, "Multivariate calibration leverages and spectral *F*-ratios via the filter factor representation", *J. Chemometrics*, vol. 24, 2010, pp 249-260.