



Error Propagation of Quantitative Analysis Based on Ratio Spectra

J. Dubrovkin

Computer Department, Western Galilee College
2421 Acre, Israel

Abstract: Error propagation of the quantitative analysis of binary and ternary mixtures based on the ratio spectra and the mean-centred ratio spectra has been studied. Gaussian doublets and triplets were used as models of the mixture pure-component spectra. The mixture spectra were disturbed by random constant and proportional noises and unknown background. The perturbations of the calibration matrix were modelled by systematic errors caused by the wavelength shifts. The least-squares estimation of the concentration vector and the estimation errors were obtained theoretically and numerically. The condition number of the matrix of the pure-component ratio spectra was theoretically evaluated for binary mixtures. The advantages and disadvantages of the ratio spectra methods are discussed.

Keywords: quantitative spectrochemical analysis, ratio spectra, mean centred ratio spectra, errors, condition number, random constant and proportional noise, unknown background.

I. Introduction

One of the main problems of the spectrochemical analysis of white multicomponent mixtures is overlapping of pure-component spectra which are known *a priori*. In the period of more than half a century, analysts have attempted to solve this problem by developing numerous smart mathematical algorithms for processing the mixture spectrum in conjunction with physical-chemical treatment of the mixture to be analyzed [1]. These algorithms may be divided into two main groups:

1. Allocation of the analytical points set (or its linear transforms) in the mixture spectrum free of overlapping with respect to a given analyte.
2. Direct and inverse calibration methods based on solving linear equation systems for data sets of calibration mixtures with known spectra and concentrations.

The most popular methods of the first group include derivative spectroscopy [2], the method of orthogonal transforms ("the net analyte signal") [3, 4], and different modifications of the optical density ratio method [1, 5-15].

A major progress in developing the methods of the second group was achieved in 1980s due to applying statistical methods of chemometrics (regularization, principal component analysis, and partial least squares regression) to combined analytical data obtained by spectroscopic and non-spectroscopic measurements [16]. The success of this approach is attributed to using increased information for analytical purposes. Unfortunately, in some real-life cases, mixtures which contain different concentrations of pure components are not available and/or the preparation of artificial mixtures is too complicated. The analysis of medicines with claimed compositions also requires special approach [17].

In view of the above, many researchers attempted to improve the "old" analytical methods developed in the "pre-computer era" by using modern instrumentation and computational tools. An interesting practical applications in this field is the centering modification of the ratio spectra (RS) method [5] (MCRS method [6-15]). However, its effectiveness was proved only by experimental studies. The goal of our work was giving rigorous mathematical foundation for studying the noise-filtering properties (error propagation) of the method.

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

II. Theory

Consider the spectrum of an additive binary mixture that obeys the Beer-Bouguer-Lambert law:

$$\mathbf{a} = \mathbf{S}\mathbf{c}^T, \quad (1)$$

where $\mathbf{a}$ is a column vector ($m \times 1$), $\mathbf{S} = [\mathbf{s}_1 \ \mathbf{s}_2]$, $\mathbf{s}_1$ and $\mathbf{s}_2$ are the column vectors ($m \times 1$) which represent the spectra of the first and the second mixture components, respectively, $\mathbf{c} = [c_1 \ c_2]$ is the concentration vector, c_1 and c_2 are the component concentrations, and T is the symbol for transpose. The path length is assumed to be equal to unit. Suppose that the error-free pure-component spectra are *a priori* known.

Multiplying Eq. 1 by the diagonal matrix

$$\mathbf{D}_1 = \text{diag} \left(\frac{1}{s_{11}}, \frac{1}{s_{12}}, \dots, \frac{1}{s_{1m}} \right), \quad (2)$$

where $s_{i1} \neq 0$ is the i -element of $\mathbf{s}_1$, we obtain the ratio spectrum:

$$\mathbf{D}_1 \mathbf{a} = c_1 \mathbf{1} + c_2 \mathbf{D}_1 \mathbf{s}_2, \quad (3)$$

where $\mathbf{1}$ is the unit vector ($m \times 1$). Constant term $c_1 \mathbf{1}$ can be eliminated by the mean centering of Eq. 3 (subtracting its average):

$$\mathbf{D}_1 \mathbf{a} - \langle \mathbf{D}_1 \mathbf{a} \rangle \mathbf{1} = c_2 \mathbf{z}_{12}, \quad (4)$$

where chevron is the mean value symbol and $\mathbf{z}_{12} = \mathbf{D}_1 \mathbf{s}_2 - \langle \mathbf{D}_1 \mathbf{s}_2 \rangle \mathbf{1}$. Matrix equation (4) consists of m linear equations for each wavelength. Earlier it was suggested to evaluate unknown concentration (c_2) at one optimal wavelength (e.g., at the point of the $\mathbf{D}_1 \mathbf{a}$ maximum) by linear calibration procedure using standard mixtures [6]. Such calibration can significantly reduce the systematic errors [18]. On the other hand, the single-point analysis results in the losses of information that is contained in the rest of the analytical points. Therefore, we prefer to solve Eq. 4 by the least squares (LS) method using the “best” combination of the analytical wavelengths (e.g., [19]).

The LS-solution of Eq. 4 is [20]

$$\hat{c}_2 = (\mathbf{D}_1 \mathbf{a} - \langle \mathbf{D}_1 \mathbf{a} \rangle \mathbf{1})^T \mathbf{l}_{12}, \quad (5)$$

where $\mathbf{l}_{12} = \mathbf{z}_{12} / (\mathbf{z}_{12}^T \mathbf{z}_{12})$.

Since $\mathbf{1}^T \mathbf{l}_{12} = 0$,

$$\hat{c}_2 = \mathbf{a}^T \mathbf{k}_2^{RM}, \quad (6)$$

where $\mathbf{k}_2^{RM} = \mathbf{D}_1 \mathbf{l}_{12}$. Similarly,

$$\hat{c}_1 = \mathbf{a}^T \mathbf{k}_1^{RM}, \quad (7)$$

where $\mathbf{k}_1^{RM} = \mathbf{D}_2 \mathbf{l}_{21}$, $\mathbf{D}_2 = \text{diag} \left(\frac{1}{s_{21}}, \frac{1}{s_{22}}, \dots, \frac{1}{s_{2m}} \right)$, $\mathbf{l}_{21}$ and $\mathbf{z}_{21}$ are obtained by changing the indexes in Eqs. 4 and 5.

The above algebraic operations represent, actually, linear transformation of Eq. 1. According to statistical concepts [20], in the presence of uncorrelated normal noise (perturbation) with zero mean, the LS-estimate is the best linear unbiased estimate with minimal dispersion, which cannot be decreased by any linear transformations. However, for other kinds of noise (e.g., proportional), this conclusion is not valid.

The LS-solution of Eq.1 gives two calibration vectors ($j = 1, 2$):

$$\mathbf{k}_j^{LS} = \mathbf{s}_j (\mathbf{S}^T \mathbf{S})^{-1}, \quad (8)$$

which differ from the corresponding vectors obtained by the MCRS method (Eqs. 6 and 7).

If a mixture spectrum contains uncorrelated normal noise with zero mean and constant dispersion σ^2 , the mean squared error of the LS estimation of the component concentration vector $[\hat{c}_1, \hat{c}_2]$ depends on the sum of the diagonal elements of inverse matrix (8) [20]:

$$MSE_{LS}^{const} = \{ \text{diag}[(\mathbf{S}^T \mathbf{S})^{-1}]_1 + \text{diag}[(\mathbf{S}^T \mathbf{S})^{-1}]_2 \} \sigma^2. \quad (9)$$

If the standard deviation of the noise is proportional to the response of the spectral instrument,

$$MSE_{LS}^{prop} = \sum_{i=1}^m (\gamma a_i)^2 \{ [(\mathbf{k}_1^{LS})_i]^2 + [(\mathbf{k}_2^{LS})_i]^2 \}, \quad (10)$$

where γ is the coefficient of the noise.

Similarly, if the MCRS method is used,

$$MSE_{RM}^{const} = \sum_{i=1}^m \{ [(\mathbf{k}_1^{RM})_i]^2 + [(\mathbf{k}_2^{RM})_i]^2 \} \sigma^2, \quad (11)$$

$$MSE_{RM}^{prop} = \sum_{i=1}^m (\gamma a_i)^2 \{ [(\mathbf{k}_1^{RM})_i]^2 + [(\mathbf{k}_2^{RM})_i]^2 \} \quad (12)$$

To compare the LS and the MCRS methods, we used the standard deviation (std) ratio:

$$\rho = \sqrt{MSE_{RM} / MSE_{LS}}. \quad (13)$$

The mathematical analysis of the mean centering of ternary mixture ratio spectra is similar to the corresponding analysis for binary mixtures (see APPENDIX A). In the case of ternary mixtures, the error term of the third component will appear in Eqs. 9-12.

The error propagation in multicomponent analysis can be studied using the condition number of matrix $\mathbf{S}$, $\text{cond}(\mathbf{S})$. It is known that the relative prediction uncertainty in quantitative analysis [21] is

$$\frac{\|\delta c\|}{\|c\|} \leq \text{cond}(\mathbf{S}) \left(\frac{\|\delta \mathbf{S}\|}{\|\mathbf{S}\|} + \frac{\|\delta \mathbf{a}\|}{\|\mathbf{a}\|} \right), \quad (14)$$

where $\frac{\|\delta S\|}{\|S\|}$ and $\frac{\|\delta a\|}{\|a\|}$ are the relative uncertainties of matrix S (calibration errors) and of the mixture spectra (measurement errors), respectively. Since the theoretical calculation of the condition number is possible only in some simple cases (see APPENDIX B), this number is generally evaluated numerically by computer modeling.

III. Computer modeling

A. Binary mixtures

To perform computer modeling, the components of a symmetrical Gaussian doublet (Fig. 1) were chosen as elements of matrix S :

$$s_{1i} = \exp(-(x_i + x_0)^2), s_{2i} = \exp(-(x_i - x_0)^2), \quad (15)$$

where $x_i = \beta_G v_i / w$, $x_0 = \beta_G v_0 / w$, $\beta_G = 2\sqrt{\ln 2}$, v_i is the abscissa of the spectrum plot (e.g., wavelength), $i = -m, -m+1, \dots, 0, 1, \dots, m$, w is full width at half maximum of Gaussian lines, and $\pm v_0$ are the positions of the component maxima.

The condition number of ratio spectra matrix $S_{rs} = D_1 S$ was evaluated (APPENDIX B) as

$$\text{cond}(S_{rs}) \approx \frac{\sqrt{12}}{p} \left(1 + \frac{p^2}{8}\right), \quad (16)$$

where $p = 4\text{Lim}x_0$, $\text{Lim} = m\Delta x$, and Δx is the sampling interval along x -axis. Since the $\text{cond}(S)$ value could not be calculated analytically, the ratio

$$\rho_{\text{cond}} = \text{cond}(S_{rs}) / \text{cond}(S) \quad (17)$$

was calculated numerically (Fig. 2).

From the curves shown in Fig. 2, it can be concluded that the RS method is a little less error-sensitive than the LS method ($\rho_{\text{cond}} < 1$) only for a large number of analytical points in the case of a strongly overlapping Gaussian doublet ($x_0 < 0.3$). For a resolved doublet ($x_0 > 0.4$), the RS method is not effective.

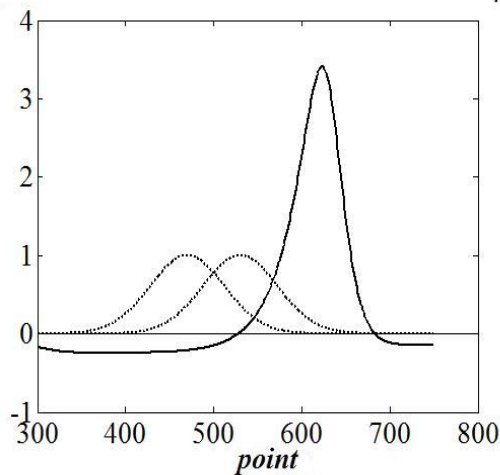
The standard deviation ratios (Eq. 13) were evaluated for different sets of analytical points. The chosen sets were located in the neighborhood of the doublet middle point (Fig. 3) and symmetrically around the point 598 of the doublet MCRS (Fig.1). Since the wings of Gaussian lines quickly decay to zero, the intensity of the MCRS approaches infinity. To partly compensate for this drawback, very small constant background (0.001) was added to the doublet components (Eq. 15).

Based on the results presented in Figs. 4 and 5, it can be concluded that it is more advantageous to apply the LS method to the MCRS than to the original spectra only in the case of proportional noise (which is typical of UV-VIS instruments). The sets of analytical points were identical for both LS and MCRS methods.

The systematic errors caused by uncompensated second-order polynomial background in the mixture spectrum, were evaluated numerically for both methods. The obtained ρ values were close to 1.

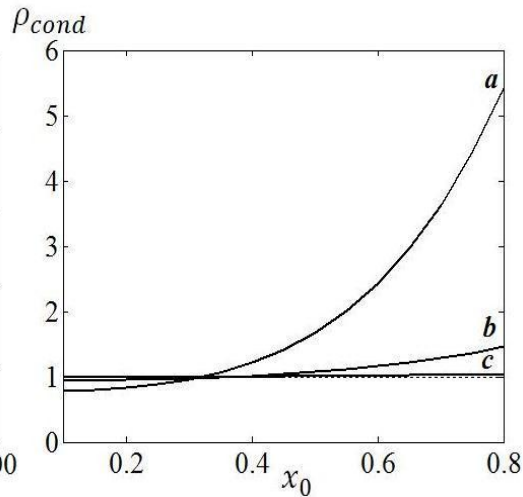
Another type of errors is caused by the systematic errors in the matrix of the pure-component spectra. While the random errors of matrix S can be significantly decreased by averaging, the systematic errors cannot be eliminated. The main source of the latter type errors is the shift of the spectral points from their original position

Figure 1. Symmetrical Gaussian doublet with constant base line and its MCRS



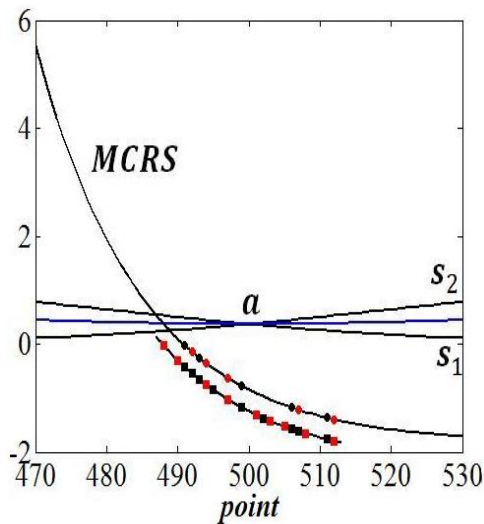
(...) Doublet ($x_0 = 0.5$, base line = 0.001);
(—) the MCRS.

Figure 2. Dependences of relative condition number on x_0 for different integration limits



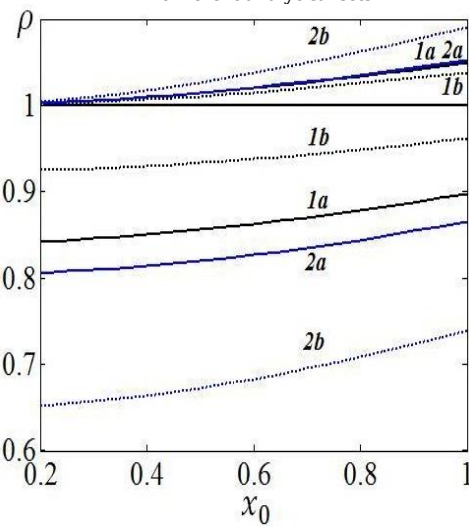
Lim: (a) 1.0, (b) 0.5 and (c) 0.2.

Figure 3. Analytical points for binary mixtures.



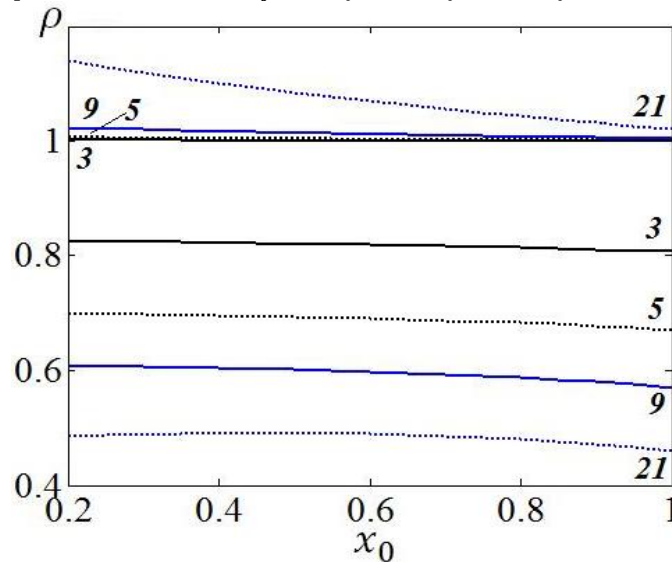
Mixture (*a*) and pure-component (*s*₁ and *s*₂) spectra.
Five-point (circles) and nine-point analytical sets (rectangles): 1a, 1b, 2a and 2b; a - black, b-red.

Figure 4. Dependences of std ratio on x_0 for different analytical sets



Constant and proportional noise (above and below the line $\rho = 1$, respectively). The designations of the analytical sets are the same as in Fig. 3.

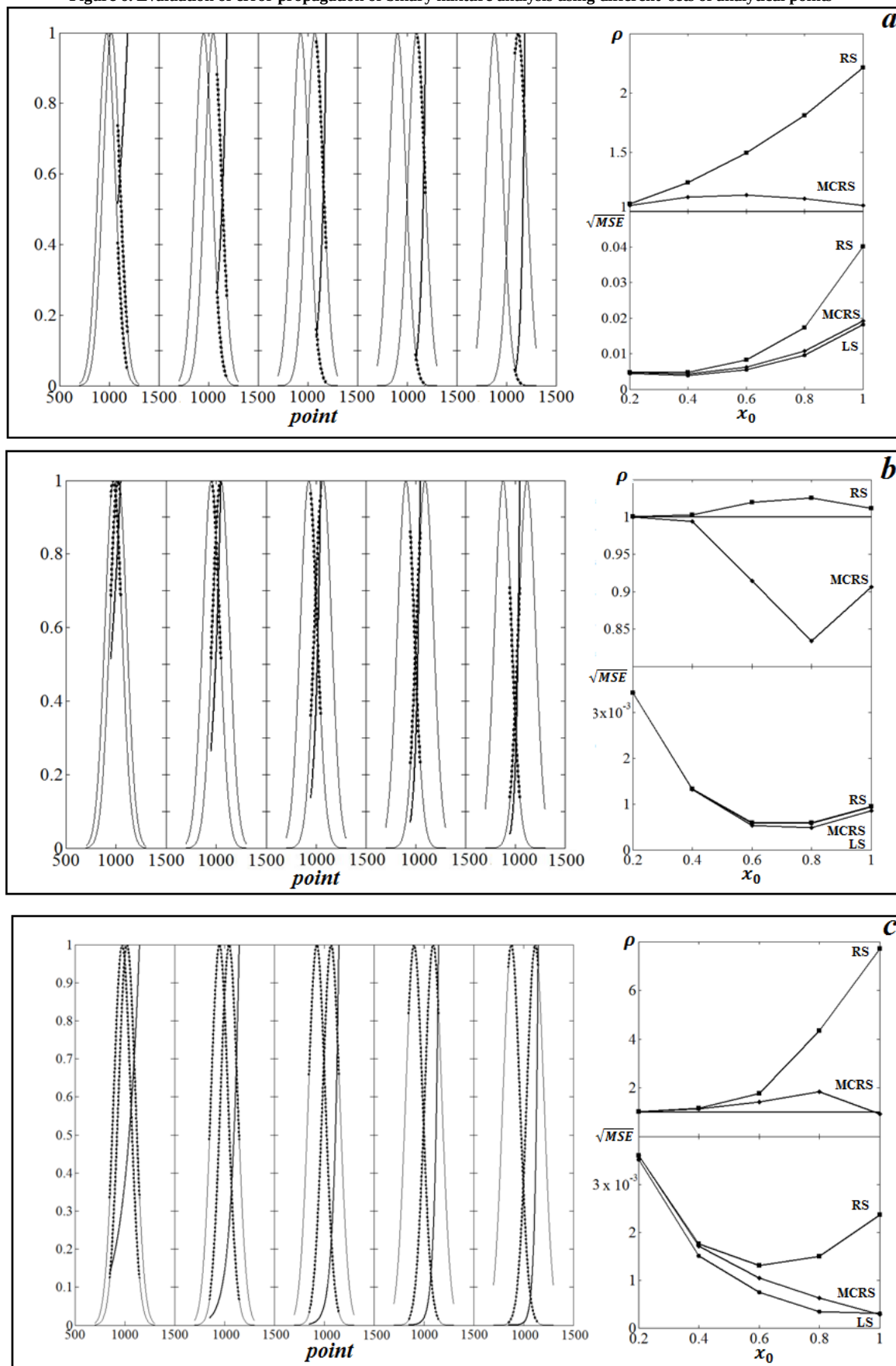
Figure 5. Dependences of std ratio on x_0 for analytical sets symmetrically located around point 598

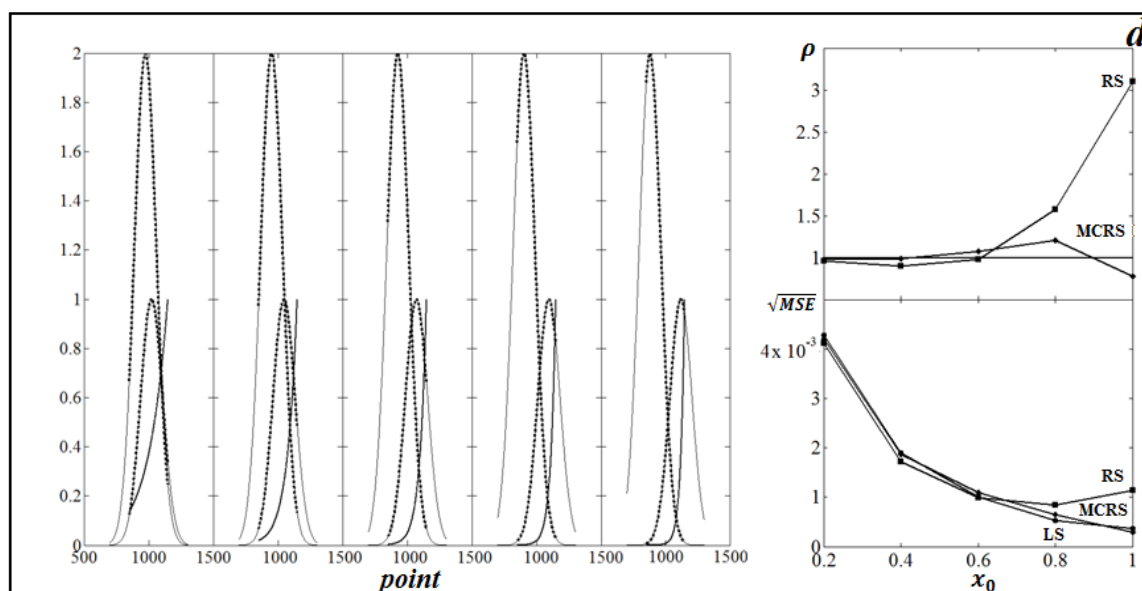


Constant and proportional noise (above and below the $\rho = 1$ line, respectively). The set size is given next to each plot.

along x -axis This shift depends on the slope of the spectral curve [1]. The theoretical study of the impact of the wavelength shift in MCRS on the concentration errors is presented in APPENDIX (Eq. C7). It was shown that the errors of the calculated concentrations strongly depend on the derivatives (slopes) of the mixture spectrum and of the pure-component spectra. Thus the metrological characteristics of the RS method can be improved by precise setting of the mixture spectrum and of the pure-component spectra at the same wavelengths. This result is in agreement with that of the well-known study of the influence of the calibration spectra wavelength shift on the validity of multivariate calibration models [22, 23]. The impact of wavelength shift on the uncertainty of the LS-based analysis of spectra and of the corresponding ratio spectra (including MCRS) was also evaluated numerically. The analytical points were chosen in the range of steep slope of the spectra (Fig. 6, the sampling interval was halved). The obtained results show that, in some cases, the mean centering procedure significantly decreases the RS analysis uncertainty (Fig. 6). However, the MCRS method has no advantages over the common LS method. Moreover, the selection of analytical points is a critical factor of the analysis. For example, for symmetrical distribution of the analytical points around the point 598 (Fig.1), the LS method gives better results than the MCLS one, especially for large data sets (Fig. 7).

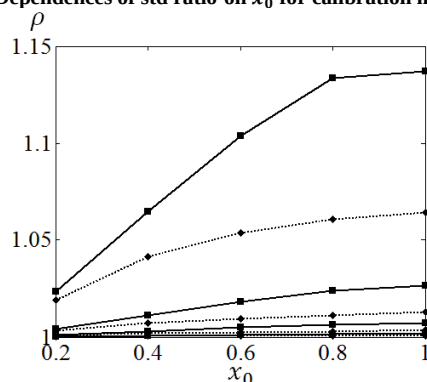
Figure 6. Evaluation of error propagation of binary mixture analysis using different sets of analytical points





Left-hand panels: Sets of the analytical points of matrix $\mathbf{S}$ (···) and of the first doublet component RS (—). Ranges of analytical points: (a) 1090 - 1190, (b) 950 - 1050, (c, d) 850 - 1150. $r = 1$, $R = 1$ (a-c) and 2 (d). Right-hand panels: $\rho(x_0)$ (top) and $\sqrt{MSE}(x_0)$ (bottom) plots.

Figure 7. Dependences of std ratio on x_0 for calibration matrix errors



LS (—) and MCRS (···) methods. The analytical sets are located symmetrically around point 598 (Fig.1). The set sizes are 3, 5, 9, and 21 from bottom to top, respectively.

B. Ternary mixtures

The components of two Gaussian triplets were taken as elements of two matrices $\mathbf{S}$ (Fig. 8, a1 and a2, top plots). The condition number of the first triplet, $\text{cond}(\mathbf{S}) = 91$, is very large due to strong overlapping of the pure-component spectra. Overlapping in the second triplet is small, consequently, $\text{cond}(\mathbf{S}) = 6.4$. From the full-region MCRSs (Fig. 8 a1 and a2, bottom plots), “the best” combination of analytical points was selected empirically (Fig. 8, b1-d1, b2-d2). The points were selected in the regions of the maximum intensity of the transformed spectra, according to the minimum error criteria for evaluated concentrations. The total prediction error (Eq. C9) was estimated by averaging over 100 statistically independent numerical experiments using a 95%-confidence interval.

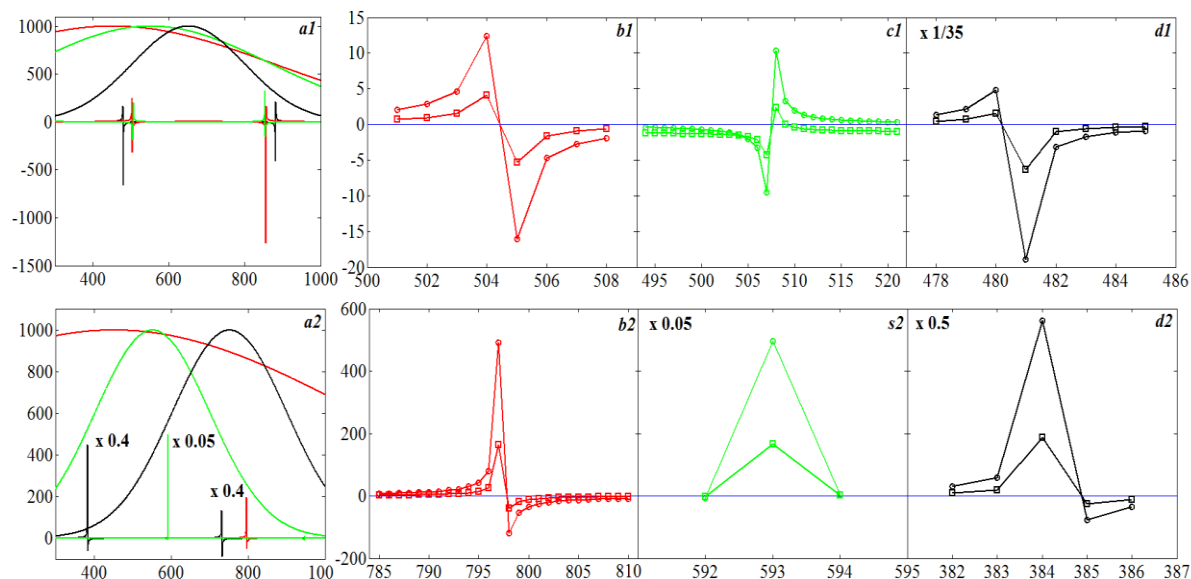
The results presented in Fig. 8 show that the advantage of the MCRS method is significant compression of spectral data. In other words, this method allows replacing full-range spectra by a relative small number of analytical points. However, the intensities of the transformed spectra decrease notably, which results in an increased impact of small errors in MCRS on the of the quantitative analysis uncertainty. It was found that data compression is achieved at the expense of biased estimation of the second component and large increase of the total prediction error of the component concentrations for ternary mixtures.

The prediction errors, calculated for 16 ternary mixtures (Table 1), are listed in Table 2. According to these results, the errors of the LS analysis based on full-range spectra are significantly less than those of the MCRS method. However, in the same spectral region, the MCRS method is more preferable than the LS one in the case of strongly overlapping pure-component spectra.

The most critical factor for the MCRS method are systematic errors of the pure-component spectra matrix. For example, for $\gamma = 0.05$, the relative errors of the total prediction error are more than 100%.

In conclusion, it should be pointed out that the MCRC method employs a very limited number of analytical points, whose location has been *a priori* chosen. In contrast to this, in the multivariate regression method, all possibly relevant spectral and non-spectral data are pooled together. Therefore, generally speaking, the latter method is preferred to the MCRC analysis. However, consuming large data volumes from analytical instruments creates a new data-management level. In this connection, traditional spectrochemical professionals often prefer "single-point" analytical methods to complex multi-point procedures.

Figure 8. Gaussian triplet component spectra and their full-region MCRC.



(a1, a2) Component spectra (top plots) and MCRC (bottom plots). Analytical point sets for: (b1-d1) a1 triplets and for (b2-d2) a2 triplets, respectively.

Table 1. Mixture concentrations

c_1	0.05	0.05	0.05	0.1	0.1	0.8	0.1	0.2	0.7	0.1	0.3	0.6	0.2	0.2	0.6	1/3
c_2	0.05	0.9	0.05	0.1	0.8	0.1	0.2	0.1	0.7	0.3	0.1	0.6	0.2	0.6	0.2	1/3
c_3	0.9	0.05	0.9	0.8	0.1	0.1	0.7	0.2	0.1	0.6	0.3	0.1	0.6	0.2	0.2	1/3

Table 2. Total prediction errors, %

Disturbance		LS (full)	LS (part)	MCRC
Constant noise ($\sigma = 1$)		0.35±0.13	83±30	5.0±1.6
		0.027±0.091	0.20±0.071	0.41±0.12
Proportional noise ($\gamma = 0.001$)		0.25±0.088	67±27	4.3±1.3
		0.017±0.0059	0.15±0.054	0.31±0.096
Constant noise ($\sigma = 1$) Background ($a_0 = 1$)		0.41±0.14	79±29	5.5±1.7
		0.16±0.059	0.25±0.081	0.46±0.14
Wavelength shift	$\gamma = 0.01$	0.019±0.0075	0.057±0.025	0.30±0.12
	$\gamma = 0.02$	0.037±0.0144	0.12±0.056	0.61±0.25
		0.016±0.0065	0.016±0.0071	0.070±0.018
	$\gamma = 0.05$	0.098±0.039	0.28±0.13	64±128
		0.040±0.016	0.041±0.017	0.70±1.3

Data for two calibration sets (Fig. 8, a1 and a2) are given in the upper and lower rows, respectively.

References

- [1] I.Ya. Bernstein and Yu.L. Kaminsky, Spectrophotometric Analysis in Organic Chemistry. Leningrad: Science, 1986.
- [2] J. M. Dubrovkin, V. G. Belikov, Derivative Spectroscopy. Theory, Technics, Application. Russia: Rostov University, 1988.
- [3] J. M. Dubrovkin, "The possibility of the discrete Fourier transform-based quantitative analysis for overlapping absorption bands", Izvestia Severo-Kavkazskogo Nauchnogo Centra Vyshey Shkoly, Yestestvennye Nauki, N1, 1981, pp. 57-60.

- [4] A. Lorber, "Error propagation and figures of merit for quantification by solving matrix equations", Anal. Chem., vol. 58, 1986, pp. 1167-1172
- [5] M. J. S. Dewar and D. S. Urch, "Electrophilic substitution. Part VIII. The nitration of dibenzofuran and a new method of ultraviolet spectrophotometric analysis of mixtures", J. Chem. Soc., N1, 1957, pp. 345-347.
- [6] A. Afkhami and M. Bahram, "Mean centering of ratio kinetic profiles as a novel spectrophotometric method for the simultaneous kinetic analysis of binary mixtures", Anal. Chim. Acta, vol. 526, 2004, pp. 211-218.
- [7] A. Afkhami and M. Bahram, "Mean centering of ratio spectra as a new spectrophotometric method for the analysis of binary and ternary mixtures", Talanta, vol. 66, 2005, pp. 712-720.
- [8] H. M. Lotfy and M. A. Hegazy, "Comparative study of novel spectrophotometric methods manipulating ratio spectra: An application on pharmaceutical ternary mixture of omeprazole, tinidazole and clarithromycin", Spectrochim. Acta, A: Molecular and Biomolecular Spectrosc., 96, 2012, pp. 259-270.
- [9] E. A. Abdelaleem and N. S. Abdelwahab, "Simultaneous determination of some antiprotozoal drugs in different combined dosage forms by mean centering of ratio spectra and multivariate calibration with model updating methods", Chem. Central J., vol. 6:27, 2012, pp.1-8.
- [10] M. M. Issa, R. M. Nejeh, A. M. Abu Shanab and N. T. Shaat, "Resolution of five-component mixture using mean centering ratio and inverse least squares", Chem. Central J., vol. 7:152, 2013, pp.1-11.
- [11] H. W. Darwish, S. A. Hassan, M. Y. Salem and B. A. El-Zeiny, "Three different spectrophotometric methods manipulating ratio spectra for determination of binary mixture of Amlodipine and Atorvastatin", Spectrochim. Acta, A: Molecular and Biomolecular Spectrosc., 83, 2011, pp. 140-148.
- [12] N. M. Bhatt, V. D. Chavada, M. Sanyal and P. S. Shrivastav, "Manipulating Ratio Spectra for the Spectrophotometric Analysis of Diclofenac Sodium and Pantoprazole Sodium in Laboratory Mixtures and Tablet Formulation", The Scientific World Journal, Vol. 2014, 2014, Article ID 495739, 10 pp.
- [13] H. M. Lotfy, N. Y. Hassan, S. M. Elgizawy and S. S. Saleh, "Comparative study of new spectrophotometric methods: an application on pharmaceutical binary mixture of ciprofloxacin hydrochloride and hydrocortisone", J. Chil. Chem. Soc. vol.58 , 2013, pp. 1651-1657 .
- [14] H. M. Lotfy, M. A.I. M. Hegazy and S. A. N. Abdel-Gawad, "Simultaneous determination of Simvastatin and Sitagliptin in tablets by new univariate spectrophotometric and multivariate factor based methods", European J. Chem. vol. 4 ,2013, pp. 414-421.
- [15] N. W. Ali, M. A. Hegazy, M. Abdelkawy and E. A. Abdelaleem, "Simultaneous determination of methocarbamol and ibuprofen or diclofenac potassium using mean centering of the ratio spectra method", Acta Pharm. vol. 62, 2012, pp. 191-200.
- [16] H. Martens and T. Næs, Multivariate Calibration. New York: Wiley, 1992.
- [17] "US Pharmacopeia <197> Spectrometric identification tests", http://www.pharmacopeia.cn/v29240/usp29nf24s0_c197.html
- [18] J. Dubrovkin, "Evaluation of the maximum achievable information content in quantitative spectrochemical analysis", International Journal of Emerging Technologies in Computational and Applied Sciences, vol. 1-6, 2013, pp. 1-7.
- [19] J. M. Dubrovkin, "Quantitative analytical spectrometry of multicomponent systems with known quantitative composition using the orthogonal projection method", Zhurnal Prikladnoi Spektroskopii, vol. 50, 1989, pp. 861-864.
- [20] G. A. F. Sever and A. J. Lee, Linear Regression Analysis, 2nd ed., New Jersey: John Wiley and Sons, 2003.
- [21] K. Danzer, M. Otto, L. A. Currie, "Guidelines for calibration in analytical chemistry. Part 2. Multispecies calibration (IUPAC Technical report)", Pure and Applied Chemistry, vol. 76, 2004, pp. 1215-1225.
- [22] H. Swierenga A.P. de Weijer, R.J. van Wijk b and L.M.C. Buydens, "Strategy for constructing robust multivariate calibration models", Chem. Intel. Lab. Syst., vol. 49, 1999, pp. 1-17.
- [23] F. Vogt and K. Booksh, "Influence of wavelength-shifted calibration spectra on multivariate calibration models", Appl. Spectrosc., 2004, vol. 58, pp.624-635.

Appendix

A. Mean centering of ratio spectra for ternary mixtures [6].

Similar to the case of binary mixtures (Eq. 1), consider the spectrum of an additive ternary mixture which obeys the Beer-Bouguer-Lambert law:

$$\mathbf{a} = c_1 \mathbf{s}_1 + c_2 \mathbf{s}_2 + c_3 \mathbf{s}_3. \quad (A1)$$

1st step: multiplication of Eq. A1 by matrix $\mathbf{D}_1(2)$:

$$\mathbf{D}_1 \mathbf{a} = c_1 \mathbf{1} + c_2 \mathbf{D}_1 \mathbf{s}_2 + c_3 \mathbf{D}_1 \mathbf{s}_3 \quad (A2)$$

and mean centering of (A2):

$$\mathbf{D}_1 \mathbf{a} - \langle \mathbf{D}_1 \mathbf{a} \rangle \mathbf{1} = c_2 (\mathbf{D}_1 \mathbf{s}_2 - \langle \mathbf{D}_1 \mathbf{s}_2 \rangle \mathbf{1}) + c_3 (\mathbf{D}_1 \mathbf{s}_3 - \langle \mathbf{D}_1 \mathbf{s}_3 \rangle \mathbf{1}) \quad (A3)$$

2nd step: multiplication of Eq. A3 by diagonal matrix $\mathbf{D}_{12}$ with non-zero elements

$1/[\mathbf{D}_1 \mathbf{s}_2 - \langle \mathbf{D}_1 \mathbf{s}_2 \rangle \mathbf{1}]_i$:

$$\mathbf{p}_{12} = c_2 \mathbf{1} + c_3 \mathbf{q}_{12}, \quad (A4)$$

where $\mathbf{p}_{12} = \mathbf{D}_{12} \{ \mathbf{D}_1 \mathbf{a} - \langle \mathbf{D}_1 \mathbf{a} \rangle \mathbf{1} \}$, $\mathbf{q}_{12} = \mathbf{D}_{12} \{ \mathbf{D}_1 \mathbf{s}_3 - \langle \mathbf{D}_1 \mathbf{s}_3 \rangle \mathbf{1} \}$, and mean centering of (A4):

$$\mathbf{p}_{12} - \langle \mathbf{p}_{12} \rangle \mathbf{1} = c_3 \mathbf{z}_{12}, \quad (A5)$$

where $\mathbf{z}_{12} = \mathbf{q}_{12} - \langle \mathbf{q}_{12} \rangle \mathbf{1}$.

The LS solution of Eq. A5 is

$$\hat{c}_3 = \mathbf{a}^T \mathbf{k}_3^{RM} - h_3^{RM}, \quad (A6)$$

where $\mathbf{k}_3^{RM} = \mathbf{D}_1 \mathbf{f}_{12}$, $\mathbf{f}_{12} = \mathbf{D}_{12} \mathbf{z}_{12} / (\mathbf{z}_{12}^T \mathbf{z}_{12})$, and $h_3^{RM} = \langle \mathbf{D}_1 \mathbf{a} \rangle \mathbf{1}^T \mathbf{f}_{12}$.

The slope and the intercept of the linear equation for the first and the second components is readily obtained in the same way.

B. Condition number of the RS matrix of the pure-components of a Gaussian doublet.

Transformed matrix $\mathbf{S}$ (Eq. 15) has the form

$$\mathbf{S}_{rs} = \mathbf{D}_1 \mathbf{S}, \quad (B1)$$

where $\mathbf{D}_1$ is defined by Eq. 2. Then

$$\mathbf{S}_{RM}^T \mathbf{S}_{RM} = \left| \begin{array}{cc} \sum_{i=-m}^m \exp(-8x_i x_0) & \sum_{i=-m}^m \exp(-4x_i x_0) \\ \sum_{i=-m}^m \exp(-4x_i x_0) & N \end{array} \right|, \quad (B2)$$

where $N = 2m + 1$. For sufficiently large N , the sums in Eq. B2 can be substituted by integrals:

$$\sum_{i=-m}^m \exp(-8x_i x_0) = \frac{1}{\Delta x} \int_{-Lim}^{Lim} \exp(-8x x_0) dx = \frac{N}{2} Z(2p), \quad (B3)$$

$$\sum_{i=-m}^m \exp(-4x_i x_0) = \frac{1}{\Delta x} \int_{-Lim}^{Lim} \exp(-4x x_0) dx = \frac{N}{2} Z(p), \quad (B4)$$

where $\Delta x = 2Lim/N$, $Z(p) = [\exp(p) - \exp(-p)]/p$, and $p = 4Limx_0$.

The condition number of matrix $\mathbf{S}_{RM}^T \mathbf{S}_{RM}$ is

$$\text{cond}(\mathbf{S}_{RM}^T \mathbf{S}_{RM}) = \lambda_1 / \lambda_2, \quad (B5)$$

where λ is the eigenvalue of matrix $\mathbf{S}_{RM}^T \mathbf{S}_{RM}$ ($\lambda_1 \geq \lambda_2$). The eigenvalues are the solution of the following equation:

$$\det(\mathbf{S}_{RM}^T \mathbf{S}_{RM} - \lambda \mathbf{I}) = 0, \quad (B6)$$

where $\det$ is the determinant symbol and $\mathbf{I}$ is the identical matrix. From Eqs. B2-B6, we obtain:

$$\text{cond}(\mathbf{S}_{RM}) = \sqrt{\frac{Z(2p)+2+\sqrt{[Z(2p)-2]^2+4Z(p)^2}}{Z(2p)+2-\sqrt{[Z(2p)-2]^2+4Z(p)^2}}}. \quad (B7)$$

Using Taylor series, it is easy to show for $p < 1$ that

$$Z(p) \approx 2 + \frac{p^2}{3}. \quad (B8)$$

Substituting Eq. B8 into Eq. B7, we obtain:

$$\text{cond}(\mathbf{S}_{RM}) \approx \frac{\sqrt{12}}{p} \left(1 + \frac{p^2}{8} \right). \quad (B9)$$

For $p \leq 0.7$ approximation B9 is very close to the precise value obtained numerically.

Due to the symmetry of the Gauss doublet the same result was obtained for $\mathbf{S}_{RM} = \mathbf{D}_{21} \mathbf{S}$.

C. Gaussian doublet. Case study: Impact of systematic errors of matrix S on the errors of binary mixture analysis

$$\text{Let } \mathbf{S}_{dist} = \mathbf{S} + \Delta \mathbf{S}, \quad (C1)$$

where $\Delta \mathbf{S}$ is disturbance of matrix $\mathbf{S}$. The elements of matrix $\Delta \mathbf{S}$ can be regarded as known constants Δs_i . Substituting Eq. C1 into Eq. 1, we obtain:

$$\mathbf{S}_{dist} \mathbf{c}_x = \mathbf{S} \mathbf{c}, \quad (C2)$$

where $\mathbf{c}_x$ is an unknown concentration vector. The LS solution of Eq. C2 is

$$\mathbf{c}_x = (\mathbf{S}_{dist}^T \mathbf{S}_{dist})^{-1} \mathbf{S}_{dist}^T \mathbf{S} \mathbf{c}. \quad (C3)$$

Suppose that the shift of point $\mathbf{s}_i$ along $\mathbf{x}$ -axis from its original position in the mixture spectrum is the main source of systematic error $\Delta \mathbf{c}_i$. This shift, being dependent on the slope of the spectral curve [1], is measured by the fraction of sampling interval $\mathbf{h}$ along $\mathbf{x}$ -axis:

$$\Delta s_i = k h s'_i, \quad (C4)$$

where k is a constant, s'_i is the derivative of the spectrum $\mathbf{s}$ in the point i .

Setting Eqs. C1 and C4 into Eq. C3 and using the matrix equation for small k , we obtain:

$$(\mathbf{A} + k\mathbf{B})^{-1} \cong \mathbf{A}^{-1} - k\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}. \quad (C5)$$

Next, neglecting the small terms that contain k^r of the order $r > 1$, we have

$$\mathbf{c}_x = \mathbf{c} - \Delta \mathbf{c}_x^{LS}, \quad (C6)$$

where $\Delta \mathbf{c}_x^{LS} = k(\mathbf{S}^T \mathbf{S})^{-1} \mathbf{S}^T \mathbf{S}' \mathbf{c}$.

Eq. C6 is identical to the equation obtained in the case of the round-off errors of the explanatory variables (the elements of the regression matrix) [20].

Setting Eq. B1 in $\Delta \mathbf{c}$, we have for the ratio method:

$$\Delta \mathbf{c}_x^{RM} = k(\mathbf{S}_{RM}^T \mathbf{S}_{RM})^{-1} \mathbf{S}_{RM}^T [\mathbf{S}'_{RM} + (\mathbf{D}_1 \mathbf{S}_{RM})'] \mathbf{c}, \quad (C7)$$

where

$$\mathbf{S}_{RM} = \begin{bmatrix} \frac{s_{11}}{s_{21}} & 1 \\ \vdots & \vdots \\ \frac{s_{1m}}{s_{2m}} & 1 \end{bmatrix},$$

$$\mathbf{S}'_{RM} = \begin{bmatrix} \frac{s'_{11}s_{21} - s_{11}s'_{21}}{(s_{21})^2} & 0 \\ \vdots & \vdots \\ \frac{s'_{1m}s_{2m} - s_{1m}s'_{2m}}{(s_{2m})^2} & 0 \end{bmatrix},$$

$$(\mathbf{D}_1 \mathbf{S}_{RM})' = \begin{bmatrix} \frac{s_{11}s'_{21}}{(s_{21})^2} & \frac{s'_{21}}{s_{21}} \\ \vdots & \vdots \\ \frac{s_{1m}s'_{2m}}{(s_{2m})^2} & \frac{s'_{2m}}{s_{2m}} \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

The first term in the brackets in Eq. C7 appears due to the shifts of the mixture pure-component spectra. The second term is connected with the errors of the mixture ratio spectra that are due to the shift of the second pure-component spectrum.

To compare errors $\Delta \mathbf{c}_x^{LS}$ and $\Delta \mathbf{c}_x^{RM}$, we used the ratio of the vector norms

$$E = \|\Delta \mathbf{c}_x^{RM}\| / \|\Delta \mathbf{c}_x^{LS}\| \quad (C8)$$

which was evaluated numerically.

Besides, total prediction error for p mixtures was calculated as

$$RSEP = \sqrt{\frac{\sum_{j=1}^p \sum_{i=1}^n [(c_x)_{ij} - c_{ij}]^2}{\sum_{j=1}^p \sum_{i=1}^n c_{ij}^2}}, \quad (C9)$$

where $p = 9, n = 2, c_{1j} = 0.9, 0.8, \dots, 0.1$ and $c_{2j} = 1 - c_{1j}$.