



## Error Analysis in Determining Parameters of Overlapping Peaks Using Separation of Complex Spectra into Individual Components. Case Study: Quartets

J. Dubrovkin

Computer Department, Western Galilee College,  
2421 Acre, Israel

**Abstract:** Based on the concept of Big Data Modeling, the errors in determining peak parameters of noisy Gaussian quartets using nonlinear least squares curve fitting have been evaluated. The probability that the relative error in estimating each model parameter is not greater than a priori given limit for a given fitting error has been found. Obtained results showed that small fitting error does not guarantee that the fitting algorithm does converge to the correct peak parameters. It was found that the mean probability is a useful measure of the effectiveness of the curve fitting procedure.

**Keywords:** spectrochemistry, overlapping peaks, curve fitting, errors of peak parameters

### I. Introduction

Instrumental spectral methods of analysis are widely used to determine the physicochemical properties of natural and synthetic materials. Data produced by spectrometers can be represented as a sum of elementary peaks. Peaks are often strongly overlapped. Instrumental methods cannot completely resolve overlapping peaks. Therefore, mathematical signal processing must be done with the purpose of extracting pure peaks.

Mathematical methods of separation of overlapping peaks have been discussed in numerous articles (see references in [1]). Fitting a model to experimental data (the curve fitting) is the dominant separation method currently used in spectrochemistry [2]. The reliability of the peak parameters obtained by the curve fitting will affect the reliability of the results of qualitative and quantitative analyses.

The possibility of obtaining inaccurate parameters was discussed in terms of the multivariate selectivity of binary mixture spectra [3]. Theoretical estimation of the upper bound of the total mean square error for the fitted parameters using different constraints in these parameters was given by [4, 5]. Unfortunately, obtained data was tested using a relatively small number of experimental or synthesized objects. There are two approaches to increase this number: (a) representation of possible multicomponent spectra as a stationary stochastic process [6-8] and (b) creation of multicomponent spectral models in a very wide range of varied parameters. However, the first method is based on Fourier analysis. The drawback of this method is assumption of stationarity of spectra (the low-order moments of the joint distribution of spectral points do not depend on the shifts along x-axis of a spectrum [9]). The second approach is more attractive since it is not too cumbersome to create very large number of models using modern hardware and software tools which allow researchers to develop and analyse very complex big data systems ("Big modeling" [10]).

Let us provide some examples to more fully illustrate this issue in chemistry:

1. The first idea of Big (chemistry) Data Management [11].
2. Identification of potential toxicants using High-Throughput Screening of million compounds obtained from public toxicity data resources [12].
3. Compressing biological data by choosing representative subsets from huge sets for development of quantitative structure-activity relationships [13]. The developed algorithm was based on a Monte Carlo/simulated annealing.
4. PLS-regression of 400MB matrices and PCA-identification of 2000 x 2,000,000 matrices using Fourier transform data [14].
5. Data reduction in multicapillary column-ion mobility spectrometry using wavelet reconstruction of 2.5 GB matrix [15].
6. Management in the cloud of distributed databases in computational quantum chemistry [16].
7. Exploratory Big Chemical Data Analysis (Big Data Visualization [17]) based on: (a) clustering and approximation techniques (The Compressed Score Plots) applied to the data sets contained up to five million observations and more than one hundred variables [18]; (b) projections of 2D big data sets obtained by mapping various physical and chemical properties of more than 2 million molecules onto the incremental version of generative topographic map. The processing time was 19 h on a desktop single processor machine [19]; (c) chemometrics tools applied to TB-data sets in Laser Induced Breakdown Spectroscopy [20].

8. Molecular structure elucidation: (a) identification of large complex molecules of up to 100 or more skeletal atoms with topological peculiarity; (b) conformational analysis of big rigid molecules containing many stereocenters [21].

All big data treatment methods are based on data compression. Since it is common to perform extraction of the separated peaks from uncompressed spectra, these spectra were used to quantify uncertainty of peak fitting parameters in our research. It is important to underline that decomposition of complex spectra is a time consuming process.

Errors in determining parameters of overlapping peaks (response variables) depend on the peak parameters (factors) in a very complex way (e. g., the uncertainties of the maximum peak positions [21-23]). Therefore, choosing specific values of the parameters suitable for estimating their uncertainties (the optimal experimental design) cannot be performed using mathematical relationships between response variables and factors [25, 26]. In our empirical approach the most common set of the parameters of the multicomponent spectral models has been studied. Although it is impossible to mimic all practical situations (a population) encountered in spectra, general qualitative and quantitative patterns can be obtained.

The goal of our work was analysis of the errors in determining peak parameters of overlapping peaks using the least squares decomposition of complex spectra into individual components. The spectra were simulated by the 771,750 combinations of four Gaussian peaks.

In what follows, for the sake of simplicity, term “line” is used instead of terms “peak, line and band”. The standard algebraic notations are used throughout the article. All calculations were performed and the plots were built using the MATLAB software.

## II. Computer Modeling

Multiline spectra were simulated by Gaussian quartets:

$$Q(x) = \sum_{i=1}^4 F(x, x_{0i}, R_i, r_i), \quad (1)$$

where  $F(x, x_{0i}, R_i, r_i) = R_i \exp\{-(x - x_{0i})/r_i\}^2\}$  is the  $i^{\text{th}}$  line;  $x = 2\sqrt{\ln 2}j, j = [-7:0.1:7]$ ,  $x_{0i}$  is the position of the  $i^{\text{th}}$  line maximum;  $R_i$  and  $r_i$  are the relative intensity and the relative width of the  $i^{\text{th}}$  line, respectively. Parameter values of quartets are given in Table 1 (total number of combinations  $N_{\text{total}} = 771,750$ ). High and low resolution spectra of quartets are shown in Fig. 1.

Table 1. Parameter values

| Parameter | $x_{02}$ | $-x_{01}, x_{03}$ | $x_{04}$ | $R_2$ | $R_1, R_3, R_4$           | $r_2$ | $r_1, r_3, r_4$   |
|-----------|----------|-------------------|----------|-------|---------------------------|-------|-------------------|
| Values    | 0        | 0.3, 0.5, 1       | 1.5, 2   | 1     | 0.2, 1/3, 0.5, 1, 2, 3, 5 | 1     | 1/3, 0.5, 1, 2, 3 |

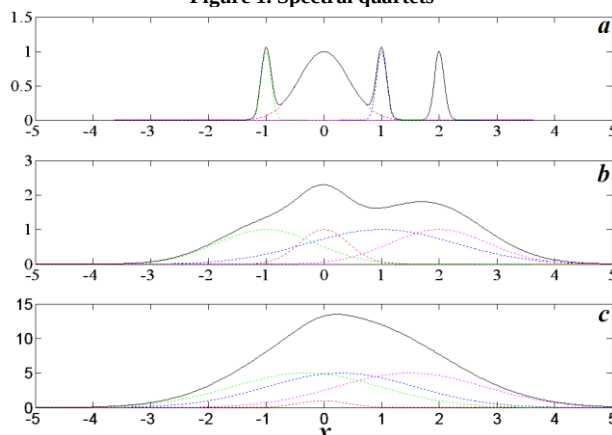
Curve fitting (non-linear regression) to four Gaussian lines was performed using the Gauss-Newton algorithm with Levenberg-Marquardt modifications (MATLAB function *nlinfit*). The initial values of parameters were uniformly distributed random variables:

$$P_i^{\text{init}} = LB + (UB - LB)\eta, \quad (2)$$

$$P_{x_{02}}^{\text{init}} = -lim + 2lim \eta, \quad (3)$$

where  $LB = P_i^{\text{true}} lLim$  and  $UB = P_i^{\text{true}} uLim$  are the low and the upper bounds, respectively;  $P_i^{\text{true}}$  is a true parameter value;  $lLim$  and  $uLim$  are  $1 \mp lim$ , respectively;  $lim=0.2$  and  $0.5$ ;  $\eta$  is a random variable uniformly distributed in the range 0-1. Since the true value of  $x_{02}$  is zero, Eq. 3 is differs from Eq. 2.

Figure 1. Spectral quartets



$[x_{01}, R_1, r_1, x_{03}, R_3, r_3, x_{04}, R_4, r_4] = [-1, 1, 1/3, 1, 1, 1/3, 2, 1, 1/3]$  (a),  $[-1, 1, 2, 1, 1, 3, 2, 1, 2]$  (b) and  $[-0.3, 5, 3, 0.3, 5, 3, 1.5, 5, 3]$  (c).

Random noise with zero mean and standard deviation  $\sigma = 0.01$  was added to the quartet (1). The quartet parameters were estimated. For randomization these two procedures were repeated 20 times. We studied two cases:

1. Unconstrained non-linear regression (UR) when whole curve-fitting data was considered valid.
2. Constrained regression (CR) with the soft constraints:

$$x_{01} < 0, x_{03} > 0, x_{04} > 0, R > 0, r > 0 \quad (4)$$

In the case of the rank deficient program warnings regression data was rejected.

The relative error (%) in estimating parameter  $P_i$  is equal to

$$\xi_{P_i} = 100(\bar{P}_i^{estim} - P_i^{true})/P_i^{true}, \quad (5)$$

where  $\bar{P}_i^{estim}$  is the mean value of the  $P_i^{estim}$ -values.

The fitting error:

$$\xi_{fit} = \sqrt{\sum_{j=1}^{141} (Q(x_j) - \hat{Q}(x_j))^2 / 141}, \quad (6)$$

where  $\hat{Q}(x_j)$  is obtained from Eq. 1 by substituting  $\bar{P}_i^{estim}$ .

The indicator for good model fit was

$$\xi_{fit} \leq K_{noise} \sigma, \quad (7)$$

where  $K_{noise}$  is the noise multiplier. The number of the good-fitting models  $N_{gf} < N_{total}$ . From these models only those quartets for which

$$\xi_{P_i} \leq K_\xi \quad (8)$$

have been chosen.  $K_\xi = [20, 10, 5, 1]$  for all  $P_i$ , except  $P_{x_{02}}$  for which  $K_\xi = [0.05, 0.025, 0.01, 0.002]$  (Fig. 2).

Number of selected quartets  $N_{sel} = N(K_{noise}, K_\xi)$ . The relative values (%)

$$\Psi_{gf} = 100N_{sel}/N_{gf} \quad (9)$$

$$\Psi_{total} = 100N_{sel}/N_{total} \quad (10)$$

have been calculated.  $\Psi_{gf}$  and  $\Psi_{total}$  may be considered as the probability that the relative error in estimating parameter  $P_i$  is not greater than  $K_\xi$  for a given  $\xi_{fit}$  over good-fitting models and over all experiments, respectively.

### III. Results and Discussion

Typical histograms of the error ( $\xi_{P_i}$ ) distributions are given in Fig. 3. These plots are usually asymmetric unimodal, however, in some cases two modes appear (e.g., for  $R_4$ ). Table 2 represents the statistical parameters (mean, variance and skewness) of the distributions. It is seen from this table that for experiments based on the unconstrained regression the absolute value of the mean errors  $|\bar{\xi}_{P_i}|$  increase if the range bounds of initial guesses increase ( $lim$  increases from 0.2 to 0.5). However, there are two exceptions  $|\bar{\xi}_{R_2}|$  and  $|\bar{\xi}_{x_{03}}|$  for  $K_{noise} = 10$ . For experiments based on constrained regression (Eq. 4), from 771,750 Gaussian quartets 745,550 quartets (96.6%) were suitable for analysis. As expected, for these quartets  $|\bar{\xi}_{P_i}|$ - values decrease compared with corresponding values obtained by unconstrained regression. The exceptions are  $|\bar{\xi}_{R_2}|$  and  $|\bar{\xi}_{R_4}|$ . No dependence of skewness upon the errors of modeling parameters has been found.

The relative number of good model fits ( $N_{gf}/N_{total}$ ) decreases significantly for  $lim = 0.5$  compared with  $lim = 0.2$  when  $K_{noise}$  is small (Table 3). The probability  $\Psi_{gf}$  (Eq. 9) is maximal for small values of  $K_{noise}$  (Fig. 4). Bad initial guesses decrease  $\Psi_{gf}$ . Since in the constrained regression a number of rejected models becomes larger than that in the unconstrained regression ( $N_{gf}$  decreases),  $\Psi_{gf}$  also increases.

The dependences of the probability of "good fitting" on  $K_{noise}$  over all numerical experiments based on the unconstrained regression ( $\Psi_{total}(K_{noise})$ ) have very similar shape for all coefficients of quartet lines (Fig. 5). These dependences are increasing. They approach to saturation for  $K_{noise} > 20$ . In all cases the probabilities increase while increasing  $K_\xi$ .

It is seen that small values of  $K_{noise}$  do not guarantee high probability of the accurate results of the fitting procedure.

The relative error probability is a very complex function  $\Psi_{gf}(P_i, lim, K_{noise}, K_\xi)$ . Some numerical values of  $\Psi_{gf}$  are given in Table 4. It is seen that in most cases  $\Psi_{gf} < 0.95 * \bar{\Psi}_{gf}$  for  $x_{02}, R_3, R_4$  and  $r_4$ . For  $r_3$  and  $x_{04}$  this condition is not always true. The  $\Psi_{gf}$ -values were calculated from the independent samples randomly extracted from the full dataset. It was found that these values are different despite the fact that the samples are very large. Two examples (sample 1 and 2 containing 148,603 and 2958 triplets, respectively) are given in Table 4. Significant differences between probabilities obtained from the entire population and from the samples are given in red in these tables. However, the mean probability ( $\bar{\Psi}_{gf}$ ) varies little. Therefore, it can be the useful measure of the effectiveness of the curve fitting procedure. It is seen from Table 4 that the ratio  $(\bar{\Psi}_{gf})_{K_{noise}=3}/(\bar{\Psi}_{gf})_{K_{noise}=10}$  increases while decreasing  $K_\xi$ . In other words, to increase probability ( $\Psi_{gf}$ ) of minimal errors in determining line parameters the noise coefficient should be taken as minimal as possible.

Figure 2. Selection algorithm

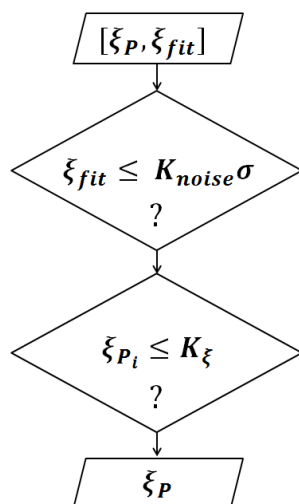
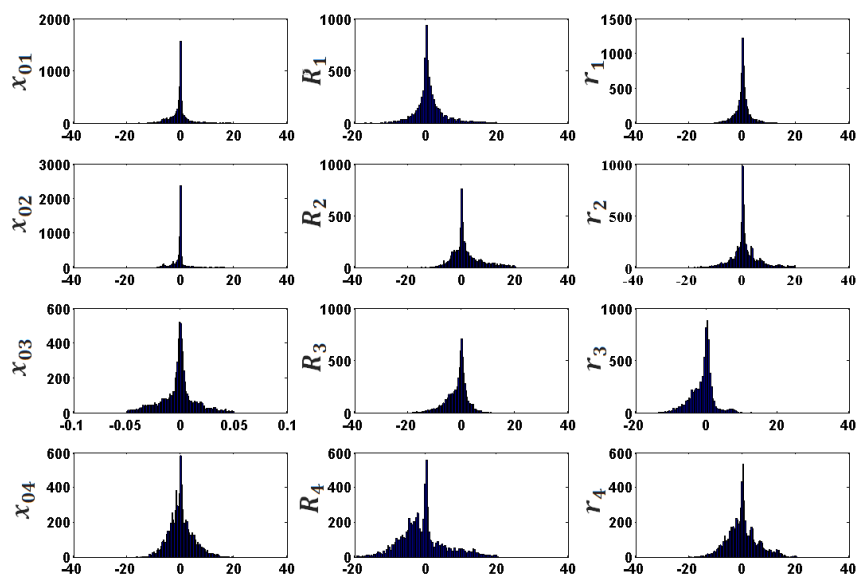


Figure 3. Histograms



$lim = 0.5$ ,  $CR$ ,  $K_{noise} = 3$

Table 2. Statistical parameters of the error distributions

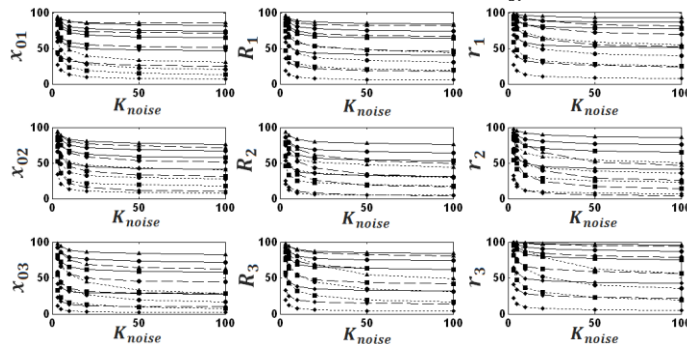
| Regression             |          | UR                |                   |                 |                   | CR                |                   |
|------------------------|----------|-------------------|-------------------|-----------------|-------------------|-------------------|-------------------|
| $lim$                  |          | 0.2               |                   | 0.5             |                   | 0.5               |                   |
| $K_{noise}$            |          | 3                 | 10                | 3               | 10                | 3                 | 10                |
| Parameters of quartets | $x_{01}$ | 0                 | 0                 | $-1.5 \pm 0.33$ | $-3.2 \pm 0.15$   | $-0.5 \pm 0.13$   | $-0.75 \pm 0.051$ |
|                        |          | 0.46              | 0.30              | -1.05           | -1.3              | 0.19              | -0.31             |
|                        | $R_1$    | 0                 | $-0.51 \pm 0.016$ | $1.5 \pm 0.32$  | $1.3 \pm 0.11$    | $0.8 \pm 0.12$    | $-0.17 \pm 0.052$ |
|                        |          | 0.24              | -0.61             | 1.08            | 0.73              | 1.3               | 0.1               |
|                        | $r_1$    | $-0.65 \pm 0.012$ | $-1.1 \pm 0.013$  | $-1.2 \pm 0.26$ | $-1.6 \pm 0.10$   | $-0.24 \pm 0.088$ | $-1.03 \pm 0.045$ |
|                        |          | -2.5              | -2.1              | -1.1            | -1.0              | -0.1              | -1.0              |
|                        | $x_{02}$ | 0                 | 0                 | $-0.6 \pm 0.27$ | $-0.9 \pm 0.12$   | 0                 | $0.25 \pm 0.063$  |
|                        |          | -0.52             | -0.087            | -0.29           | -0.43             | 1.2               | 0.55              |
|                        | $R_2$    | $-0.39 \pm 0.014$ | $-0.57 \pm 0.016$ | $1.1 \pm 0.32$  | $0.3 \pm 0.12$    | $1.6 \pm 0.17$    | $-0.53 \pm 0.075$ |
|                        |          | -0.75             | -0.63             | 0.69            | 0.17              | 1.6               | -0.2              |
|                        | $r_2$    | $-0.60 \pm 0.009$ | $-1.1 \pm 0.013$  | $-1.3 \pm 0.29$ | $-2.6 \pm 0.12$   | $0.6 \pm 0.14$    | $-2.22 \pm 0.085$ |
|                        |          | -2.2              | -1.7              | -1.2            | -1.3              | 0.8               | -1.0              |
|                        | $x_{03}$ | $\pm 0.015$       | $0.22 \pm 0.016$  | 0               | 0                 | 0                 | 0                 |
|                        |          | 0.22              | 0.49              | -0.18           | -0.23             | -0.8              | 0                 |
|                        | $R_3$    | $0.21 \pm 0.017$  | $-0.49 \pm 0.018$ | $-2.2 \pm 0.29$ | $-3.1 \pm 0.12$   | $-2.0 \pm 0.14$   | $-1.84 \pm 0.062$ |
|                        |          | 0.58              | -0.50             | -1.3            | -1.2              | -1.6              | -1.2              |
|                        | $r_3$    | $-0.84 \pm 0.015$ | $-1.6 \pm 0.016$  | $-2.0 \pm 0.25$ | $-3.0 \pm 0.11$   | $-1.5 \pm 0.11$   | $-1.45 \pm 0.048$ |
|                        |          | -2.3              | -1.9              | -1.2            | -1.2              | -1.4              | -1.0              |
|                        | $x_{04}$ | 0                 | 0                 | $-1.0 \pm 0.21$ | $-1.66 \pm 0.085$ | 0                 | 0                 |
|                        |          | 0.75              | 0.38              | -1.2            | -1.3              | 0.30              | 0.26              |
|                        | $R_4$    | $-0.46 \pm 0.014$ | $-0.98 \pm 0.014$ | $-0.8 \pm 0.23$ | $-1.3 \pm 0.10$   | $-1.0 \pm 0.17$   | $-3.4 \pm 0.11$   |
|                        |          | -1.2              | -1.4              | -0.75           | -0.81             | -0.38             | -1.1              |
|                        | $r_4$    | $-0.85 \pm 0.013$ | $-1.2 \pm 0.013$  | $-1.6 \pm 0.23$ | $-2.50 \pm 0.093$ | $-0.2 \pm 0.14$   | $-1.28 \pm 0.082$ |
|                        |          | -3.0              | -2.4              | -1.7            | -1.6              | 0                 | -0.67             |

$\mu_1 \pm 2\mu_2^{1/2}$  and  $\gamma_{as} = \mu_3/\mu_2^{3/2}$  (skewness) are given in the upper and lower rows, respectively.  $\mu_i$  is the  $i^{\text{th}}$ -order central moment.

Table 3. Relative number of good model fits ( $N_{gf}/N_{total}, \%$ )

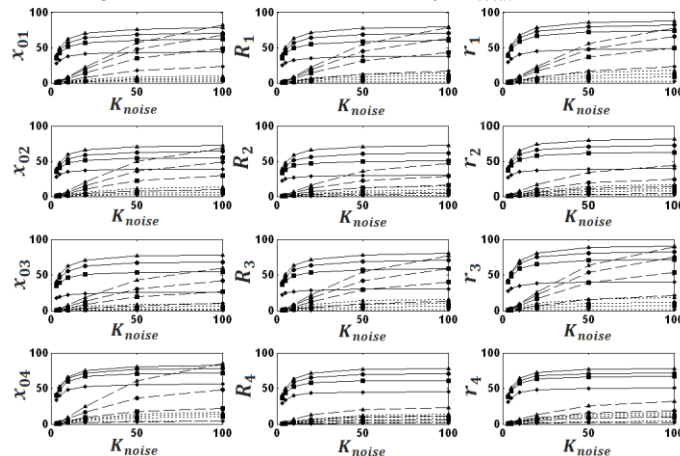
| $lim$ | Regression | $K_{noise}$ |      |      |      |      |      |
|-------|------------|-------------|------|------|------|------|------|
|       |            | 3           | 5    | 10   | 20   | 50   | 100  |
| 0.2   | UR         | 42.3        | 54.1 | 70.8 | 82.4 | 91.4 | 94.8 |
| 0.5   | UR         | 0.4         | 1.4  | 5.4  | 12.6 | 25.3 | 32.8 |
| 0.5   | CR         | 1.1         | 2.8  | 9.9  | 26.3 | 66.4 | 95.3 |

Figure 4. Relative error probability ( $\psi_{gf}$ )



$K_{\xi}=20$  (▲), 10 (●), 5 (■) and 1 (◆).  $\lim=0.2$  (—) and 0.5 (···) (UR); 0.5 (---) (CR).

Figure 5. Relative error probability ( $\psi_{total}$ )



$K_{\xi}=20$  (▲), 10 (●), 5 (■) and 1 (◆).  $\lim=0.2$  (—) and 0.5 (···) (UR); 0.5 (---) (CR).

Table 4. Relative error probability ( $\psi_{gf}$ )

| All data  |            |       |       |             |             |             |             |             |             |             |             |             |                   |
|-----------|------------|-------|-------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------------|
| $K_{\xi}$ | Parameters |       |       |             |             |             |             |             |             |             |             |             | $\bar{\psi}_{gf}$ |
|           | $x_{01}$   | $R_1$ | $r_1$ | $x_{02}$    | $R_2$       | $r_2$       | $x_{03}$    | $R_3$       | $r_3$       | $x_{04}$    | $R_4$       | $r_4$       |                   |
| 20        | 93.5       | 94.8  | 97.9  | 93.0        | 98.0        | 99.8        | 94.4        | <b>86.5</b> | 93.2        | 99.2        | 92.7        | 93.0        | 94.7              |
|           | 87.6       | 86.3  | 92.6  | <b>75.0</b> | 89.0        | 98.1        | 80.5        | <b>69.2</b> | <b>75.2</b> | 95.1        | <b>67.6</b> | <b>64.0</b> | 81.7              |
| 10        | 84.6       | 87.2  | 93.9  | <b>76.9</b> | 91.2        | 96.9        | 86.1        | <b>74.6</b> | 82.5        | 93.1        | <b>77.4</b> | 82.6        | 85.6              |
|           | 76.3       | 73.9  | 83.5  | <b>55.7</b> | 73.4        | 87.4        | 64.0        | <b>52.0</b> | <b>52.0</b> | 74.4        | <b>40.4</b> | <b>47.2</b> | 65.0              |
| 5         | 70.9       | 74.5  | 83.6  | <b>54.3</b> | 75.1        | 81.5        | 73.1        | <b>59.1</b> | 66.3        | 72.3        | <b>53.8</b> | <b>60.2</b> | 68.7              |
|           | 58.8       | 58.0  | 68.9  | <b>35.3</b> | 53.4        | 66.5        | 45.0        | <b>33.8</b> | <b>33.4</b> | 45.1        | <b>22.3</b> | <b>28.5</b> | 45.8              |
| 1         | 41.5       | 35.6  | 46.0  | <b>22.7</b> | 33.0        | 41.3        | 51.2        | <b>25.2</b> | 32.4        | <b>23.5</b> | <b>20.3</b> | <b>22.2</b> | 32.9              |
|           | 32.6       | 24.7  | 32.7  | <b>13.2</b> | 19.3        | 28.0        | 21.6        | <b>10.3</b> | <b>11.7</b> | <b>10.8</b> | <b>6.1</b>  | <b>8.2</b>  | 18.3              |
| Sample 1  |            |       |       |             |             |             |             |             |             |             |             |             |                   |
| 20        | 93.1       | 95.1  | 97.5  | 93.8        | 98.0        | 99.8        | 94.0        | <b>86.0</b> | 92.9        | 99.2        | 92.8        | 91.9        | 94.5              |
|           | 87.8       | 86.4  | 92.4  | 80.1        | <b>69.0</b> | <b>75.3</b> | <b>75.3</b> | 89.1        | 98.3        | 95.0        | <b>67.9</b> | <b>63.4</b> | 81.7              |
| 10        | 84.6       | 87.4  | 93.8  | <b>76.7</b> | 91.9        | 97.1        | 85.8        | <b>74.3</b> | 82.8        | 93.3        | <b>77.6</b> | 81.1        | 85.5              |
|           | 76.9       | 73.9  | 83.2  | 63.9        | <b>51.9</b> | <b>52.3</b> | <b>55.4</b> | 73.1        | 86.8        | 74.2        | <b>40.5</b> | <b>46.4</b> | 64.9              |
| 5         | 70.7       | 74.3  | 83.7  | <b>52.3</b> | 73.8        | 82.1        | 73.7        | <b>59.2</b> | 66.4        | 72.2        | <b>54.3</b> | <b>60.6</b> | 68.6              |
|           | 59.4       | 57.9  | 68.3  | 45.4        | <b>33.9</b> | <b>33.7</b> | <b>35.2</b> | 53.0        | 66.1        | 45.2        | <b>22.2</b> | <b>27.5</b> | 45.7              |
| 1         | 40.9       | 35.1  | 46.0  | <b>23.2</b> | 33.2        | 41.0        | 50.9        | <b>25.7</b> | 31.0        | <b>22.8</b> | <b>21.0</b> | <b>22.0</b> | 32.7              |
|           | 32.9       | 24.7  | 33.1  | 21.6        | <b>10.7</b> | <b>11.7</b> | <b>13.0</b> | 19.5        | 27.9        | <b>10.8</b> | <b>6.2</b>  | <b>7.9</b>  | 18.3              |

Table 4 (continued)

| Sample 2 |      |             |      |      |             |             |             |      |      |             |             |             |      |
|----------|------|-------------|------|------|-------------|-------------|-------------|------|------|-------------|-------------|-------------|------|
| 20       | 93.2 | <b>88.6</b> | 95.5 | 90.9 | <b>81.8</b> | 97.7        | 93.2        | 97.7 | 100  | 95.5        | <b>88.6</b> | 97.7        | 93.4 |
|          | 87.0 | 85.2        | 92.6 | 83.0 | <b>70.0</b> | <b>75.9</b> | <b>72.8</b> | 87.7 | 98.5 | 95.4        | <b>69.2</b> | <b>65.7</b> | 81.9 |
| 10       | 88.6 | 86.4        | 86.4 | 86.4 | <b>70.5</b> | 84.1        | <b>79.5</b> | 88.6 | 97.7 | 86.4        | <b>79.5</b> | 88.6        | 85.2 |
|          | 76.9 | 73.1        | 79.6 | 67.0 | <b>51.5</b> | <b>55.2</b> | <b>48.8</b> | 73.1 | 88.6 | 72.2        | <b>44.4</b> | <b>50.9</b> | 65.1 |
| 5        | 77.3 | 79.5        | 81.8 | 68.2 | <b>61.4</b> | <b>61.4</b> | <b>54.5</b> | 77.3 | 81.8 | <b>61.4</b> | <b>54.5</b> | <b>54.5</b> | 67.8 |
|          | 61.4 | 60.5        | 65.7 | 47.5 | <b>34.6</b> | <b>37.7</b> | <b>30.6</b> | 53.4 | 66.4 | 45.1        | <b>24.1</b> | <b>29.9</b> | 46.4 |
| 1        | 50.0 | 40.9        | 52.3 | 45.5 | <b>13.6</b> | 31.8        | 25.0        | 31.8 | 45.5 | <b>29.5</b> | <b>18.2</b> | <b>15.9</b> | 33.3 |
|          | 32.7 | 27.2        | 32.7 | 22.5 | <b>9.0</b>  | <b>13.3</b> | <b>19.8</b> | 19.8 | 26.2 | <b>13.3</b> | <b>6.5</b>  | <b>9.0</b>  | 19.2 |

$lim = 0.5$ , CR.  $K_{noise} = 3$  and 10 for upper and lower rows, respectively. Values  $\psi_{gf} < 0.95 * \bar{\psi}_{gf}$  are given in bold italic

In practice it is easily to evaluate the relative error probability in determining line parameters of badly resolved measured curves using above method. For this goal a set of simulated spectra which line parameters are varying near calculated parameters must be created. The values of the coefficients  $K_{noise}$  and  $K_{\xi}$  should be chosen according to the specific requirements of the curve-fitting problem.

#### References

- [1] J. Dubrovkin, "Critical Analysis of Spectral Deconvolution Methods", International Journal of Emerging Technologies in Computational and Applied Sciences, vol. 1-10, 2014, pp. 40-49.
- [2] P. Pelikan, M. Ceppan and M. Liska, Applications of Numerical Methods in Molecular Spectroscopy, CRC, 1994.
- [3] G. Vivó-Truyols, J. R. Torres-Lapasió and M. C. García-Alvarez-Coque, "Net analyte signal as a deconvolution-oriented resolution criterion in the optimisation of chromatographic techniques", J. Chromatogr. A, vol. 991, 2003, pp. 47-59.
- [4] G. A. F. Seber and C. J. Wild, Nonlinear Regression, Hoboken, NJ: Wiley-Interscience, 2003.
- [5] V. A. Lorenz-Fonfri and E. Padros, "Curve-fitting overlapped bands: quantification and improvement of curve-fitting robustness in the presence of errors in the model and in the data", Analyst, vol. 129, 2004, pp. 1243 – 1250.
- [6] A. Felinger, L. Pasti and F. Dondi, "Fourier analysis of multicomponent chromatograms. Theory and models", Anal. Chem, 1990, vol.62, pp. 1846-1853.
- [7] A. Felinger, L. Pasti, P. Reschiglian and F. Dondi, "Fourier Analysis of multicomponent chromatograms. Numerical evaluation of statistical parameters", Anal. Chem, 1990, vol.62, pp. 1854-1860.
- [8] A. Felinger, L. Pasti and F. Dondi, "Fourier Analysis of multicomponent chromatograms. Theory of nonconstant peak width model", Anal. Chem, 1991, vol. 63, pp. 2627-2733.
- [9] S. M. Ross, Stochastic processes. 2nd Ed. John Wiley & Sons, 1996.
- [10] "Big Modeling Up Front (BMUF) Anti-Pattern", <http://www.agilemodeling.com/essays/bmuf.htm>
- [11] B. Gibb, "Big (chemistry) data", Nature chemistry, vol 5, 2013, pp. 248-249.
- [12] Hao Zhu, Jun Zhang, M. T. Kim, A. Boison, A. Sedykh and K. Moran, "Big Data in Chemical Toxicity Research: The Use of High-Throughput Screening Assays To Identify Potential Toxicants", Chem. Res. Toxicol., vol. 27, 2014, pp 1643–1651.
- [13] A. Yosipof and H. Senderowitz, "Optimization of Molecular Representativeness", J. Chem. Inf. Model., vol. 54, 2014, pp 1567–1577.
- [14] J. Dubrovkin, "Big Data approach to analytical chemistry", International Journal of Emerging Technologies in Computational and Applied Sciences, vol. 2-7, 2014, pp. 205-210.
- [15] E. Szymańska, E. Brodrick, M. Williams, A. N. Davies, H. J. van Manen and L. M. Buydens, "Data size reduction strategy for the classification of breath and air samples using multicapillary column-ion mobility spectrometry", Anal Chem., vol. 87, 2015, pp.869-875.
- [16] M. Álvarez-Moreno, C. de Graaf, N. López, F. Maseras, J. M. Poblet and C. Bo, "Managing the computational chemistry big data problem: The ioChem-BD platform", J. Chem. Inf. Model., vol. 55, 2015, pp 95–103.
- [17] P. Simon, The visual organization: data visualization, big data, and the quest for better decisions, Wiley, 2014.
- [18] J. Camacho, "Visualizing big data with compressed score plots: approach and research challenges", Chemom.Intel. Lab. Syst., vol.135, 2014, pp. 110–125.
- [19] H. A. Gaspar, I. I. Baskin, G. Marcou, D. Horvath and A. Varnek, "Chemical data visualization and analysis with incremental generative topographic mapping: big data challenge", J. Chem. Inf. Model., vol. 55, 2015, pp 84–94.
- [20] C. McManus, N. McMillian and J. Dowe, "Analysis of "Big" Spectral Data Chemometrics Without Preconceived Notions", Conference : SCIX 2014.
- [21] M. Elyashberg, K. Blinov, S. Molodtsov, Y. Smurnyy, A. J. Williams and T. Churanova, "Computer-assisted methods for molecular structure elucidation: realizing a spectroscopist's dream", Journal of Cheminformatics, vol. 1 :3,2009, pp.1-26.
- [22] J. Dubrovkin, "Evaluation of the peak location uncertainty in spectra. Case study: symmetrical lines", Journal of Emerging Technologies in Computational and Applied Sciences, vol. 1-7, 2014, pp. 45-53.
- [23] J. Dubrovkin, "Evaluation of the peak location uncertainty in second-order derivative spectra. Case study: Symmetrical lines", Journal of Emerging Technologies in Computational and Applied Sciences, vol. 3-9, 2014, pp. 248-255.
- [24] J. Dubrovkin, "Identification of peak positions using second-order derivative spectra and Tikhonov deconvolution method. Comparison study", Journal of Emerging Technologies in Computational and Applied Sciences, vol. 2-10, 2014, pp.192-197.
- [25] V. V. Fedorov and P. Hackl, Model-oriented design of experiments. Springer-Verlag, 1997.
- [26] R. L. Mason, R. F. Gunst and J. L. Hess, Statistical Design and Analysis of Experiments With Applications to Engineering and Science. 2nd Ed. John Wiley & Sons, 2003.