

AN ALGORITHM FOR SPECTROMETRIC DATA CORRECTION WITH BUILT-IN ESTIMATION OF UNCERTAINTY

T. Szafranski, P. Sprzeczak and R.Z. Morawski

Warsaw University of Technology

Faculty of Electronics and Information Technology, Institute of Radioelectronics
ul. Nowowiejska 15/19, 00-665 Warszawa, Poland

Abstract: Practical usefulness of the result of spectrometric data correction depends considerably on how precise its uncertainty is assessed. If the algorithm used for correction is linear with respect to the data and all intermediate results of computation, and if it does not use the data or intermediate results of computation more than once, then the linear propagation of uncertainties represented by intervals, executed step by step for each operation, yields a satisfactory estimate of the measurement uncertainty. Otherwise, a tendency to overestimate the uncertainty appears. An effective interval-based method for dealing with this problem in weakly non-linear algorithms has been recently proposed by the authors. In this paper, it is implemented in a new algorithm for spectrometric data correction. Computational efficiency of this algorithm is demonstrated using synthetic spectrometric data.

Keywords: spectrometry, spectrometric data correction, measurement uncertainty, uncertainty evaluation

1 INTRODUCTION

The spectrometric data correction is a special case of measurand reconstruction [1, 2], i.e. an operation aimed at transforming the raw measurement data $\tilde{\mathbf{y}} = [\tilde{y}_1 \tilde{y}_2 \dots]^T$, acquired by means of a spectrometer, into an estimate $\hat{\mathbf{x}} = [\hat{x}_1 \hat{x}_2 \dots]^T$ of the measured spectrum $\mathbf{x} = [x_1 x_2 \dots]^T$. This operation can be compactly defined by an operator R :

$$\hat{\mathbf{x}} = R[\tilde{\mathbf{y}}; \hat{\mathbf{p}}] \quad (1)$$

where $\hat{\mathbf{p}} = [\hat{p}_1 \hat{p}_2 \dots]^T$ is an estimate of the vector of parameters $\mathbf{p} = [p_1 p_2 \dots]^T$, determined during calibration of the spectrometric system under consideration. Numerous examples of the operators R and corresponding algorithms for spectrometric data correction may be found in [3] and [4-18].

The result of correction $\hat{\mathbf{x}}$ is subject to uncertainty "inherited" both from the measurement data $\tilde{\mathbf{y}}$ and from the estimate of the vector of parameters $\hat{\mathbf{p}}$. Its practical value depends on how precise these uncertainties are assessed [19]. Such an assessment is not a trivial task if the structure of the algorithm contains loops - then basic methods of error propagation may become useless because of overestimation of uncertainty due to inter-correlation of the intermediate results of computation. In the paper [20], the authors proposed some means for dealing with this problem in a numerically efficient way. The main objective of this paper is to apply those means for designing a new algorithm for correction of spectrometric data with built-in estimation of uncertainty.

The following rules of notation are used throughout the paper:

- bold-face symbols denote vectors and matrices;
- "·" over a symbol indicates an exact value of a quantity denoted by this symbol;
- "~" over a symbol indicates a disturbed value of a quantity denoted by this symbol;
- "^" over a symbol indicates an estimate of the quantity denoted by this symbol;
- "Δ" preceding a symbol indicates the absolute error of the quantity denoted by this symbol;
- underlined symbols denote random variables, e.g. $\underline{\hat{\mathbf{x}}}$ denotes the random vector modelling errors in the final result of measurement $\hat{\mathbf{x}}$;
- symbols in brackets $[\cdot]$ denote intervals, e.g. $[\Delta \hat{x}_n] = [\Delta \hat{x}_n^{\text{inf}}, \Delta \hat{x}_n^{\text{sup}}]$ denotes the interval modelling the uncertainty corrupting the n -th component of the final result of measurement $\hat{\mathbf{x}}$.

The abbreviation ACSD denotes an algorithm for correction of spectrometric data, transforming the vector of uncertain data $\tilde{\mathbf{d}} = [\tilde{\mathbf{y}}^T \hat{\mathbf{p}}^T]^T$ into the vector of uncertain results $\hat{\mathbf{x}}$. It is assumed that this

algorithm may be decomposed into some algebraic operations - $v_0 = f(v_1, v_2, \dots)$ - transforming the intermediate results of computation - $v_1, v_2, \dots$ - into an intermediate or final result of computation v_0 . The values of the derivatives of the function $f(v_1, v_2, \dots)$ are denoted as follows:

$$f'_i = \left. \frac{\partial f(v_1, v_2, \dots)}{\partial v_i} \right|_{\substack{v_1=\hat{v}_1 \\ v_2=\hat{v}_2}} \quad \text{for } i = 1, 2, \dots \quad f''_{i,j} = \left. \frac{\partial^2 f(v_1, v_2, \dots)}{\partial v_i \partial v_j} \right|_{\substack{v_1=\hat{v}_1 \\ v_2=\hat{v}_2}} \quad \text{for } i, j = 1, 2, \dots \quad (2)$$

The uncertainties related to the errors corrupting the data:

$$\Delta \tilde{\mathbf{d}} = \tilde{\mathbf{d}} - \mathbf{d} \quad (3)$$

as well as those related to the final results of computation:

$$\Delta \hat{\mathbf{x}} = \hat{\mathbf{x}} - \mathbf{x} \quad (4)$$

and intermediate results of computation:

$$\Delta \hat{v}_i = \hat{v}_i - v_i \quad \text{for } i = 0, 1, 2, \dots \quad (5)$$

are characterised by intervals:

$$[\Delta \tilde{\mathbf{d}}], [\Delta \hat{\mathbf{x}}], \text{ and } [\Delta \hat{v}_i] \text{ for } i = 0, 1, 2, \dots$$

2 FORMULATION OF THE RESEARCH PROBLEM

If the analysed ACSD:

- is linear with respect to the data $\tilde{\mathbf{y}}$ and $\hat{\mathbf{p}}$, and with respect to all intermediate results,
 - does not use the data or intermediate results more than once,
- then the linear propagation of uncertainties [19], executed step by step for each operation of ACSD, yields a satisfactory solution to the problem of uncertainty estimation. Otherwise, the following difficulties appear:
- if ACSD contains non-linear operations, then the results of those operations are biased, and - consequently - biases of errors corrupting the intermediate results should be propagated together with the random uncertainties;
 - if ACSD uses some data or intermediate results more than once, then a tendency for overestimating the uncertainties appears.

The first difficulty may be relatively easily overcome if the interval analysis is applied; the second one requires the monitoring of correlation among the errors corrupting the data ($\tilde{\mathbf{y}}$ and $\hat{\mathbf{p}}$) and intermediate results of computation during the execution of an ACSD. The methodology for solving this problem, developed in [20], is based on the following assumptions:

- The expanded uncertainties of the data may be adequately expressed by the linearly independent intervals $[\Delta \tilde{\mathbf{d}}_k] = [\Delta \tilde{\mathbf{d}}_k^{\text{inf}}, \Delta \tilde{\mathbf{d}}_k^{\text{sup}}]$ with $\Delta \tilde{\mathbf{d}}_k^{\text{inf}} = -\Delta \tilde{\mathbf{d}}_k^{\text{sup}}$ ($k = 1, 2, \dots$)
- The analysed ACSDs are only weakly non-linear, i.e. they are composed of the operations which may be locally approximated by the second-order Taylor expansion:

$$\Delta \hat{v}_0 \cong f'_1 \Delta \hat{v}_1 + f'_2 \Delta \hat{v}_2 + \dots + \frac{1}{2} f''_{11} (\Delta \hat{v}_1)^2 + \frac{1}{2} f''_{22} (\Delta \hat{v}_2)^2 + f''_{12} \Delta \hat{v}_1 \Delta \hat{v}_2 + \dots \quad (6)$$

- The errors in the data are small enough to keep satisfied the conditions:

$$|\Delta \hat{v}_0 - f'_1 \Delta \hat{v}_1 + f'_2 \Delta \hat{v}_2 + \dots| \ll |f'_1 \Delta \hat{v}_1 + f'_2 \Delta \hat{v}_2 + \dots| \quad (7)$$

for all the operations the analysed ACSD is composed of.

Moreover, it is assumed that only the uncertainty component "inherited" from the errors in the data is considered. Another significant component - always present in the result of correction obtained by means of an ACSD with an built-in mechanism of regularisation - is caused by this mechanism; its analysis requires mathematical means - as a rule - specific for each ACSD so it is not considered here.

3 IMPLEMENTED METHOD OF UNCERTAINTY ESTIMATION

A linear ACSDs is composed of the operations of the form:

$$v_0 = f'_1 v_1 + f'_2 v_2 + \dots + \text{const.} \quad (8)$$

The method for tracking the correlation among the errors corrupting the intermediate results of computation, proposed in [20], is based on expressing those errors by linear combinations of the errors in the data:

$$[\Delta \hat{v}_i] = \mathbf{a}_i^T [\Delta \tilde{\mathbf{d}}] \quad \text{for } i = 0, 1, 2, \dots \quad (9)$$

The expanded uncertainties of two intermediate results of computation, e.g. $\hat{v}_i$ and $\hat{v}_j$, could be then expressed (in terms of intervals) as:

$$[\Delta \hat{v}_0] = \mathbf{a}_0^T [\Delta \tilde{\mathbf{d}}], \quad [\Delta \hat{v}_1] = \mathbf{a}_1^T [\Delta \tilde{\mathbf{d}}] \quad \text{and} \quad [\Delta \hat{v}_2] = \mathbf{a}_2^T [\Delta \tilde{\mathbf{d}}] \quad (10)$$

where:

$$\mathbf{a}_0 = f'_1 \mathbf{a}_1 + f'_2 \mathbf{a}_2 + \dots \quad (11)$$

For a weakly non-linear ACSD, composed of the operations that may be locally approximated by the second-order Taylor expansion, the formula for modelling the errors is generalised in the following way:

$$[\Delta \hat{v}_i] = \mathbf{a}_i^T [\Delta \tilde{\mathbf{d}}] + b_i \quad \text{for } i = 0, 1, 2, \dots \quad (12)$$

where $\mathbf{a}_i$ are the same vectors as in Eq.(9), and b_i are displacements of the intervals with respect to zero. The key idea underlying this solution is the choice of vectors $\Delta \tilde{\mathbf{d}}^{\min}$ and $\Delta \tilde{\mathbf{d}}^{\max}$ which generate a realistic estimate of the interval $[\hat{v}_0]$:

$$\hat{v}_0^{\inf} = f(\dot{v}_1 + \mathbf{a}_1^T \Delta \tilde{\mathbf{d}}^{\min} + b_1, \dot{v}_2 + \mathbf{a}_2^T \Delta \tilde{\mathbf{d}}^{\min} + b_2, \dots) \quad \text{and} \quad \hat{v}_0^{\sup} = f(\dot{v}_1 + \mathbf{a}_1^T \Delta \tilde{\mathbf{d}}^{\max} + b_1, \dot{v}_2 + \mathbf{a}_2^T \Delta \tilde{\mathbf{d}}^{\max} + b_2, \dots) \quad (13)$$

The vectors $\Delta \tilde{\mathbf{d}}^{\min}$ and $\Delta \tilde{\mathbf{d}}^{\max}$ are constructed using information contained in the vector $\mathbf{a}_0$ computed according to Eq.(11), viz.:

$$\Delta \tilde{\mathbf{d}}_k^{\max} = -\Delta \tilde{\mathbf{d}}_k^{\min} = \begin{cases} \Delta \tilde{\mathbf{d}}_k^{\sup} & \text{if } a_{0,k} > 0 \\ 0 & \text{if } a_{0,k} = 0 \\ \Delta \tilde{\mathbf{d}}_k^{\inf} & \text{if } a_{0,k} < 0 \end{cases} \quad (14)$$

for $k = 0, 1, 2, \dots$. The interval displacement corresponding to Eq.(12) may be then estimated using the formula:

$$\hat{b}_0 = \frac{1}{2}(\hat{v}_0^{\inf} + \hat{v}_0^{\sup}) - f(\dot{v}_1, \dot{v}_2, \dots) \quad (15)$$

Example: Let us consider a simple non-linear operation:

$$v_0 = f(v_1, v_2) = \frac{v_1^2}{1 + v_2} \quad \text{for } \dot{v}_1 = 2, \text{ and } \dot{v}_2 = 1.$$

Let us assume that:

$$\Delta \tilde{\mathbf{d}}_k \in [-0.005, +0.005] \quad \text{for } k = 1, 2, 3, \quad \mathbf{a}_1 = [5 \ -5 \ 5]^T, \quad b_1 = 0.01, \quad \mathbf{a}_2 = [11 \ -11 \ 10]^T, \quad \text{and } b_2 = -0.01$$

The corresponding interval representing the uncertainty of the result of computation is $[\Delta \hat{v}_0] = [0.0201, 0.0460]$; thus $b_0 = 0.0331$. According to the applied methodology of analysis, one obtains $f'_1 = 2$ and $f'_2 = -1$, and consequently:

$$\mathbf{a}_0 = 2\mathbf{a}_1 - \mathbf{a}_2 = [-1 \ 1 \ 0]^T$$

$$\Delta \tilde{\mathbf{d}}^{\max} = [-0.005 + 0.005 \pm 0]^T \quad \text{and} \quad \Delta \tilde{\mathbf{d}}^{\min} = [+0.005 - 0.005 \pm 0]^T$$

$$\hat{v}_0^{\sup} = \frac{(\dot{v}_1 + \mathbf{a}_1^T \Delta \tilde{\mathbf{d}}^{\max} + b_1)^2}{1 + \dot{v}_2 + \mathbf{a}_2^T \Delta \tilde{\mathbf{d}}^{\max} + b_2} = 2.0434 \quad \text{and} \quad \hat{v}_0^{\inf} = \frac{(\dot{v}_1 + \mathbf{a}_1^T \Delta \tilde{\mathbf{d}}^{\min} + b_1)^2}{1 + \dot{v}_2 + \mathbf{a}_2^T \Delta \tilde{\mathbf{d}}^{\min} + b_2} = 2.0208$$

The corresponding estimate of the displacement b_0 is $\hat{b}_0 = 0.0321$. The only reason for the difference $\hat{b}_0 \neq b_0$ is the non-informative element of the vector $\mathbf{a}_0$, $a_{0,3} = 0$, which makes impossible precise identification of the vectors:

$$\Delta \tilde{\mathbf{d}}^{\max} = [-0.005 \ 0.005 \ -0.005]^T \quad \text{and} \quad \Delta \tilde{\mathbf{d}}^{\min} = [+0.005 \ -0.005 \ -0.005]^T$$

Nevertheless, the obtained result of analysis is significantly more precise than that yielded by the standard interval formula, viz.:

$$[\hat{v}_0] = \frac{(\dot{v}_1 + \mathbf{a}_1^T [\Delta \tilde{\mathbf{d}}] + b_1)^2}{1 + \dot{v}_2 + \mathbf{a}_2^T [\Delta \tilde{\mathbf{d}}] + b_2} = [1.6968, 2.4328]$$

The corresponding estimate of the displacement b_0 , $\hat{b}'_0 = 0.064$, would be subject to 100 % error, and the span $v_0^{\sup} - v_0^{\inf}$ would be overestimated even more! ♣

4 PROPOSED ALGORITHM FOR CORRECTION OF SPECTROMETRIC DATA

The following non-linear, recursive formula was used for designing a new ACSD with built-in estimation of uncertainty:

$$\hat{x}_n = c_{-1}\hat{x}_{n-1} + \frac{1}{1 + \left[\sum_{k=-3}^3 b_k (\tilde{y}_{n-k}^2 + a_1 y_{n-k} + a_0) \right]^2} - \frac{1}{2} \quad (16)$$

where:

$$\mathbf{p} = [a_0 \ a_1 \ b_{-3} \ b_{-2} \ b_{-1} \ b_0 \ b_1 \ b_2 \ b_3 \ c_{-1}]^T \quad (17)$$

is a vector of parameters to be estimated during calibration of the spectrometer. This formula contains a nonlinear transformation of a quadratic FIT-type filter. The obvious superiority of the operator defined by Eq.(16) over this quadratic filter is that it imposes a physically justified constraint on the solution, viz.: $\hat{x}_n - c_{-1}\hat{x}_{n-1} \in [-0.5, +0.5]$.

5 RESULTS OF STUDY OBTAINED FOR SYNTHETIC DATA

First, a genetic-algorithm-based procedure described in [18], as well as the test signals $\{\dot{x}_n^{cal}\}$ and $\{\tilde{y}_n^{cal}\}$ shown in Fig. 1, were used for calibration which yielded the following result:

$$\hat{\mathbf{p}} = [2.8455 \ -3.3464 \ -0.0048 \ 0.9940 \ -0.3536 \ -0.5958 \ -0.3467 \ 1.5121 \ -0.8536 \ 0.6867]^T \quad (18)$$

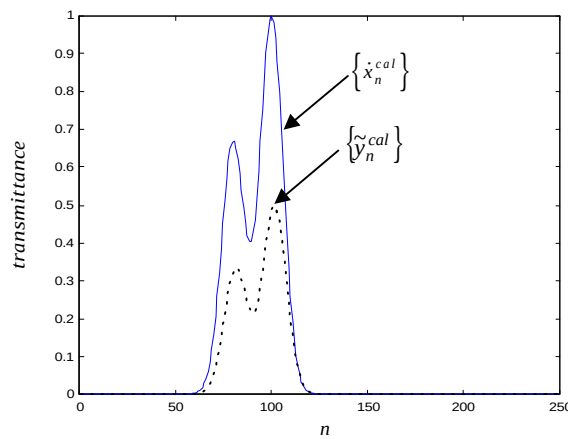


Fig. 1 Spectrometric-type synthetic signals used for estimating the parameters of the proposed ACSD

Next, the spectrum $\{\dot{x}_n\}$ shown in Fig. 2 was reconstructed on the basis of the synthetic data $\{\tilde{y}_n\}$ - also shown in Fig. 2 - using the proposed ACSD. The built-in procedure for analysis of the uncertainty was based on the assumption that:

- the raw measurement data are subject to errors $\Delta \tilde{y}_n = \sup \{|\dot{y}_n|\} h_n$, where $|h_n| \leq 10^{-3}$;
- the estimates of the parameters are subject to errors $\Delta \hat{p}_k = \dot{p}_k e_k$, where $|e_k| \leq 10^{-3}$.

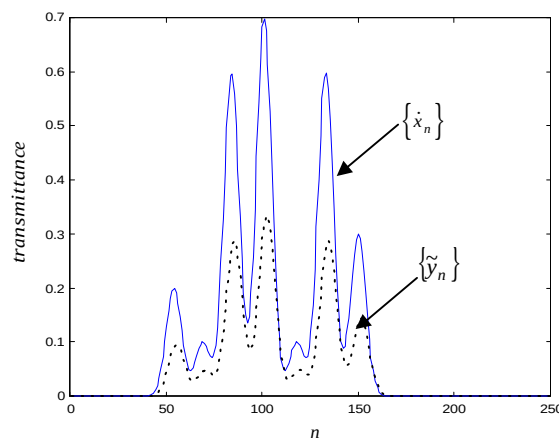


Fig. 2 Synthetic signals used for testing the proposed ACSD with built-in evaluation of uncertainty

Finally, the expanded uncertainty of correction obtained by means of the proposed method was compared with the corresponding estimate obtained by means of the Monte Carlo simulation technique; the result of comparison is shown in Fig. 3.

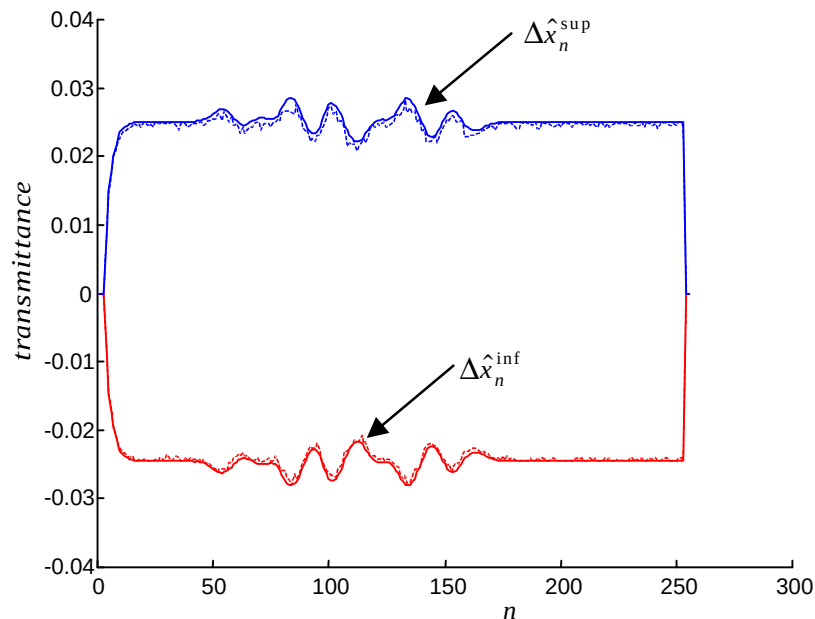


Fig. 3 Expanded uncertainty of correction $[\Delta\hat{\chi}_n^{\text{inf}}, \Delta\hat{\chi}_n^{\text{sup}}]$: an estimate obtained by means of the proposed method (solid line) compared with the corresponding estimate obtained by means of the Monte Carlo technique (dashed line)

6 CONCLUSION

A new algorithm for correction of spectrometric data with built-in evaluation of the uncertainty of the results of correction has been proposed. The main conclusions which may be drawn from its investigation, based on synthetic spectrometric data, are the following:

- It is able to effectively deal with overestimation of uncertainty characteristic of widely used methods of error and uncertainty propagation.
- It is much less time consuming than highly reliable Monte Carlo techniques or fully reliable worst-case simulation technique (the proposed method required ca. 30 times less computer time than the Monte Carlo technique).

Acknowledgement:

This work was supported by the State Committee for Scientific Research, Poland (grants # 8 T10C 025 11 and # 8 T10C 040 17), and by the Institute of Radioelectronics, Faculty of Electronics and Information Technology, Warsaw University of Technology (from the statutory fund).

REFERENCES

- [1] R. Z. Morawski: "Unified Approach to Measurand Reconstruction". *IEEE Trans. Instrum. & Meas.*, Vol. 43, No 2, April 1994, pp. 226-231.
- [2] A. Miękina, R. Z. Morawski, A. Podgórski: "Dynamic Reconstruction of Measurands and Calibration of Measuring System", *Measurement - Journal of IMEKO*, Sept. 1994, Vol. 14, No. 1, pp. 63-72.
- [3] P. A. Jansson (ed.), *Deconvolution of Spectra and Images*, Academic Press Inc., New York, 1997.
- [4] R. Z. Morawski, M. Ben Slima, M. Milewski, A. Barwicz: "Application of a Digital Signal Processor for Correction of Spectrophotometric Measurements", *IEEE Trans. Instrum. & Meas.*, June 1993, Vol. 43, No. 3, pp. 778-782.

- [5] L. Szczeciński, R. Z. Morawski, A. Barwicz: "Original-domain Tikhonov Regularisation and Non-negativity Constraint Improve Resolution of Spectrometric Analysis", *Proc. XIIIth IMEKO World Congress (Torino, Sept. 5-9, 1994)*, pp. 441-446.
- [6] R. Z. Morawski, B. Pawłowski: "Improving Resolution of Spectrometric Analysis by Means of Adjoint-operator Method and B-splines", *Proc. 6th Int. Conf. Industrial Metrology CIMI'95 (Zaragoza, October 25-27, 1995)*, pp. 382-390.
- [7] R. Z. Morawski, L. Szczeciński, A. Barwicz: „Deconvolution Algorithms for Instrumental Applications: A Comparative Study”, *Journal of Chemometrics*, Vol. 9, 1995, pp. 3-20.
- [8] D. Massicotte, R. Z. Morawski, A. Barwicz.: "Incorporation of a Positivity Constraint into a Kalman-filter-based Algorithm for Correction of Spectrometric Data", *IEEE Trans. Instrum. & Meas.*, Vol. 44, No.1, Feb. 1995, pp. 2-7.
- [9] R. Z. Morawski, A. Miękina, A. Barwicz: „Combined Use of Tikhonov Deconvolution and Curve Fitting for Spectrogram Interpretation”, *Instrum. Science & Technol.*, Vol. 24, No 3, August 1996, pp.155-167.
- [10] D. Massicotte, R. Z. Morawski, A. Barwicz: "Kalman-filter-based Algorithms of Spectrophotometric Data Correction - Part I: An Iterative Algorithm of Deconvolution". *IEEE Trans. Instrum. & Meas.*, Vol. 46, No. 3, June 1997, pp. 685-689.
- [11] M. Ben Slima, R. Z. Morawski, A. Barwicz: "Kalman-filter-based Algorithms of Spectrophotometric Data Correction - Part II: Use of Splines for Approximation of Spectra". *IEEE Trans. Instrum. & Meas.*, Vol. 46, No. 3, June 1997, pp. 685-689.
- [12] L. Szczeciński, R. Z. Morawski, A. Barwicz: "A Cubic FIR-type Filter for Numerical Correction of Spectrometric Data". *IEEE Trans. Instrum. & Meas.*, Vol. 46, No. 4, August 1997, pp. 922-928.
- [13] A. Miękina, R. Z. Morawski, A. Barwicz: "The Use of Deconvolution and Iterative Optimization for Spectrogram Interpretation". *IEEE Trans. Instrum. & Meas.*, Vol. 46, No. 4, August 1997, pp. 1049-1053.
- [14] L. Szczeciński, R. Z. Morawski, A. Barwicz, "Variational Algorithms for Spectrogram Correction Based on Entropy-like Criteria", *J. Chemometrics*, Vol. 12, issue 6, 1998, pp. 397 - 403.
- [15] L. Szczeciński, R. Z. Morawski, A. Barwicz: "Numerical Correction of Spectrometric Data Using a Rational Filter", *J. Chemometrics*, Vol. 12, issue 6, 1998, pp. 379 - 395.
- [16] C. Niedziński, R. Z. Morawski: "Bayesian Approach to Spectrophotometric Analysis of Multicomponent Substances", *Proc. IEEE Instrum. & Meas. Technol. Conf. - IMTC'99 (Venice, Italy, May 24-26, 1999)*, str. 1901-1906, (lub CD-ROM #MEE602 Z-6427 A).
- [17] M. P. Wiśniewski, R. Z. Morawski, A. Barwicz: "An Adaptive Rational Filter for Interpretation of Spectrometric Data", *Proc. IEEE Instrum. & Meas. Technol. Conf. - IMTC'99 (Venice, Italy, May 24-26, 1999)*, str. 1907-1912 (lub CD-ROM #MEE602 Z-6427 A).
- [18] P. Sprzączak, R. Z. Morawski: "Calibration of a Spectrometer Using a Genetic Algorithm", *Proc. IEEE Instrum. & Meas. Technol. Conf. - IMTC'99 (Venice, Italy, May 24-26, 1999)*, str. 1027-1032 (lub CD-ROM #MEE602 Z-6427 A).
- [19] *Guide to the Expression of Uncertainty in Measurement*, BIMP-IEC-IFCC-ISO-IUPAC-IUPAP-OIML, ISO 1993.
- [20] T. Szafranski, R. Z. Morawski: "Dealing with Overestimation of Uncertainty in Algorithms of Mesurand Reconstruction", *Proc. XVth IMEKO World Congress (Osaka, Japan, June 13-18, 1999)*, CD-ROM #IFPI L263 IMEKO-15 MT C01

AUTHOR(S): TOMASZ SZAFRAŃSKI, PIOTR SPRZĘCZAK, ROMAN Z. MORAWSKI: Warsaw University of Technology, Faculty of Electronics and Information Technology, Institute of Radioelectronics, ul. Nowowiejska 15/19, 00-665 Warsaw, Poland, Phone: (+48 22) 660-7721, Fax: (+48 22) 825-5248, E-mail: R.Morawski@ire.pw.edu.pl.