



Using of the undetectable perturbations for determining confidence intervals of parameters of overlapping asymmetrical peaks estimated by nonlinear curve fitting

J. Dubrovkin

Multidisciplinary Department, Western Galilee College
2421 Acre, Israel

Abstract: Decomposition of overlapping peaks is one of the most commonly used computer-enhanced procedures of instrumental analytical chemistry which is usually performed by nonlinear fitting a predefined model to the measured data. The minimum of the squared norm of the residuals between the model and the measured data is commonly served as a measure of goodness of fit. However, the minimum does not guarantee a unique solution of the fitting. Also confidence interval radii of the estimated model parameters are often overestimated.

To solve these problems we have developed a new method of determining the uncertainty (named the undetectable perturbations (UPs) method) of the correlated peak parameters based on a predetermined norm of the residuals. A multiple band-decomposition procedure based on the variation of the initial peak parameters was chosen to guarantee the unique solution with a minimal uncertainty.

The fluorescence spectra of solutions of fluorescein, trans-stilbene and benzene were decomposed. The polynomial modified Gaussians (PMG), weighted sum of Lorentzian and PMG; double asymmetrical sigmoid and asymmetrical log-normal functions were used as the models. The estimated model parameters were constrained by the values which had physical meaning. The confidence intervals of the asymmetrical peak parameters were evaluated using the linear approximation, the bootstrap and the UPs methods. It was found that the linear approximation and the bootstrap methods often overestimated radii of the confidence intervals. Use the UPs method allowed us to control the radii according to a predetermined level of the squared norm of the residuals as opposed to the common methods.

Mathematical and physical justifications of the evaluated model parameters were discussed.

Keywords: overlapping peaks, curve fitting, uncertainties in parameter estimates, undetectable perturbations

I. Introduction

In the last decades instrumental methods of analytical chemistry have gained a dominant position in the process analytical technology and in academic research. Information obtained by analytical methods is called analytical signals [1]. The most commonly used one-dimensional signals are modelled as a set of the bell-shaped peaks and some background. Peak parameters (maximum position and intensity, width, area and asymmetry coefficients) are the main sources of analytical information. Unfortunately, these peaks are often overlapping [1]. Since in many practical cases the instrumental peak separation is impossible (or very time consuming), the mathematical processing of the overlapping peaks is needed.

For this goal two main groups of the mathematical methods are commonly used. In the first group the required chemical-physical property may be evaluated without estimating the peak parameters or only some parameters are needed. E.g., the amounts of mixture components were calculated by solving systems of equations [1] or using the ratio methods [2]. The maximum positions of the overlapping peaks may be determined using the second order derivative of the measured signal or Fourier transform for deconvolution of noisy signals [3].

In the second group all peak parameters are estimated, that is, the signal is separated into its "pure" peaks. For this processing a predefined mathematical model of the measured signal must be created [1, 3 and references within]. The model parameters should be chosen so that the difference between the measured signal and the model (the curve-fit residuals) was minimal. Such model is called "the best-fit" model. The mathematical measure of the fitting is a squared norm of this difference. The fitting procedure is usually performed by the nonlinear curve fitting least squares algorithms which are searching for the parameters of the best-fit model [3, 4]. These algorithms are included in software of the modern analytical instruments [5]. Unfortunately, curve fitting does not guarantee the unique solution since the solution strongly depends on the initial guesses of the peak parameters needed for decomposition procedures [3, 6].

The uncertainty in determining peak parameters obtained by the separation algorithms is usually expressed in terms of the standard deviations of the estimated model parameters using a linear approximation of nonlinear models [3]. However, this method often overestimates the standard deviations [7].

To overcome this drawback we have developed new method in which the relative errors of the peak parameters of the best-fit model (called undetectable perturbations-UPs) were estimated for a predetermined norm of residuals

[5]. In the present article the *UPs*-method was modified by taking into account correlations between the errors of the peak fitting parameters.

A multiple band-decomposition procedure based on the random variation of the initial peak parameters and on the subsequent choice of competing models was performed to guarantee a unique solution with a minimal uncertainty [6].

The fluorescence spectra of solutions of fluorescein, trans-stilbene and benzene were decomposed. The following asymmetrical functions: the polynomial modified Gaussians (PMG), the weighted sum of Lorentzian and PMG (WLPMG), the double asymmetrical sigmoid (DASIGM) and the asymmetrical log-normal function (ALOGN) were used as the peak models. The estimated model parameters were constrained by the values which had physical meaning. The confidence intervals of asymmetrical peak parameters were evaluated using the linear approximation [4], bootstrap [7] and the *UPs* methods. It was found that the linear approximation and bootstrap methods often overestimate radii of the confidence intervals. Use *UPs* allowed us to control the radii according to a predetermined level of the squared norm of residuals of curve fitting as opposed to the former methods.

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

II. Experimental

The spectra were taken from the Open DataBasePhotochemCAD 2.1 [8-10] with permission granted to me by J.S. Lindsey. For convenience of comparison the spectra were normalized to the maximum unit intensity. The peak models (see Appendix) were empirically selected from the large set of the well-known models [11] according to the following competing criteria: a minimum fitting error to the measured spectrum and a minimum number of peak parameters. Rough initial guesses of the peak parameters were estimated by inspection of the measured and Fourier-regularized second-order derivative spectrum [3] (Figs.1, 2). These parameters were refined by trial and error method by minimizing the norm of the residuals. At the first stage, the refined parameters were taken as initial values using decomposition by the Gauss-Newton algorithm with Levenberg-Marquardt modifications (MATLAB function *nlinfit*). All of the peak parameters have been constrained to be positive numbers. In the case of the rank deficient program warnings of the Gauss-Newton algorithm and abnormal large diagonal values of Jacobian, the obtained data were rejected.

At the second stage, for randomization, the fitting procedure was repeated 5,000 times with varying parameters obtained at the first stage [6]. In each repetition random normal noise with zero mean and standard deviation $\sigma = 0.01$ was added to the spectra. The 95%-confidence intervals of the relative values of these parameters which standard deviations were 0.2 were built. In each repetition of the process the initial guesses were taken from these intervals. This method has allowed us to find the set of the competing best-fit models [6].

The “best” estimated parameters corresponded to the residuals less than 110% of the minimum residual were averaged. The reliability of curve fitting was also controlled by visual inspection of the second order derivative of the measured and model spectra. Bootstrapping was performed 10,000 times.

III. Results and Discussion

Graphical representations of the decomposition are given in Figs. 1 and 2. They also include the second order derivative spectra used only to visually compare the model spectra with the measured spectra. Three best-fit models were obtained for the trans-stilbene spectrum (Fig.1). However, the applicable model cannot be justified using only mathematical consideration based on the minimum root mean squared fitting error:

$$RMSE = \sqrt{\sum_{l=1}^L [F_{ms}(l) - F_{bestfit}(l)]^2 / L}, \quad (1)$$

where F_{ms} and $F_{bestfit}$ are the measured spectrum and the best-fit model of the spectrum, respectively, $l = 1, 2, \dots, L$ is the discrete sample number.

For example, $RMSE = 0.0014$ (DASIGM model) is the minimum value obtained for the entire models. However, the first short-wavelength wide peak (Fig. 1b) is seen very strange. We suppose that the only ALOGN model can be physically justified by visual comparing the pure peaks of the trans-stilbene spectrum with those of the benzene spectrum. In both cases the spectra are the envelope of the progression in the ν_2 vibration of the benzene ring (in the ground electronic state) [12]. The pure peaks of the trans-stilbene and benzene spectra have approximately similar forms.

The parameters of the pure peaks of the fluorescein spectra depend on the model function (Fig.2).

Minimum $RMSE$ – value was obtained for the WLPMG model. However, the correct model cannot be mathematically derived from the decomposition procedure only. Maybe the cumbersome experimental matrix method [13] would help to solve such problem but it is out of the scope of the present paper.

Our estimations showed that in all decomposition procedures the systematical discrepancy between the measured spectrum and its model had a dominant impact on the RMSE – value, that is, RMSE was significantly large than the standard deviation of the noise of the measured spectrum.

Valuable information about the uncertainties of the peak parameters was obtained from the Figs. 3 and 4.

The uncertainties of the parameters of the tiny first peak of the trans-stilbene spectra (Fig. 1c) estimated by the linear approximation (LA) and bootstrap (BS) method (Fig. 3) were so large that they cast doubt on the reliability of these parameters.

This is a well-known fact [3] that the uncertainties in determining parameters of a weak peak strongly overlapped by adjacent intensive peak may be very large. In this case the linear approximation of the non-linear model is unsatisfactory [4]. In other words, the estimated peak parameters are very "vague".

Large correlations between the peak parameters lead to the fact that the derivatives of the fitting spectrum over peak parameters are strongly correlated too, and the estimates of the confidence interval using the inverse matrix in Eq. (A14) are bad.

These considerations are valid for the direct confidence intervals estimation in LA method and for the indirect estimation in the BS resampling process when a set of the direct estimations is processed (see Appendix).

The relative radii of the parameters of the strong long-wavelength 4th peak of the trans-stilbene spectra estimated by LA and BS methods are acceptable since they are not larger than 10% (Fig. 3). This observation can be explained qualitatively by the fact that this peak is overlapped only by the 3rd peak (Fig. 1c). However, significant degree of overlapping may be served only as a possible (and not dominant) reason of large uncertainties. E.g., the only confidence interval radius of the intensity of the 3rd trans-stilbene peak is large although this peak is overlapped by the 2nd and by the 4th peaks (Fig. 1c)

In order to shed some light on the drawbacks of the LA and BS methods, the following numerical experiment design to validate the model spectra obtained by the fitting procedures was carried out. The goal of the experiment was to find all feasible solution of the decomposition of the trans-stilbene spectrum. The peak parameters of the ALOGN model were taken randomly from the confidence intervals estimated by LA method. Only those model spectra were chosen for which RMSE- values were not larger than 0.010 and 0.020, respectively. This procedure was repeated 10⁷ times. Only 189 (0.0019%) and 52818 (0.53%) valid combinations of the peak parameters were found in the first and in the second case, respectively. In other words, the probability that the peak parameters randomly chosen from their confidence intervals will give well-fitting models was negligible.

The estimated confidence intervals of the peak parameters are given in Table 1A (see Appendix). Fig. 1A represents the feasible acceptable models of the trans-stilbene spectrum. In full concordance with Fig. 3 the relative radii of the confidence intervals of the all parameters of the 1st peak and of the parameters w_2, ρ_2 and R_3 are very large.

These considerations prompted us instead of LA and BS confidence intervals to use the intervals estimated by undetectable perturbations (UPs) method (see Appendix). In the UPs method the effective root mean squared error is set to

$$RMSE_{ef} = k_{ef} \sigma / \sqrt{L} \geq RMSE, \quad (2)$$

where $k_{ef} \geq 1$ is a constant, σ is the standard deviation of the noise in the measurement signal.

Mathematical basis of the UPs method ensures that the parameters of the peaks randomly selected from the estimated confidence intervals will provide the correct $RMSE_{ef}$ values in a much larger number of cases than the LA method. E.g., for the ALOGN model of the trans-stilbene spectrum 24.4% and 28.8% of the valid peak parameters combinations were found for $RMSE_{ef} \leq 0.01$ and 0.02, respectively.

Moreover, the probability of the selection of the correct combinations of the peak parameters may be increased up to more than 99% if the correlation between the peak parameters were taken into account [5].

So, the confidence intervals estimated by the UPs method are more reliable than those obtained by the LA method. The overestimation observed using LA and BS is supported by the fact that for $RMSE_{ef} \leq 0.020 \gg RMSE = 0.0059$, the UPs values are significantly less than the radii obtained by LA and BS (Fig. 3).

Abnormally large radii obtained by LA and BS for the fluorescein spectrum are due to the fact that the number of the peak parameters has increased from 8 to 12 when the WLPMG model was used instead of the PMG model (Fig. 4). In this case the sensitivity of the mathematical processing to the small errors in data significantly increases which lead to the instability of the estimated radii [4].

From Figs. 3 and 4 it is clearly seen that the smaller $RMSE_{ef}$, the less UPs.

We conclude that the UPs values for the $RMSE_{ef}$ close to the RMSE may be used as the lower boundaries of the confidence interval radii.

Figure 1. Decomposition of the trans-stilbene (a-b) and the benzene (d) spectra.

$A^{(2)}$ is the second order derivative spectrum.

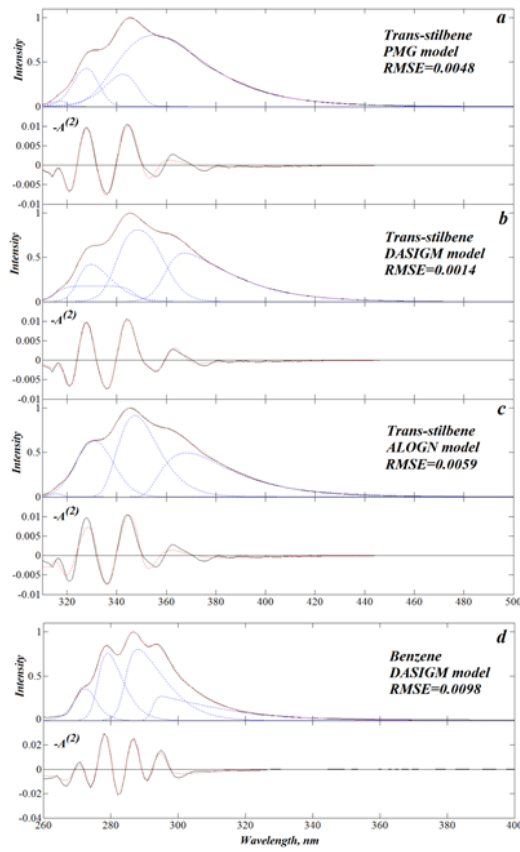


Figure 2. Decomposition of the fluorescein spectrum.

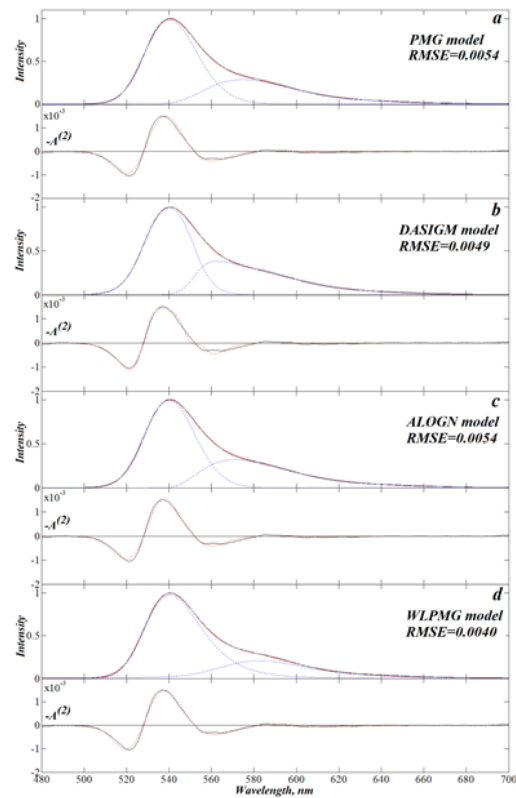


Figure 3. The trans-stilbene spectrum. ALOGN model.

Relative radii of the confidence intervals of the estimated peak parameters.

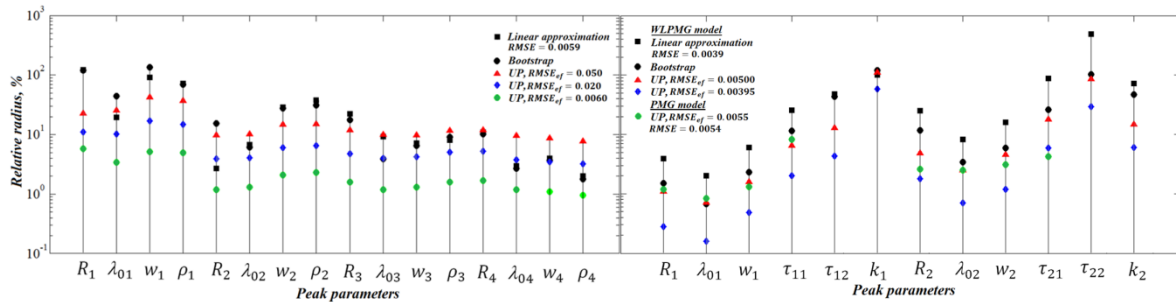


Figure 4. The fluorescein spectrum.

Acknowledgements

The author grateful to J. S. Lindsey (Department of Chemistry North Carolina State University Raleigh) for permission to publish all spectra presented in the article.

Appendix

1. Brief theory of the undetectable perturbation

Suppose that a fix-regressor nonlinear model ($F_{fit}(l, \mathbf{p})$) fits to the analytical signal ($F_{ms}(l)$) disturbed by the normal noise with a mean of zero and a standard deviation σ :

$$F_{ms}(l) = F_{fit}(l, \mathbf{p}_0) + \epsilon_l, \quad (A1)$$

where $l = 1, 2, \dots, L$ is the index of the analytical points, $\mathbf{p}_0$ is a vector of the true model parameters, ϵ_l is a residual.

Also suppose, for simplicity, that the baseline - free model consists of N peaks of the same shape

$$F_{fit}(l, \mathbf{p}) = \sum_{i=1}^N f(l, \mathbf{p}_i), \quad (A2)$$

where $\mathbf{p}_i = p_1, p_2, \dots, p_t$, t is the number of shape parameters, N corresponds strictly to the number of peaks in the analytical signal. The least-squares estimate ("the best-fit") of $\mathbf{p}_0$ denoted by $\hat{\mathbf{p}}_0$, minimizes the objective function over vector $\mathbf{p}$ of the model parameters:

$$\Phi(\hat{\mathbf{p}}_0) = (1/L) \|\Delta F_{bf}(l, \hat{\mathbf{p}}_0)\|^2 = \varepsilon_{bf}^2, \quad (A3)$$

where $\Delta F_{bf}(l, \hat{\mathbf{p}}_0) = F_{ms}(l) - F_{fit}(l, \hat{\mathbf{p}}_0)$, $\varepsilon_{bf} = k_{bf}\sigma$ and k_{bf} is the best-fit empirical coefficient.

ΔF_{bf} (Eq. (A3)) may be approximated by a first-order Taylor polynomial [4]:

$$\Delta F_{bf}(l, \hat{\mathbf{p}}_0) \cong \sum_{i=1}^N \sum_{j=1}^t \frac{\partial F_{fit}(l, \mathbf{p}_{0F})}{\partial p_{ij}} \Delta p_{ij}, \quad (A4)$$

where $\Delta p_{ij} = (p_0)_{ij} - (\hat{p}_0)_{ij}$ is the increment of a peak parameter.

$$\text{The relative increments are equal to } \delta p_{ij} = \Delta p_{ij} / (p_0)_{ij} \quad (A5)$$

To eliminate the influence of the peak position on its relative increment, the last value was calculated relative to the peak width.

By substituting Eqs. (A5) into Eq. (A3) we have

$$\left\| \sum_{i=1}^N \sum_{j=1}^t \frac{\partial F_{fit}(l, \mathbf{p}_{0F})}{\partial p_{ij}} (p_0)_{ij} \delta p_{ij} \right\|^2 = \boldsymbol{\delta}_p \mathbf{A} \boldsymbol{\delta}_p^T \approx L * RMSE^2, \quad (A6)$$

where $\boldsymbol{\delta}_p$ is the vector of the relative increments δp_{ij} , $RMSE = \sqrt{\sum_{l=1}^L [F_{ms}(l) - F_{fit}(l)]^2 / L}$ is the minimum root squared fitting error. The diagonal and the non-diagonal elements of symmetric matrix $\mathbf{A}$ are equal to $\left\| \frac{\partial F_{fit}(l, \mathbf{p}_{0F})}{\partial p_{ij}} \right\|^2 (p_0)_{ij}^2$ and $(p_0)_{ij} (p_0)_{kq} \sum_{l=1}^L \frac{\partial F_{fit}(l, \mathbf{p}_{0F})}{\partial p_{ij}} \frac{\partial F_{fit}(l, \mathbf{p}_{0F})}{\partial p_{kq}}$, respectively.

According to the method of the undetectable perturbations the relative increments of the model parameters must satisfy the inequality:

$$\boldsymbol{\delta}_p \mathbf{A} \boldsymbol{\delta}_p^T = \boldsymbol{\delta}_p \mathbf{U} \boldsymbol{\Lambda} \mathbf{V}^T \boldsymbol{\delta}_p^T = \mathbf{y} \boldsymbol{\Lambda} \mathbf{y}^T \leq L * RMSE_{ef}^2, \quad (A7)$$

where $\mathbf{U}$ and $\mathbf{V}$ are the orthogonal matrices of the left- and the right-singular vectors of the matrix $\mathbf{A}$, $\boldsymbol{\Lambda}$ is diagonal matrix of the singular values (Λ_i), $\mathbf{y} = \boldsymbol{\delta}_p \mathbf{U}, RMSE_{ef} \geq RMSE$ (Eq. (A6)).

Eq. (A7) with the equality sign may be represented as a n -dimensional ellipsoid which center is in the origin and the half-axes are equal to [15]: $\theta_i =$

$$RMSE_{ef} \sqrt{L / \Lambda_i} \quad (A8)$$

The components of the vector $\mathbf{y}$ in the spherical coordinates are [15]:

$$\begin{aligned} y_1 &= \theta_1 \cos(\varphi_1) \\ y_2 &= \theta_2 \sin(\varphi_1) \cos(\varphi_2) \\ y_3 &= \theta_3 \sin(\varphi_1) \sin(\varphi_2) \cos(\varphi_3) \end{aligned} \quad (A9)$$

... ..

$$y_{n-1} = \theta_{n-1} \sin(\varphi_1) \dots \sin(\varphi_{n-2}) \cos(\varphi_{n-1})$$

$$y_n = \theta_n \sin(\varphi_1) \dots \sin(\varphi_{n-2}) \sin(\varphi_{n-1}),$$

where $[\varphi_1, \dots, \varphi_{n-2}] \in [0, \pi]$, $\varphi_n \in [0, 2\pi]$.

Unlike the relative errors, the impact of each component y_i on the $\boldsymbol{\delta}_p \mathbf{A} \boldsymbol{\delta}_p^T$ (Eq. (A7)) does not depend on other components.

Suppose that the angles φ are uniformly distributed in the boundaries given by Eq. (A9) for each y_i independently and the vector

$$\boldsymbol{\delta}_p = \mathbf{y} \mathbf{U}^{-T} \quad (A10)$$

is calculated. Then according to the Central Limit Theorem [15] the relative increment of each parameter $q(\delta p_q)$ will be normally distributed with a zero mean.

The standard deviation $\sigma_{\delta p}$ was considered as a measure of the undetectable relative increment of the corresponding peak parameter. For symmetrical peaks radii of the 95% confidence intervals of the undetectable increments are

$$1.96 \sigma_{\delta x_0} w, 1.96 \sigma_{\delta R} R \text{ and } 1.96 \sigma_{\delta w} w, \quad (A11)$$

where x_0 is the peak maximum, R is the maximum intensity and w is the full width.

The mean value of the squared norm of residuals [2]

$$\overline{\boldsymbol{\delta}_p \mathbf{A} \boldsymbol{\delta}_p^T} = \sum_{i=1}^I A_{ii} \overline{\delta_{p_i}^2} + 2 \sum_{i=1}^I \sum_{j=1}^I A_{ij} \overline{\delta_{p_i} \delta_{p_j}} = \sum_{i=1}^I A_{ii} D(\delta_{p_i}, \delta_{p_i}) + 2 \sum_{i=1}^I \sum_{j=1}^I A_{ij} D(\delta_{p_i}, \delta_{p_j}), \quad (A12)$$

where $D(\delta_{p_i}, \delta_{p_j})$ is the i, j -element of the covariance matrix $\mathbf{D}(\boldsymbol{\delta}_p)$.

In the computer algorithm the angles are randomly selected from the boundaries defined by Eq. (A9). For a given model (matrix $\mathbf{A}$) the vector $\boldsymbol{\delta}_p$ (Eq. (A10)) is calculated under constraints (Eq. (A7)). Data obtained by repeating this procedure $10^5 - 10^6$ times are used for estimating of the sample covariance matrix $\mathbf{D}$.

Eq. (A12) allows estimating the relative weights of $\overline{\delta_{p_i}^2}$ of each peak parameter in residuals.

2. Linear approximation (LA)[4]

For a given peak Eq. (A4) can be represented using Jacobian ($\mathbf{J}$):

$$\Delta F_{bf}(l, \hat{\mathbf{p}}_0) \cong \sum_{j=1}^L \frac{\partial F_{fit}(l, \mathbf{p}_{0F})}{\partial p_j} \Delta p_j = \mathbf{J}(\mathbf{p}) \Delta \mathbf{p}, \quad (\text{A13})$$

$$\text{where } \mathbf{J} = \begin{bmatrix} \left(\frac{\partial F_{fit}}{\partial p_1} \right)_1 & \dots & \left(\frac{\partial F_{fit}}{\partial p_1} \right)_L \\ \vdots & \ddots & \vdots \\ \left(\frac{\partial F_{fit}}{\partial p_t} \right)_1 & \dots & \left(\frac{\partial F_{fit}}{\partial p_t} \right)_L \end{bmatrix}, \Delta \mathbf{p} = \begin{bmatrix} p_1 - p_{01} \\ \vdots \\ p_t - p_{0t} \end{bmatrix}$$

Then the confidence interval radius of an estimated parameter k is equal to

$$Rad_k \cong t_{L-t}^{\alpha/2} (\mathbf{J}^T \mathbf{J})_{kk}^{-1} s, \quad (\text{A14})$$

where $t_{L-t}^{\alpha/2}$ is the Student coefficient for the confidence interval $1 - \alpha$ and $L - t$ degrees of freedom, s is the root mean squared fitting error. In the case of the ill-conditioned matrix $\mathbf{J}^T \mathbf{J}$ the inverse procedure is very sensitive to the computational errors; therefore, the radius may be incorrect.

3. Bootstrap method (BM)

BM is based on the resampling procedure [14]. First of all, from a given dataset the estimate ($\hat{\theta}$) of some unknown parameter (θ) and its standard deviation are calculated. The residuals $\epsilon_l = F_{ms}(l) - F_{bestfit}(l)$ are evaluated. Then a new dataset of the residuals ϵ_l^* is build using the samples of the old dataset of the residuals. Same sample of the first set may appear in the second set many times. The mean and the standard deviation of θ are estimated for new bootstrap set: $F_{ms}^*(l) = F_{ms}(l) + \epsilon_l^*$. This resampling is repeated many times. If the 95% confidence interval on $\hat{\theta}$ is calculated, then the "true" estimate $\hat{\theta}$ must be in the percentile range [2.5 - 97.5] % of the bootstrap samples. This range is estimated by processing of the all samples.

The BM gives exact estimates for the confidence intervals if the number of the repetitions is very large [7].

Therefore, this method is very time consuming. The MATLAB function *bootci* was used for calculation of the bootstrap confidence interval.

4. Peak models.

Notation:

R is the intensity of the peak maximum.

$x = \lambda/w$ is a dimensionless abscissa of the spectrum, w is the peak width,

r is the peak width on the x -scale.

$\lambda_l = l\Delta h$ is the wavelength, $l = [-T, -T + 1, \dots, T - 1, T]$, Δh is the constant sampling interval, $x_0 = \lambda_0/w$ is the position of the peak maximum on the x -scale, λ_0 is the position of the peak maximum,

τ and ρ are the asymmetry coefficients.

The polynomial modified Gaussian [11].

$$PMG = R \exp\{-[y/(1 - \tau_1 y)]^2\}, \quad (\text{A15})$$

where $y = 2\sqrt{\ln 2}(x - x_0)/r$.

The weighted sum of Lorentzian and extended PMG

$$WLPMG = R \left\{ \frac{k}{[1 + 4((x - x_0)/r)^2]} + (1 - k) \exp\left\{-\left[\frac{y}{(1 - \tau_1 y + \tau_2 y^2)}\right]^2\right\} \right\}, \quad (\text{A16})$$

where $k \in [0.1]$.

The double asymmetrical sigmoid

$$DASIGM = R / \max(\Psi), \quad (\text{A17})$$

where $\Psi = \{[1 + \exp(-arg1)][1 + \exp(arg2)]\}^{-1}$,

$arg1 = (x - x_0 + 0.5)/r$, $arg2 = (x - x_0 - 0.5)/(\tau r)$, sp_1 and sp_2 are the left and the right form parameters, respectively, $r = sp_1/w$, $\tau = sp_2/sp_1$

The asymmetrical log-normal function

$$ALOGN = R \exp\left\{-\left(\frac{\ln 2}{\ln^2 \rho}\right) \left(\ln \left[\frac{(x - x_0)(\rho^2 - 1)}{r\rho} + 1\right]\right)^2\right\}, \quad (\text{A18})$$

where $\rho \neq 1$.

5. Selection of the feasible acceptable ALOGN models of the trans-stilbene spectrum.

Two sets of the model parameters corresponding to $RMSE=0.010$ and 0.020 , respectively, were randomly chosen from the LA confidence intervals.

Figure 1A. The model spectra for $RMSE = 0.010$ (a) and 0.020 (b), respectively

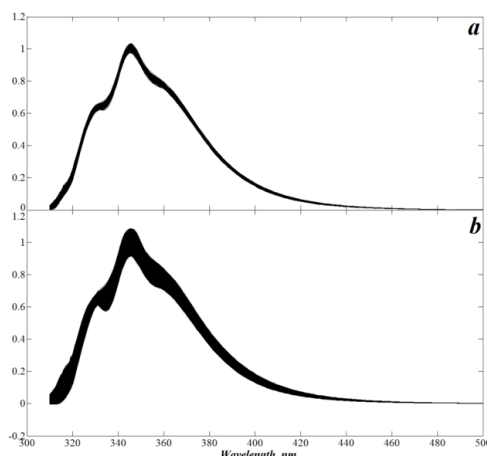


Table 1A. 95% Confidence intervals of the peak parameters.

Parameter	<i>RMSE</i>		
	0.0059(min)	0.010	0.020
R_1	0.0309 ± 0.0380 (123%)	0.0424 ± 0.0405 (96.6%)	0.0327 ± 0.0430 (132%)
λ_{01}	315.5 ± 1.1 (19.3%*)	315.5 ± 1.3 (20.8%*)	315.5 ± 1.3 (21.7%*)
w_1	5.7 ± 5.2 (90.9%)	6.3 ± 5.4 (85.6%)	5.9 ± 5.9 (102%)
ρ_1	0.98 ± 0.71 (72.3%)	1.30 ± 0.38 (29.2%)	1.34 ± 0.42 (31.2%)
R_2	0.63 ± 0.017 (2.7%)	0.63 ± 0.018 (2.9%)	0.63 ± 0.020 (3.1%)
λ_{02}	330.9 ± 1.2 (6.8%*)	331.2 ± 1.1 (5.9%*)	331.0 ± 1.3 (6.8%*)
w_2	17.9 ± 5.2 (29.0%)	19.3 ± 3.9 (20.3%)	19.4 ± 4.0 (20.7%)
ρ_2	1.02 ± 0.38 (37.7%)	1.1 ± 0.23 (20.6%)	1.2 ± 0.25 (21.6%)
R_3	0.92 ± 0.21 (22.5%)	0.86 ± 0.13 (15.4%)	0.86 ± 0.13 (15.5%)
λ_{03}	347.3 ± 1.8 (9.2%*)	347.5 ± 0.98 (5.0%*)	347.5 ± 1.2 (6.3%*)
w_3	19.3 ± 1.4 (7.2%)	19.5 ± 1.4 (7.1%)	19.4 ± 1.5 (7.6%)
ρ_3	1.4 ± 0.11 (8.1%)	1.4 ± 0.11 (7.7%)	1.4 ± 0.13 (9.1%)
R_4	0.50 ± 0.055 (11.1%)	0.49 ± 0.046 (9.3%)	0.50 ± 0.050 (10.1%)
λ_{04}	368.1 ± 1.0 (3.0%*)	368.0 ± 1.1 (3.2%*)	368.0 ± 1.1 (3.3%*)
w_4	33.4 ± 1.3 (4.0%)	33.5 ± 1.3 (4.0%)	33.4 ± 1.5 (4.5%)
ρ_4	2.0 ± 0.04 (2.0%)	2.0 ± 0.04 (2.2%)	2.0 ± 0.05 (2.3%)

The percent values were evaluated using not rounded data. * the relative radius was calculated relative to the peak width.

References

- [1] Danzer K., Analytical Chemistry, Theoretical and Metrological Fundamentals, Springer, 2008.
- [2] Hassan S.A, Elzanfaly E.S, Salem M.Y, et al., Mean centering of double divisor ratio spectra, a novel spectrophotometric method for analysis of ternary mixtures, SpectrochimActa, A. Mol.Biomol.Spectrosc., 2016, 153,132 – 42.
- [3] Dubrovkin J., Mathematical Methods for Separation of Overlapping Asymmetrical Peaks in Spectroscopy and Chromatography. Case study: One-Dimensional Signals, International Journal of Emerging Technologies in Computational and Applied Sciences, 2015, 11 (1), 1-8.
- [4] Seber G. A. F., Wild C. J., Nonlinear Regression, Hoboken, NJ: Wiley-Inter science. 2003.

- [5] Dubrovkin J., Evaluation of undetectable perturbations of the peak parameters estimated by the least square curve fitting of analytical signal consisting of overlapping peaks, Chemom. and Intel. Lab. Syst. 2016,153, 9 - 21.
- [6] Dubrovkin J, Tomin V, Ushakou D., Analytical approximation of the deconvolution of strongly overlapping broad fluorescence bands, J. Appl. Spectr., 2016, 83, 534 -540.
- [7] Pomerantsev A.L., Confidence intervals for nonlinear regression extrapolation, Chemom. Intel. Lab. Syst. 1999, 49, 41-48.
- [8] Open Chemistry Database. 2015. <http://pubchem.ncbi.nlm.nih.gov>;
- [9] Du H, Fuh R.-C.A., Li J, Corkan L.A., Lindsey J.S., Computer-Aided Design and Research Tool in Photochemistry and Photobiology, Photochem. Photobiol, 1998,68,141-142.
- [10] Dixon J.M., Taniguchi M., Lindsey L.S .,PhotochemCAD 2. A Refined Program with Accompanying Spectral Databases for Photochemical Calculations Photochem. Photobiol., 2005,81, 212-213.
- [11] Di Marco V.B, Bombi G.G., Mathematical functions for the representation of chromatographic peaks, J. Chromatogr., A., 2001,931, 1-30.
- [12] Harry ten Blink, Energy Research Centre of the Netherlands, private communication.
- [13] Sjoback R, Nyrgen J, Kubista M., Absorption and fluorescence properties of fluorescein, Spectr. Acta, 1995, A51, L7-L21.
- [14] Efron B. ,Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods, Biometrika, 1981, 68, 589-599.
- [15] Korn G.A., Korn T.M., Mathematical Handbook for Scientists and Engineers, NY: McGraw- Hill, 1961.