# Calibration transfer in model based analysis of second order consecutive reactions

Maryam Khoshkam, Mohsen Kompany-Zareh *

Institute for Advanced Studies in Basic Sciences (IASBS), Department of Chemistry, Zanjan 45137-66731, Iran

## ARTICLE INFO

## ABSTRACT

In the present work, UV–VIS spectroscopic data from a second order consecutive reaction between ortho-amino benzoeic acid (o-ABA) and diazonium ions (DIAZO) with one intermediate was studied. Since o-ABA was not absorbing species in visible region of interest the closure rank deficiency problem did not exist. Analysis of simulated and experimental data shows that in the presence of variations between spectra of pure species in different data matrices, applying the model based methods to augmented datasets leads to inaccurate results. The application of a calibration transfer method as an additional step in the hard modeling procedure improves the precision of results, and accurate estimation of reaction rate constants is obtained. Effect of different types of spectral variations including intensity, shift or broadening is tested in simulated data. The proposed method is compared to local spectra mode of analysis (LSMA) which is proposed by Puxty et al. A comparison of the results shows that the proposed method is more efficient than LSMA and leads to less uncertainty in estimated rate constants and less percent error in the relative residuals.

© 2012 Elsevier B.V. All rights reserved.

## 1. Introduction

In the analysis of kinetic systems, model based methods are rather preferred methods if the model of the kinetic system is known and all variation in the system can be explained in the model [1–4]. These methods are used to determine the kinetic parameters (e.g. rate constants) of a chemical reaction. Two types of data can be used in model based data analysis: 1) a single data matrix or a two-way array and 2) column augmented data matrices [5,6]. Compared to individual data analysis, working with augmented data matrices has some advantages. First, there is typically a significant decrease in the ambiguity of the results because of having more information about the system under investigation: one procedure is monitored under different situations, analyzed simultaneously in the column-wise augmented data matrices [7–9]. Second, using augmented data the problem of rank deficiency can be solved [10–12]. Third, employing augmented data uncertainty in the estimated parameters reduces considerably. In the case of augmentation, the main assumption is that the pure spectral profiles (rows) or concentration profiles of absorbing species (columns) are the same in all augmented data matrices. In other words, the main assumption is the bilinearity of the augmented data [6]. Wavelength shifts or spectroscopic intensity variation due to the misalignment of the monochromator and/or light source can occur within one spectrometer over time or as variation between different spectrometers. This

leads to non-bilinear data after augmentation [13]. These deviations can also occur when spectra are measured under different environmental conditions [14]. One practical limitation to these non-bilinear augmented data sets is the degradation in the performance of the analysis [5,15]. However, it seems that the experimental conditions should be controlled externally in order to avoid the presence of these non-ideal situations when two or more data sets are going to be analyzed simultaneously. Since the external control of experimental parameters is difficult or impossible, development of modern modeling software attempts to incorporate these non-ideal situations into the computations [1]. One of these methods is calibration transfer [16–22]. Calibration transfer makes the measured response obtained from one instrument similar to that which would be obtained on a second instrument [23,24], or in a different environmental condition [25]. Calibration transfer was applied successfully in electrochemical methods [26,27] and for the correction of matrix effects [28]. Calibration transfer entails various methods, both univariate and multivariate, which are discussed in the literature [14,29,30]. In our previous study it was shown that in the presence of non-bilinear augmented data in a first order kinetic system, inaccurate results are obtained. In order to correct for spectral differences a calibration transfer method was used as an intermediate step during the hard modeling procedure [5].

In this study a second order consecutive reaction was followed spectroscopically. It is shown that similar to a previous study, in the case of augmentation of two or more data matrices, ignoring spectral variations leads to systematic error in estimated parameters. In addition it is shown that since the system is very sensitive to errors in the initial concentration of reactants, the system is very sensitive to intensity changes, as well. This was not observed in the previous study since the reaction was a first order one and was not sensitive

* Corresponding author. Tel.: +98 241 4153123; fax: +98 241 4153232.
E-mail address: kompanym@iasbs.ac.ir (M. Kompany-Zareh).

to initial concentrations; so intensity changes were not important in the previous study. The experimental system is a second order consecutive reaction between o-ABA and DIAZO, with one intermediate. In the described system, o-ABA was not absorbing in the visible region of interest and thus, closure rank deficiency problem did not exist. It is shown that analysis of augmented data sets leads to systematic residual without applying calibration transfer.

## 2. Theory

### 2.1. Model based analysis

There are many detailed studies in model based analysis [3,4,31–39]. In a kinetic case the model is the reaction mechanism, concentration matrix $\mathbf{C}$ is defined as a function of time, and the rate constants are parameters. There are many algorithms to optimize the parameters in hard modeling methods [3,6,9,37,38]. In model based analysis it is very important to know the appropriate or correct mechanism under investigation. A brief description of the algorithm used here is as follows: 1) initial estimation of the parameters — accuracy of the initial estimates is not important because they will be optimized through the nonlinear optimization algorithm. 2) Calculation of concentration matrix $\hat{\mathbf{C}}$ ($\hat{\mathbf{C}}$ is an estimation of concentration matrix C from initial estimates of rate constants) — for a specific reaction model the mathematical function is known and the concentration matrix can be estimated. 3) Calculation of matrix $\hat{\mathbf{S}}$ — according to Beer's law, pure spectra of absorbing species are linear parameters and can be estimated by least square of concentration matrix $\hat{\mathbf{C}}$ and data matrix $\mathbf{D}$ ($\hat{\mathbf{S}} = \hat{\mathbf{C}}^{+}\mathbf{D}$) where $\hat{\mathbf{C}}^{+}$ is the pseudo inverse of $\hat{\mathbf{C}}$ [12]. 4) Calculation of the residual matrix $\mathbf{R}$ ($\mathbf{R} = \mathbf{D} - \hat{\mathbf{C}}\hat{\mathbf{S}}$) — if $\mathbf{R}$ is minimal, it means that the optimized parameters are the best estimates; if not, dependent on the optimization algorithm used, a shift vector for the parameters is calculated and a new iteration starting from step (2) is made. This process is continued until convergence where the process is terminated and the optimized parameters are reported [12]. If the data matrices are column wise augmented, it is crucial that the data are bilinear in the column direction.

#### 2.1.1. Error calculations in estimated parameters

Knowing Jacobian matrix ($\mathbf{J}$) the error value for $i_{th}$ estimated parameter, $\sigma_i$, can be calculated as follows:

$$\sigma_i = \sigma_Y \sqrt{h_{ii}} \tag{1}$$

in which $h_{ii}$ is an element of Hessian matrix $\mathbf{H} = \mathbf{J}^T\mathbf{J}$ and $\sigma_Y$ is standard deviation of elements of residual matrix, $\mathbf{R},$ which can be calculated from the following equation:

$$\sigma_Y = \sqrt{\frac{RSS}{(n_t \times n_\lambda) - \left((n_c \times n_\lambda) + n_p\right)}}. \tag{2}$$

In Eq. (2), $RSS$ is the sum of squares of $\mathbf{R}$ matrix elements and the dominator of equation is degrees of freedom and equals to the number of residual elements which represent the number of experimental parameters ($n_t \times n_\lambda$) minus the number of fitted parameters. Scalars $n_t$, $n_\lambda$, $n_c$, and $n_p$ indicate the number of time points, wavelengths, components, and significant principal components, respectively. Error calculations are described completely by Puxty et al. [37].

### 2.2. Calibration transfer

There is much literature explaining the principles behind calibration transfer or standardization methods [13–15,40]. Only a brief description of the Direct Standardization (DS) method will be given here [14]. Three different datasets are required for calibration transfer: 1) calibration set ($\mathbf{D}_1$) which is used to build the calibration model. These samples

are measured on primary instrument or condition. 2) Standardization subset samples $\mathbf{B}_1$ ($n_{ct} \times n_w$) and $\mathbf{B}_2$($n_{ct} \times n_w$) are measured on primary and secondary instruments respectively. These samples are used to calculate the transformation matrix, $\mathbf{T}$. 3) Prediction set ($\mathbf{d}_2$) is measured on secondary instrument or condition. This sample is the unknown sample which should be predicted from calibration model.

The aim of calibration transfer is to transfer all the information of primary instrument or condition into a secondary instrument or condition without explicitly measuring the response matrix of the whole calibration set $\mathbf{D}_2$ on the secondary instrument or condition

$$\mathbf{B}_1 = \mathbf{B}_2\mathbf{T} \tag{3}$$

where $n_{ct}$ is the number of samples in subset, $n_w$ is the number of wavelength and $\mathbf{T}$ ($n_w \times n_w$) is a transformation matrix [16]. In the direct standardization method (DS) $\mathbf{T}$ is calculated from linear regression according to

$$\mathbf{T} = \mathbf{B}^{+}_2\mathbf{B}_1. \tag{4}$$

If $\mathbf{d}_2$ ($1 \times n_w$) is the response vector of an unknown sample from the second instrument, then for prediction of the concentration using a calibration model made from $\mathbf{D}_1$ it is necessary to calculate $\mathbf{d}_1$ according to the following equation:

$$\mathbf{d}_1 = \mathbf{d}_2\mathbf{T}. \tag{5}$$

There are many alternative algorithms for calibration transfer to be found in literature [13,14,16–26].

In this study calibration transfer algorithm is incorporated in the hard modeling process as an additional step, as shown in Fig. 1. In this step, in order to correct the spectral differences between two augmented data sets, calibration transfer is performed between two estimated pure spectral matrices, $\hat{\mathbf{S}}_1$ ($n_s \times n_w$) and $\hat{\mathbf{S}}_2$ ($n_s \times n_w$). The estimated transformation matrix $\mathbf{T} = (\hat{\mathbf{S}}_2^{+} \ \hat{\mathbf{S}}_1)$ can correct the spectral differences between two conditions. By multiplying the second data matrix $\mathbf{D}_2$ by $\mathbf{T}$ spectral differences are corrected and augmentation can be done between $\mathbf{D}_1$ and corrected $\mathbf{D}_2$, as shown in Fig. 1. More explanations about applying calibration transfer in hard modeling procedure are described in our previous work [5].

## 3. Experimental

### 3.1. Simulation conditions

A second order consecutive reaction was simulated.

$$A + B \xrightarrow{k_1} C \xrightarrow{k_2} D$$
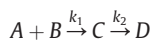
The concentration profiles were obtained from solving the ordinary differential equation in MATLAB [38]. Two sets of concentration profiles ($\mathbf{C}$) were simulated. The initial concentrations were [1 1.5 0 0] in $\mathbf{C}_1$ and [2 1.2 0 0] in $\mathbf{C}_2$ for A, B, C and D species and the simulated rate constants were 0.02 and 0.003 for $k_1$ and $k_2$, respectively. Since in the considered experimental system, the species A is not an absorbing species, three, instead of four, spectral profiles are generated in simulations using Gaussian functions. In order to have non-bilinear data in augmented structure, the second set of simulated spectral profiles (for slave condition) were simulated by modifying the first set of spectral profiles with just intensity changes, 4 nm shifts or broadenings. The concentration profiles and the spectral profiles for master condition are shown in Fig. 2(a–d).