Chemometrics and Intelligent Laboratory Systems 162 (2017) 203–213

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemometrics

Error propagation along the different regions of multivariate curve resolution feasible solutions

Mahsa Dadashi[a,*], Hamid Abdollahi[a], Romà Tauler[b]

[a] Institute for Advanced Studies in Basic Sciences, P.O. Box 45195-1159, Zanjan, Iran
[b] IDAEA-CSIC, Jordi Girona 18-26, Barcelona 08034, Spain

ARTICLE INFO

Keywords:
Rotational ambiguity
Feasible solutions
multivariate curve resolution
Error propagation
Profiles overlapping
Uncertainty
Monte Carlo

ABSTRACT

Evaluation of uncertainties due to rotational ambiguities and noise propagation is essential to ascertain the reliability of Multivariate Curve Resolution estimations. When ambiguity is present, every resolved profile can be represented by a band of feasible solutions instead of by a unique profile. In the presence of experimental noise the estimation of this band of feasible solutions is more difficult and uncertain. The aim of this work is to show how experimental noise and profiles overlapping affect the reliability of feasible solutions. For this purpose, feasible solutions of several two and three components data systems with different profiles overlapping and noise levels have been systematically investigated. Results obtained using simulated and experimental data showed that increasing noise levels and profiles overlapping in one mode (e.g. concentration profiles) make more uncertain the estimation of feasible solutions in the other mode (e.g. spectral profiles), and this uncertainty is propagated in a non-uniform and complex way.

1. Introduction

Multivariate curve resolution (MCR) methods are powerful tools to investigate complex chemical systems whenever there is little or no knowledge about the system [1–4]. Different type of data sets can be analyzed with MCR methods, for instance, spectroscopic multi-wavelength measurements performed on a chemical processes evolving with time [5], hyphenated chromatographic analysis [6], bioanalytical experiments [7], environmental monitoring studies [8], hyperspectral image analysis [9] etc. MCR methods offer a fast and economic tool for mathematical resolution of unknown mixtures using a constrained bilinear model of instrumental multivariate (e.g. multiwavelength) measurements, organized in data tables or data matrices. Often, mathematical resolution of these data matrices does not provide unique solutions; rather there is a range or band of feasible solutions fitting equally well the data and fulfilling the constraints of the system. No matter what MCR method is used, results can be subjected to ambiguities. There are three types of ambiguities in MCR methods: rotational, intensity/scale and permutation ambiguities. These ambiguities can be explained with plain words; the correct data set can be reproduced with component profiles that have different shapes (rotational ambiguity), different magnitude (intensity ambiguity) and different ordering (permutation ambiguity) from the true ones [10,11]. These ambiguities cause uncertainty in MCR results. The amount of ambiguity in MCR solutions is mostly dependent on the degree of complexity of the investigated data set and on the applied constraints. Manne [12] already stated in mathematical terms the local rank and selectivity necessary conditions for correct resolution of concentration and spectral profiles in liquid chromatography diode array coeluted mixtures [13]. Manne published his resolution theorems giving the necessary and sufficient conditions of unique resolution of the component profiles. Rajko and his coworkers further investigated general procedures for uniquely recovering of component profile(s) on the basis of data set structures [14]. To evaluate the amount of ambiguity present in MCR results, different procedures have been proposed [15–21]. In addition to this uncertainty, data noise makes even more difficult the estimation of the effect of rotational ambiguities on the MCR solutions. Finding a way of quantifying the amount of uncertainty in MCR results due to both aspects, rotational ambiguities and noise propagation, is important to further extend the use of MCR methods for analytical purposes and determinations [22].

Different statistical methods have been suggested to estimate the uncertainty of parameters when their direct analytical evaluation is not possible. One strategy for solving this problem is using resampling [23] methods. Resampling and Monte Carlo simulation methods have been already proposed for the estimation of uncertainties of multivariate curve resolution profiles for cases where rotational ambiguities were negligible [24,25]. However, an aspect which is still little documented

* Corresponding author.
E-mail address: m.dadashi357@gmail.com (M. Dadashi).

http://dx.doi.org/10.1016/j.chemolab.2017.01.009
Received 21 September 2016; Received in revised form 20 December 2016; Accepted 10 January 2017
Available online 11 January 2017

in the literature is the reliability and increase of the amount of ambiguity in presence of experimental noise. Noise on experimental data sets influences the calculation of feasible solutions as it has been previously shown [18–20]. For a system in absence of noise, feasible solutions and the extent of ambiguity can be evaluated using different approaches [15,16,21]. When the noise level increases, estimation of rotational ambiguities is more difficult and they are affected. These previous studies however, have only shown the effect of noise on feasible solutions in qualitative terms. It is therefore still necessary to find out ways to evaluate more precisely how experimental errors are propagated to feasible solutions quantitatively. Ideally, evaluation of error propagation effects should be done by experimental replication and recalculation of feasible bands several times. Unfortunately this ideal approach is time consuming and difficult to apply in practice. Noise addition and resampling methods are popular alternative ways to perform error propagation studies due to their relatively easy implementation on modern computers and to their relatively fast application. Fast analytical based algorithm is preferred for these tasks against slow approximation numerical algorithms.

The present paper describes how to perform error propagation studies to calculate quantitatively their effects on the estimation of MCR feasible bands by using Monte Carlo methods. Different HPLC-DAD simulated data sets with two and three components were used as case examples for this study. Additionally, several experimental data sets from the replicates of the UV spectroscopic titration of a the polynuclotide poly(I)-poly(C) were studied to confirm the results obtained in the detailed study of the previously simulated data sets. The more interesting aspects of this work are the proposal of a method for the evaluation of the effect of noise propagation in the different regions of the feasible band solutions and the effect of overlapping profiles on them.

2. Simulated data sets

In this work, different simulated data sets emulating coelution problems in hyphenated chromatography were used to investigate the effect of profiles overlapping and noise propagation on the calculation of Multivariate Curve Resolution feasible solutions. Two and three component systems were simulated. In every case, different data sets with different amount of overlapping (coelution) between concentration profiles were studied. Fig. 1 shows the concentration, **C**, and spectra, $\mathbf{A}^{\mathbf{T}}$, profiles used for the simulation of noiseless data matrices, $\mathbf{D}_{\boldsymbol{O,sim}}$.

$$\mathbf{D}_{0,\mathrm{sim}} = \mathbf{C}\,\mathbf{A}^{\mathbf{T}} \tag{1}$$

where $\mathbf{D}_{\boldsymbol{O,sim}}$ is the data matrix of dimensions (I, J) containing the information about the chemical system under investigation, **C** is the concentration matrix of dimension (I, N) where N is the number of components and $\mathbf{A}^{\mathbf{T}}$ is the spectral matrix with dimensions (N, J). I and J are respectively the number of rows and columns of the data matrix, which for the spectroscopic case refer for instance to the number of spectra and to the number of wavelengths.

2.1. Data sets used to study the effect of profiles overlapping

In Fig. 1, two (upper part) and three (lower part) component concentration and spectra profiles are shown. Concentration (elution) profiles have different overlapping (coelution). These data sets are used to study the effect of profiles overlapping on the calculation of feasible solutions. For this purpose, in this case, the level of noise was fixed to 0.5% of the maximum intensity of the measured signal in each data set. In all cases the same noise level was applied two-hundred times, and added to data set, to give two-hundred new $\mathbf{D}_{\boldsymbol{sim}}$ data matrices. Spectra profiles (last ones on the right of Fig. 1) were the same for the two-components (upper part) and three components (lower part) systems. The correlation coefficient values for simulated data sets are given in Table 1.

2.2. Data sets to study the effect of noise propagation

Two data systems, one with two-components and the other with three components, having the highest overlapping concentration profiles in Fig. 1 (those more on the right of the plot) were selected to investigate the effect of noise propagation on the calculation of feasible solutions. Homoscedastic noise levels with standard deviations of 0.3%, 0.5% and 0.7%, of the maximum data value were considered. Similarly to pervious data in Section 2.1, noise contributions were generated two-hundred times, and used to generate two-hundred new $\mathbf{D}_{\boldsymbol{sim}}$ data matrices.

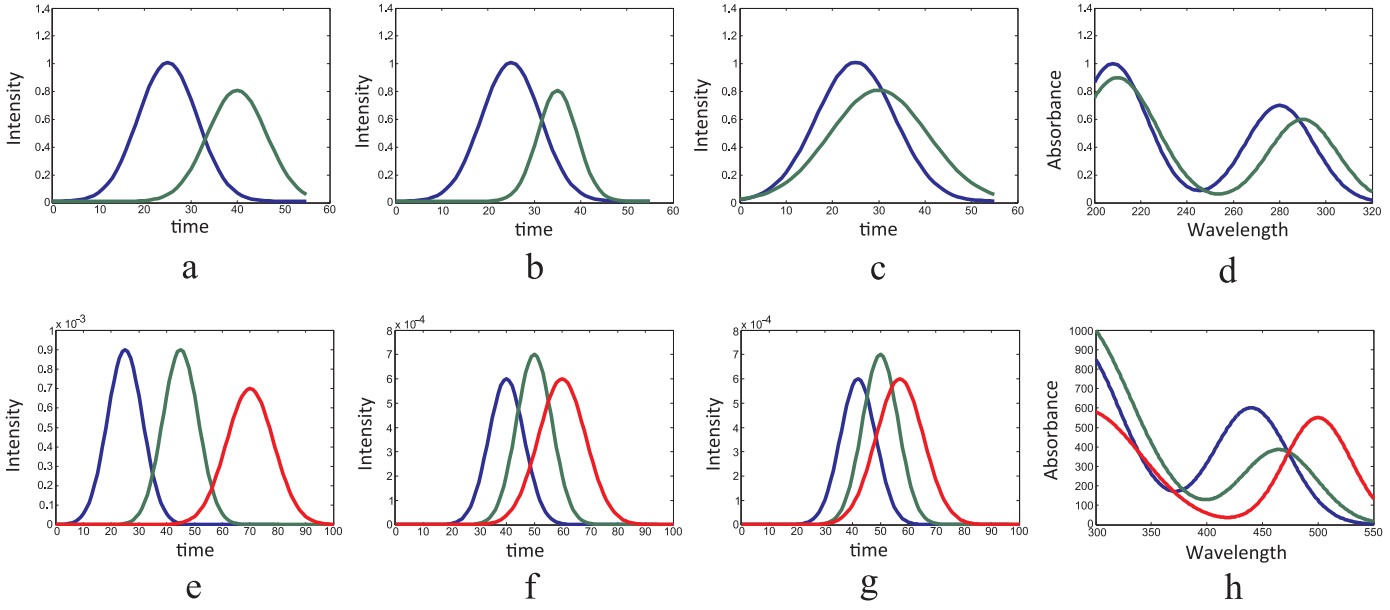

Fig. 1. up) Simulation of the elution/concentration (a, b and c) and spectra (d) profiles of different two-component systems. Spectra are the same in the three cases while elution/concentration profiles overlapping is increasing from left to right. down) Simulation of the elution/concentration (e, f and g) and (h) spectra profiles of different three component systems. Spectra are the same in the three cases while elution/concentration profiles overlapping is increasing from left to right.