Analytica Chimica Acta 911 (2016) 1–13

Contents lists available at ScienceDirect

Analytica Chimica Acta

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

Review

A review of recent methods for the determination of ranges of feasible solutions resulting from soft modelling analyses of multivariate data

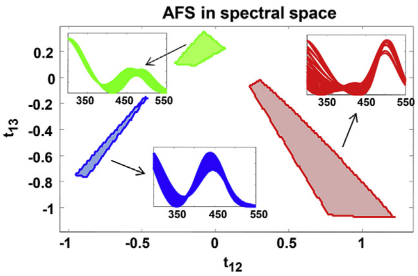
Azadeh Golshan [a, *], Hamid Abdollahi [b], Samira Beyramysoltan [b, c], Marcel Maeder [a], Klaus Neymeyr [d, e], Robert Rajkó [c], Mathias Sawall [e], Romá Tauler [f]

[a] *Department of Chemistry, University of Newcastle, Callaghan, NSW 2308, Australia*
[b] *Faculty of Chemistry, Institute for Advanced Studies in Basic Sciences, P.O. Box 45195-1159, Zanjan, Iran*
[c] *Institute of Process Engineering, Faculty of Engineering, University of Szeged, H-6725 Szeged, Moszkvai krt. 5-7, Hungary*
[d] *Leibniz-Institut für Katalyse, Albert-Einstein-Strasse 29a, 18059 Rostock, Germany*
[e] *Universität Rostock, Institut für Mathematik, Ulmenstrasse 69, 18057 Rostock, Germany*
[f] *IDAEA-CSIC, Jordi Girona 18-24, Barcelona 08034, Spain*

HIGHLIGHTS

- A review on available methods for determination of area feasible solutions (AFS).
- Comparison of the ability of each method in determining the range of feasible bands.
- Summarising the results from analysis of simulated and experimental data sets.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 31 October 2015
Received in revised form
23 December 2015
Accepted 4 January 2016
Available online 11 January 2016

"Dedicated to Prof. Marcel Maeder by the co-authors on the occasion of his retirement"

Keywords:
Soft modelling
Rotational ambiguity
Unique solutions
Band boundaries of feasible solutions

ABSTRACT

Soft modelling or multivariate curve resolution (MCR) are well-known methodologies for the analysis of multivariate data in many different application fields. Results obtained by soft modelling methods are very likely impaired by rotational and scaling ambiguities, i.e. a full range of feasible solutions can describe the data equally well while fulfilling the constraints of the system. These issues are severely limiting the applicability of these methods and therefore, they can be considered as the most challenging ones. The purpose of the current review is to describe and critically compare the available methods that attempt at determining the range of ambiguity for the case of 3-component systems. Theoretical and practical aspects are discussed, based on a collection of simulated examples containing noise-free and noisy data sets as well as an experimental example.

---

* Corresponding author.
E-mail address: azadeh.golshan@uon.edu.au (A. Golshan).

http://dx.doi.org/10.1016/j.aca.2016.01.011

Contents

1. Introduction

Soft modelling analysis methods attempt at the extraction of useful chemical information using a minimum amount of prior information about the process under investigation. In contrast, model-based analyses require well defined prior knowledge about the chemical process. A typical example would be a chemical reaction where the reaction mechanism is prior knowledge required for model-based analysis. In that example the determination of the rate constants for the individual reactions steps is the goal of the analysis. Model-based analyses usually result in unique solutions.

Soft modelling methods always involve multivariate data sets, typically, but not exclusively absorption spectra acquired as a function of the progress of the chemical process. In the following we will use that kind of measurement for the description of the different alternative soft modelling data analysis methods. Such a data set is conveniently written as a matrix, **D**, with the convention that the rows are formed by the spectra and the columns thus are the absorption traces at the different wavelengths. If *ns* spectra are measured at *nl* wavelengths, the dimensions of **D** are $ns \times nl$. Bouguer-Beer-Lambert's law for absorption measurements stipulates a linear relationship between the absorption and the concentrations of the chemical species in the solutions. This of course is the case at all wavelengths. The proportionality factor is the molar absorptivity. This fact allows us to decompose the matrix **D** into the product of two much smaller matrices **C** and **A**.

$$\mathbf{D} = \mathbf{C}\mathbf{A} + \mathbf{E} \tag{1}$$

In this decomposition the columns of **C** are formed by the concentration profiles of the chemical components that take part in the process and the matrix **A** contains in its rows the molar absorption spectra of those components. Also, residuals (noise) or non-modelled parts of **D** are collected in the matrix **E**, which has the same dimension as **D**. If there are *nc* components, the dimensions for **C** are $ns \times nc$ and those of **A** are $nc \times nl$.

Given the measurement **D**, the goal of soft modelling analysis is the determination of the two product matrices **C** and **A** without, or rather with a minimal amount of prior knowledge. The minimal a priori knowledge that is commonly imposed is that both concentrations and molar absorptivities must be positive, thus all elements of **C** and **A** must be positive. Equation (1) is a system of $ns \times nl$ equations, one for each element of **D**, with $(ns + nl) \times nc$ unknowns, the number of elements in **C** and **A**. Since the number of equations is typically much larger than the number of unknowns, i.e. $ns \times nl > (ns + nl) \times nc$, having a unique solution is possible. Unfortunately this is often not the case. In most instances the elements of **C** and **A** are not uniquely defined but there is a range of values for each of these elements that results in the correct matrix **D** after multiplication. Thus, instead of unique profiles, the result of the analysis consists of bands of variable width for the concentration profiles and absorption spectra. This is in contrast to the principal component analysis (PCA) where elements of the resolved profiles are orthogonal and thus generally unique. This problem is commonly known under the expression ‘rotational ambiguity’, a rather unfortunate expression as there is no rotation involved at all.

In this contribution we only apply the standard non-negativity constraint to all elements of **C** and **A**. Often it is possible to apply additional constraints such as unimodality for the concentration profiles, selectivity or closure. If such additional constraints can be imposed, the width of the bands will be reduced and ideally a unique solution can be achieved. Implementation of additional constraints has been discussed in the references [1–4].

The determination of these ranges is not straightforward for $nc > 2$ and the difficulty increases dramatically with increasing number of components. In fact the determination for the 2-component systems has been published in 1971 in a groundbreaking publication by Lawton and Sylvestre which is one of the initializing publications in chemometrics [5]. The extension to 3- and 4-component system has only been published recently [6–13]. In this contribution we will critically analyse and compare these methods.

Soft modelling analysis of chemical data is of course tremendously attractive, resulting in ‘free’ or effortless analysis results. Analysis that can be done completely automatically without user input, are an old dream of the analytical chemist: “The ultimate goal of curve resolution would be to be able to determine the number of components in an overlapping chromatographic peak as well as the spectrum and concentration profile of each compound, without assumption regarding peak shape, location, or identity [14].”

Therefore, not surprisingly, soft modelling analysis methods are well established and have found large numbers of applications in